Dietary metabolite profiling brings new insight into the relationship between nutrition and metabolic risk: An IMI Direct study

Author names: Rebeca Eriksen[1], Isabel Garcia Perez[1], Joram M Posma[2,20], Mark Haid[16], Sapna Sharma[12], Louise Thomas[9], Robert Koivula[3,5], Roberto Bizzotto[3,4], Andrea Mari[4] , Giuseppe N. Giordano[3], Imre Pavo[17], Jochen M. Schwenk[19], Federico De Masi[10], Konstantinos Tsirigos[10], Søren Brunak[10], Ana Viñuela[7], Anubha Mahajan[14], Timothy J. McDonald[11], Tarja Kokkola[15], Harriet Teare[14], Tue H. Hansen[13], Juan Fernandez[14], Angus Jones[11], Chris Jennison[21], Mark Walker[8], Mark I McCarthy[5,14], Oluf Pedersen[13], Hartmut Reutten[18], Ian Forgie[6], Jimmy Bell[9], Ewan Pearson[6], Paul W. Franks[3], Jerzy Adamski[16, 23, 24], Elaine Holmes[1], Gary Frost[1*]

1. Section for Nutrition Research, Division of Digestive Diseases, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, UK

2. Section of Bioinformatics, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College, London, UK

3. Genetic and Molecular Epidemiology Unit, Lund University Diabetes Centre, Department of Clinical Sciences, Lund University, Skåne University Hospital, Malmö, Sweden

4. Institute of Neuroscience - National Research Council, Padova, Italy

5. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Radcliffe Department of Medicine, Oxford, UK

6. Population Health & Genomics, Medical Research Institute, University of Dundee, Dundee, UK

7. Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland

8. Institute of Cellular Medicine (Diabetes), Newcastle University, Newcastle upon Tyne, UK

9. Research Centre for Optimal Health, Department of Life Sciences, University of Westminster, London, UK

10. Department of Bio and Health Informatics, Technical University of Denmark, Lyngby, Denmark

11. Medical School, Exeter, UK NIHR Exeter Clinical Research Facility, University of Exeter

12. German Centre for Diabetes Research, Ingolstädter, Neuherberg, Germany

13. The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Science, University of Copenhagen, Copenhagen, Denmark

14. Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

15. Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland

16. Research Unit Molecular Endocrinology And Metabolism, Helmholtz Zentrum Muenchen, German Research Center for Environemental Health (GmbH), Neuherberg, Germany

17. Eli Lilly Regional Operations GmbH, Vienna, Austria

18. Sanofi-Aventis Deutschland GmbH, R&D, Frankfurt am Main, Germany

19. Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH - Royal Institute of Technology, Stockholm, Sweden

20. Health Data Research UK, London, UK

21. Department of Mathematical Sciences, University of Bath, Bath, UK

22. NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford, UK

23. Lehrstuhl für Experimentelle Genetik, Technische Universität München, 85350 Freising-Weihenstephan, Germany

24. Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597, Singapore

Corresponding authors: Professor Gary Frost, Nutrition Research Section, Division of Digestive Diseases, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, W12 0NN UK. *Telephone:* +44 (0) 20 7594 0959 *Fax number:* +44 (0) 20 8383 8320 *Email:* g.frost@imperial.ac.uk

Summary

Background: Dietary advice remains the cornerstone of prevention and management of type 2 diabetes (T2D). However, understanding the efficacy of dietary interventions is confounded by the challenges inherent in assessing free living diet. Here we profiled dietary metabolites to investigate glycaemic deterioration and cardiometabolic risk in people at risk of or living with T2D.

Methods: We analysed data from plasma collected at baseline and 18-month follow-up in individuals from the Innovative Medicines Initiative (IMI) Diabetes Research on Patient Stratification (DIRECT) cohort 1 n=403 individuals with normal or impaired glucose regulation (prediabetic) and cohort 2 n=458 individuals with new onset of T2D. A dietary metabolite profile model (T_{pred}) was constructed using multivariate regression of 113 plasma metabolites obtained from targeted metabolomics assays. The continuous T_{pred} score was used to explore the relationships between diet, glycaemic deterioration and cardio-metabolic risk via multiple linear regression models.

Findings: A higher T_{pred} was associated with healthier diets high in wholegrain (β=0.004 g, p=0.02 and β=0.003 g, p=0.03) and lower energy intake (β=-0.0002 kcal, p=0.04 and β=-0.0002 kcal, p=0.003), and saturated fat (β=-0.03 g, p<.0001 and β=-0.03 g, p<.0001), respectively for cohort 1 and 2. In both cohorts a higher T_{pred} score was also associated with lower total body adiposity and improved lipid profiles HDL-cholesterol (β =0.07 mmol/L, p<.0001), (β =0.08 mmol/L, p=0.0002), and triglycerides (β =-0.1 mmol/L, p=0.003), (β =-0.2 mmol/L, p=0.0002), respectively for cohort 1 and 2. In cohort 2, the T_{pred} score was negatively associated with liver fat content (β =-0.74 %, p<.0001), and lower fasting concentrations of HbA1c (β =-0.9mmol/mol, p=0.02), glucose (β =-0.2 mmol/L, p=0.01) and insulin (β =-11.0 pmol/mol, p=0.01). Longitudinal analysis showed at 18-month follow up a higher T_{pred} score was also associated lower total body adiposity in both cohorts and lower fasting glucose (β =-0.2 mmol/L, p=0.03) and insulin (β =-9.2 pmol/mol, p=0.04) concentrations in cohort 2.

Interpretation: Plasma dietary metabolite profiling provides objective measures of diet intake, showing a relationship to glycaemic deterioration and cardiometabolic health.

Funding: This work was supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115317 (DIRECT), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies.

Keywords: Metabolic profiling, dietary patterns, type 2 diabetes, cardiometabolic health

Research in context

Evidence before this study

We searched PubMed and Google Scholar for all studies on metabolic profiling of dietary patterns and/or single nutrient intake. Studies were restricted to those with abstracts in English. Our background review showed a wide range of RCT measuring single nutrients metabolites based on small sample sizes and distinct diets. Few studies investigated metabolic profiling derived from multiple metabolites associated with dietary patterns and none was found in our review on a population level. To our knowledge no similar study has been conducted before. We also searched studies investigating metabolomics and risk T2D. A substantial literature exists on metabolites linked with glycaemic and associated metabolic traits, though the majority of these studies were limited to single or a few metabolites specific to phenotypic traits, and none of these included dietary metabolic profiling.

Added value of this study

To our knowledge this study provides the first assessment of combined dietary metabolite profiling and risk of T2D. We show that application of dietary metabolic profiling at the population level provides an objective measurement of dietary patterns and is associated with glycaemic and cardiometabolic risk profiles.
The methodology of the model T_{pred} demonstrate to capture distinct dietary patterns and intake of single nutrients both in plasma samples. Metabolic profiling is a novel and pragmatic approach, which may serve as validation tool for self-reported diet recording and strengthened the precision of diet-disease relationships in epidemiological studies.

Implications of all the available evidence

Finding from our background review and this study show that use of metabolomics is a novel approach in profiling individuals cardiometabolic risk in epidemiological studies. The metabolic profile model T_{pred} is an

objective measurement tool, which should be utilised in nutritional studies to help reduce misreporting and measurement bias existing in traditional nutritional analysis methods.

Introduction

Worldwide, over 425 million people are estimated to be living with type 2 diabetes (T2D) (1). People with T2D have a five-fold risk of developing cardiovascular disease (CVD) and are 1.6 times more likely to die prematurely life expectancy reduced by at least 10 years, compared with those without T2D (2). Identifying high-risk individuals and intervening before diabetes is manifest may disrupt the deterioration of the pancreatic beta cells and minimize damage to the vasculature associated with chronic hyperglycaemia (1).

The aetiology of T2D is multi-factorial, with obesity, poor diet quality, physical inactivity and genetic factors being some of the driving forces (2). Diet is a key modifiable component in the development and management of T2D and associated cardiometabolic risk factors (2-5). The World Health Organisation (WHO) (2) report along with other randomised controlled trials (RCTs) and epidemiological studies, has found the effectiveness of diet in relation to management and prevention of T2D is related to global dietary profiling of numerous nutrients and food groups rather than change in individual nutrients (6-10).

However, assessing a person's diet can be challenging as diet recording is prone to numerous sources of error and bias, such as estimating portion sizes and misreporting by participants (11-13). The extent of underreporting of energy intake in nutritional studies is estimated to be between 30-88 % (11-13). This contributes to data inaccuracy and misinterpretation, which adversely affects the extent to which the effects of diet in health and disease can be estimated(13, 14) Diet underreporting have shown a strong direct relationship with obesity (11-14). A large health survey found analysis based only on plausible respondents re-establishes the theoretical relationship between energy intake and body weight, which a was lost in analysis using samples including misreporters(14). Although it is possible to objectively assess dietary biomarkers, these are generally constrained to a few specific nutrients such as sodium, potassium and nitrogen and are not suitable for the overall assessment of dietary patterns. Evolving discoveries of dietary metabolites may serve as a novel tool in nutritional epidemiology for measuring nutrients and foods in our diets (15-17). Garcia-Perez *et al*. (17) developed a predictive model (T_{pred}) using metabolic profiles to classify diets by training urinary metabolomics data on diet allocation in a cross-over trial of four WHO-defined diets (2). However, many of these studies are limited by their small numbers of participants and use of specific diets. Furthermore, the studies have not investigated the use of the models with health outcomes. More evidence is needed from larger samples size and testing its relationship with dietary patterns in free-living populations and health outcomes.

Here we applied the T_{pred} diet classifier to plasma metabolomics from a free-living population in the Innovative Medicines Initiative (IMI) Diabetes Research on Patient Stratification (DIRECT) Consortium cohorts (https://www.direct-diabetes.org). The richly phenotyped IMI DIRECT cohorts were designed for the discovery of biomarkers for glycaemic deterioration in individuals at risk of or diagnosed with T2D (18). In this analysis, we applied the T_{pred} score in the IMI DIRECT cohorts to explore the relationships between diet, glycaemic deterioration and cardiometabolic risk.

Materials and methods

Study design and participants

The IMI DIRECT multicentre study is a European Union Innovative Medicines Initiative project collaborating among investigators from European leading academic institutions and pharmaceutical companies. The overarching objective of the DIRECT study is to discover and validate biomarkers of glycaemic deterioration before and after onset of T2D and has been reported in detail elsewhere(19). DIRECT established two multicentre prospective cohort studies comprised of adults of Northern European-ancestry; cohort 1 n=2127 participants with normal or impaired glucose regulation (pre-diabetic) and cohort 2 n=789 participants with new onset T2D. Study inclusion and exclusion criteria for cohort 1 and 2 are outlined in table 1. Screening examinations at baseline and 18 months follow up were carried out the morning after a 10-hour overnight fast in the DIRECT study centres by trained nurses. Study protocol have been described in details elsewhere (18). Loss of follow up at 18 months were 138 (6.5%) from cohort 1 and 121 (15.3%) from cohort 2.
This manuscript analyses included all participants from the five study centres Malmö Sweden, Copenhagen Denmark, Exeter UK, Newcastle UK and Dundee UK with baseline and 18 months follow up data available n=861 participants (cohort 1 n=403, cohort 2 n= 458). Participants from study centres Kuopio Finland (cohort 1 n=1236) and Amsterdam the Netherlands (cohort 1 n=500; cohort 2 n=121) were excluded due to the required data was not yet available.

Ethical approval

All participants provided written informed consent and the study protocol was approved by the regional research ethics review boards. The research conformed to the ethical principles for medical research involving human participants outlined in the declaration of Helsinki. Clinical Trial Registration NCT03814915.

Data collection

Biochemistry assays

Fasting plasma glucose and insulin assays were analysed using the enzymatic glucose hexokinase method and photometric measurement on Konelab 20 XT Clinical Chemistry analyser (Thermo Fisher Scientific, Vantaa, Finland). Fasting HbA1c was measured by ion-exchange high-performance liquid chromatography using Tosoh G8 analysers (Tosoh Bioscience, San Francisco, CA, USA). Fasting blood lipids; cholesterol, triacylglycerol and HDL-cholesterol were measured using a Roche MODULAR P analyser (Roche Diagnostics, Indianapolis, IN, USA). LDL-cholesterol was calculated from the Friedewald formula. Each biochemical assay was performed using validated standard methods. Reference samples were included in all procedures to control for inter-assay variation and laboratories regularly participated in international external quality assessment schemes. Methodology is reported elsewhere (19).

Body composition

BMI was calculated as weight in kg divided by height in meters squared (kg/m^2), and waist circumference was measured at the level of the umbilicus at mid-respiration.

Magnetic resonance imaging

Whole body tissue composition was assessed using magnetic resonance imaging (MRI). Local protocols standardised across study centres by an experienced radiographer to harmonise the scan methodology. Multi-echo imaging sequencing was applied to identify liver fat. Methodology has been described in detail elsewhere (20).

Dietary data

Self-reported dietary intake was assessed by 24-hour multi-pass method and a food habit questionnaire, which was filled in by each participant the day before the study visit. The methods for the dietary record and the food habit questionnaire have been validated as part of the Euroaction Study (21). The method is structured into three levels of dietary questioning or 'passes'. The first pass aims to document a 'usual' day's meal. The second pass aims to give the respondent the time to reflect and add to the foods recorded in the first pass. The third pass of the food record aims to obtain information about portion size and method of preparation using a food portion size atlas. Nutritional analysis was undertaken using Dietplan-7 software (Forestfield Software Ltd, Horsham, UK) based on the McCance and Widdowson's 7th Edition Composition of Foods UK Nutritional Dataset. All diet coders were trained by a lead research dietician/nutritionist using a study specific operational manual protocol. Detailed description of the coding and diet analysis protocol are reported elsewhere (22).
Dietary patterns were assessed as concordance with WHO dietary guidelines using the validated 'Healthy Diet Indicator' (HDI) (9). The HDI priori score include assessment of six nutritional variables; fruits/vegetables, dietary fibre, total fats, saturated fat, sugars and salt. Mean daily intake of each of the seven components is assessed and scored according to their concordance with WHO dietary guidelines; 1 point represent +/- two

standard deviation of criteria, 0 point if intake was worse than WHO criteria and 2 points if intake was better than WHO criteria. The points are summed to calculate an overall score between 0-12 points; a higher score indicates a more favourable diet. The score was calculated from the dietary intakes of all food and drink consumed except alcohol, which was analysed separately and adjusted for with other known confounders.

Metabolic profiling

Fasting plasma blood samples were collected from participants at their baseline visit and processed using a targeted metabolomic assay Absolute*IDQ*™ p150 Kit (BIOCRATES Life Sciences, Innsbruck, Austria) quantifying 163 metabolites (amino acids, acylcarnitnes, sugars, glycerophospholipids, sphingolipids) (Supplementary table S5) (23). Samples were processed and quality controlled according to established protocols described in supplementary material, 116 metabolites passed quality control (Supplementary material). Plasma blood samples from a controlled clinical trial (17) (CCT) of participants undergoing four dietary interventions (with different levels of adherence to WHO guidelines for healthy eating) in random order were analysed using the same Absolute*IDQ*™ p180 Kit. This targeted metabolomic data was used to build a regression model (see below) to predict the stepwise adherence of participants to WHO guidelines. This model was constructed using the same methodology as exemplified previously by Garcia-Perez *et al*. for urine (16,000 ^{1}H NMR variables measured, 28 significant metabolites identified) (17) and applied here to fasting plasma (113 metabolites) to develop the predictive metabolomic score, T_{pred} for healthiness of diets using the 113 plasma metabolites. The 113 metabolites included in the model were selected based on metabolites found both from p180 Kit and from the 116 metabolites which passed the quality control in p150 Kit.

Statistical analysis

Baseline characteristics of participants were analysed across the two cohorts using a t-test.

The T_{pred} score was modelled with the 113 plasma metabolites of the CCT data with multivariate regression (partial least squares, PLS) and Monte Carlo cross-validation (MCCV) to assess model robustness (24). The MCCV model consisted of 1,000 iterations and the data were centred and scaled to account for the repeated-measures design of the CCT. The regression coefficients of each of the 1,000 MCCV iterations were used to calculate the predicted score (mean of all scores, T_{pred}). The samples of CCT individuals left out of each iteration of the training model, were used as test set for that iteration. This resulted in a goodness of prediction value of 0.94 for the CCT test data. The T_{pred} is indicative of how a metabolite profile relates to the plasma metabolite profiles of two dietary interventions with different levels of concordance with WHO healthy eating guidelines that were consumed in a highly controlled environment that assured full adherence to intervention diet (17). The T_{pred} (trained on plasma metabolomics data) ranged roughly from -3.5 to 3.5; a more positive T_{pred} indicate that the metabolite profiles have a greater resemblance to the diet with higher concordance with WHO healthy eating guidelines, whereas a negative T_{pred} is reflective of lower concordance with WHO guidelines. This model was then used to predict the DIRECT samples and obtain a T_{pred} for each sample. The T_{pred} score was log-transformed for analysis due to a right-skewed distribution (Supplementary figure S1).Associations between baseline data metabolite score T_{pred}, dietary pattern, single nutrients and food groups were analysed via multivariable generalised linear models. Models were adjusted for covariates including age, gender, BMI, cigarette smoking, alcohol consumption, energy intake, study centre and cohort.

Associations between baseline data T_{pred} score and glycaemic and cardiometabolic traits were analysed via multivariable generalised linear models. Models were adjusted for covariates including age, gender, smoking, alcohol consumption, energy intake and study centre; analyses where glycaemic traits outcomes were additionally adjusted for usage of glucose-lowering medication.

Relationships between baseline T_{pred} score and association with longitudinal changes in glycaemic traits and body composition were analysed at 18 months using generalised estimating equation regressions models (25). All models were adjusted for baseline age, gender, cigarette smoking, study centre and models with glycaemic traits (fasting glucose, insulin and HbA1c) were additionally adjusted for usage of glucose lowering medication. Sensitivity analyses were conducted using models adjusted for BMI. Additional longitudinal data analyses were conducted using mathematically modelled glycaemic progression rates for fasting plasma glucose and HbA1c concentrations from 36-month data adjusted for changes in BMI and usage of glucose-lowering medication (25). Each trajectory was described with a conditional linear mixed-effect model, in which the longitudinal component of the data was described as a proportional function of time, with normally distributed slopes describing individual progression rates. The slopes were additionally adjusted for age, gender and study centre. (Supplementary data).

Discordant and concordant analyses between T_{pred} and HDI quartiles were compared using a t-test. To assess discordant; upper HDI diet quartile (healthiest diet) and lower T_{pred} quartile (least favourable metabolic profile) and concordant; upper HDI diet quartile (healthiest diet) and upper T_{pred} quartile (most favourable metabolic profile).

Variable which did not follow a normal statistical distribution and was log transformed for the purpose of this study and coefficients were exponentiated. SAS version 9.4 (SAS Institute Inc. VX, Cary, NC, USA) was used for all analyses. The statistical significance threshold was set at $p<0.05$.

Results

Baseline characteristics of participants in the DIRECT cohorts

Table 2 shows descriptive characteristics for the two cohorts in DIRECT; Cohort 1 (participants with normal or impaired glucose regulation) and cohort 2 (participants diagnosed with T2D). Participants in cohort 2 were younger with a higher percentage of women than in cohort 1. Cohort 2 had higher BMI and a worse glycaemic profile compared to cohort 1 (unadjusted). No differences were observed across the two cohorts for smoking status, quality of diets or metabolite profile score T_{pred}.

T_{pred} metabolite profile score association with dietary patterns

Table 3 shows the linear regression coefficients, which represent mean changes in the nutritional variable per one-unit increase in T_{pred} score. A higher T_{pred} score (healthier metabolite profile) was associated with healthier dietary patterns based on WHO HDI score in cohort 2 (β=0.05, p= 0.0002) and a higher intake of fibre (β=0.03 g, p=0.004), fruit and vegetables (β=0.0007 g, p=0.002), and lower intake of added sugars (β=-0.02 g, p<.0001) in cohort 2. In both cohort 1 and 2, a higher T_{pred} score was associated with a higher intake of wholegrain (β=0.004 g, p=0.02 and β=0.003 g, p=0.03) and fish (β=0.004 g, p=0.0002 and β=0.004 g, p<.0001) and lower intake of energy (β=-0.0002 kcal, p=0.04 and β=-0.0002 kcal, p=0.003), and saturated fat (β=-0.03 g, p<.0001 and β=-0.03 g, p<.0001), respectively.

T_{pred} metabolite profile score association with glycaemic and cardiometabolic traits

Table 4 shows the mean changes in the phenotypic traits per one unit increase in T_{pred} score in the baseline data. A higher T_{pred} was associated with a lower weight in both cohorts β=-1.8 kg, p=0.01 and β=-2.1 kg, p=0.03, respectively for cohort 1 and 2. In cohort 2 a higher T_{pred} was also associated with a lower liver fat content (β=-0.74 %, p<.0001), and lower fasting HbA1c (β=-0.9 mmol/mol, p=0.02), glucose (β=-0.2 mmol/L, p=0.01) and insulin (β=-11.0 pmol/mol, p=0.01), lower TG (β=-0.8 mmol/L, $p=0$.0002) and higher HDL cholesterol (β=0.08 mmol/L, p=0.0002). In cohort 1, a higher T_{pred} was associated with a lower HbA1c (β=-0.4 mmol/mol, p=0.04), lower TG (β=-0.1 mmol/L, p=0.003) and a higher HDL cholesterol (β=0.07 mmol/L, p<.0001). No associations were observed for fasting LDL-cholesterol in either of the cohorts.

HDI dietary pattern score association with glycaemic and cardiometabolic traits

Supplementary table S2 shows the mean changes in phenotypic traits per one point increase in HDI diet score (indicating a healthier diet). Baseline data showed a higher HDI score was associated with lower weight in both cohorts. In cohort 2 a higher HDI was also associated with a lower waist circumference (β=-0.7 cm, p=0.002), a lower BMI (β=-0.3 kg/m^2, p=0.0002), liver fat (β=-0.7 %, p=0.0001). A higher HDI score was associated with a lower fasting glucose and insulin in both cohorts, in cohort 1 (β=-0.04 mmol/L, p=0.002) and cohort (β=-0.2 pmol/mol, p=0.04), respectively. In cohort 2 this was only significant for fasting glucose (β=-0.06 mmol/L, p=0.03) and insulin (β=-1.9 pmol/mol, p=0.09). No associations were observed between HDI score and lipid profile fasting HDL cholesterol, LDL-cholesterol and triglycerides in either cohort.

T_{pred} metabolite profile score effect on adiposity and glycaemic traits changes

Table 5 shows the generalised estimating equation regression coefficient representing the mean changes in phenotypic traits for one unit increase in T_{pred} score. A higher baseline T_{pred} score (healthier metabolite profile) was associated with decreasing body adiposity in both cohorts at 18 months follow up. In cohort 1 T_{pred} score was associated with a decrease in weight by -1.6 kg (p=0.02) and BMI -0.5 kg/m^2 (p=0.03). Cohort 2 also showed a reduction in waist circumference -1.6cm (p=0.04). A higher T_{pred} score was also associated with lower glycaemic traits in cohort 2; per one unit increase in T_{pred} score glucose was reduced by -0.2 mmol/L (p=0.03) and insulin was reduced by 9.2 pmol/mol (p=0.04). No other significant changes were observed for other glycaemic traits in cohort 1. Sensitivity analysis with glycaemic traits was done using same models with additional BMI adjustment (Supplementary table 3). The results showed per one point increase in the T_{pred} score glucose was reduced by 0.2 mmol/L (p=0.05) in cohort 2. No other significant changes were observed in either cohort.

The T_{pred} score showed no significant effect with the two glycaemic progression slopes modelled on fasting glucose (β=-0.007 mmol/L, p=0.6) or HbA1c (β=-0.07 mmol/mol, p=0.3) at 36-month follow up data in cohort 2 (Supplementary table 4).

Discordance analysis between T_{pred} metabolite profile score and HDI score

Discordance analyses of participants who were both in the lowest quartile of the metabolite score T_{pred} (worst metabolic profile) and in the highest diet score quartile (healthiest diets) were worse metabolically compared to those participants who were concordant with both scores i.e. highest T_{pred} score (healthiest metabolic profile) and highest diet score (healthiest diets) (supplementary material table 1). Participants concordant with both T_{pred} and HDI diet score had a lower BMI (mean=27.8 kg/m^2) compared to those who were discordant (mean= 30.3 kg/m^2) p=0.02, a lower mean fasting HbA1c (41.6 mmol/mol vs (42.9 mmol/mol) p=0.03, a lower mean fasting glucose (5.9 mmol/L vs 6.3 mmol/L) p<.0001, and a lower mean fasting insulin (42.6 pmol/mol vs 65.8 pmol/mol) p<.0001. Those concordant with the scores also had a better lipid profile: a lower mean fasting triglycerides (1.09 mmol/L vs 1.55 mmol/L) p<.0001, and a higher mean HDL-cholesterol (1.28 mmol/L vs 1.11 mmol/L) p=0.02. No significant differences were observed in body measurement or liver fat between the two groups.

Discussion

Dietary advice remains one of the cornerstone of prevention and management of T2D. The effectiveness of dietary intervention is related to global dietary profiling of numerous nutrients and food groups (increased fruit, vegetable, wholegrain and dietary fibre, decreased added sugars and total fat intake) rather than change in individual nutrients. However, understanding the relationship between diets and T2D is challenging because of the inaccuracy of traditional dietary assessment tools (15, 17, 26). We recently demonstrated the utility of urinary metabolite profiling in diet assessment (17). Here, we applied similar methodology to derive a plasma metabolite profiling score T_{pred} and demonstrated that it captures distinct dietary patterns in free-living populations. Furthermore, a more favourable T_{pred} score was associated with a better glycaemic homeostasis and body composition, liver fat, triglycerides and HDL-cholesterol. Our findings demonstrate that metabolic profiling validated self-reported diet recording and strengthened the precision of diet-disease relationships in epidemiological studies.

Firstly, we showed that the plasma metabolite profile score T_{pred} was associated with four distinct dietary patterns as defined by WHO dietary guidelines (HDI diet score), replicating findings from our previous RCT (17). Furthermore, a higher T_{pred} score (favourable metabolic profile) was associated with higher intake of healthy foods and nutrients including fibre, wholegrains, fish, fruit and vegetables and a lower intake of unhealthy nutrients including saturated fat, added sugars and lower energy intake despite the metabolites included in the T_{pred} score are not direct dietary metabolites. Similar findings were reported in the TwinsUK cohort which analysed same set of metabolites (AbsoluteIDQ^{TM} p150 Kit) separately and found correlations between two thirds of the metabolites and dietary patterns driven by food groups including meat intake, energy intake, fruit and vegetables (27).
We have demonstrated using discordant analyses that participants who both were in the lowest quartiles of T_{pred} (least favourable metabolic profile) and the highest diet score quartile (healthiest diet) are worse metabolically both in their glycaemic traits and lipid profiles compared to those concordant in both scores top quartiles (healthiest metabolite profile and healthiest diet). Few studies have conducted discordance analyses. We found two other cohorts (Cardiovascular Risk in Young Finns Study and Whitehall II) reporting similar findings when studying participants in concordance with higher adherence to a healthy diet and associated metabolites. The studies found participants had a better (more favourable) blood lipid profile and lower incidence of CVD (28), which was argued was driven by healthy dietary patterns.
Our findings may further imply that T_{pred} score can capture dietary misreporting known to confound the association between diet and health outcomes. This aligns with other nutritional studies, which show that participants "classified" as misreporters or energy underreporters in their diet recordings are more likely to have a worse cardio-metabolic profile compared to those who do not under-report their diet intake (11-13, 29). Metabolic profiling may serve as a more accurate and unbiased diet assessment method compared to existing methods relying mainly on BMI and physical activity reporting. Studies with high levels of underreporting/misreporting should adjust for such bias in their analyses or conduct sensitivity analyses to assess the impact on the relationship between diet and health outcomes (14, 30). In this study, we applied a statistical method using quartiles to identify the discordant and concordant groups in our dataset as both the T_{pred} and HDI score are arbitrary and no known or validated cut-off value exist. However, due to lack of published metabolomics discordance analysis in nutritional epidemiological studies such statistical principals have not yet been set for discordance as for bias of energy under-reporting.

A key finding of this study, the T_{pred} score was strongly associated with glycaemic and other cardiometabolic traits at baseline in both people at risk of diabetes and those living with the disease. Similar findings were observed in two other cohorts investigating the association between metabolic profiling with dietary patterns and CVD risk (28). They found 41 metabolites (amino acids, glycerophospholipids and sphingolipids) associated with healthy dietary patterns also predicted a favourable lipid (fatty acids) profile and lower

incidence of CVD risk. However, glycaemic and adiposity traits were not included in these studies. We also show that self-report diet data (HDI score) was associated with some of the same cardiometabolic traits as the T_{pred} score, but these relationships were less robust. This may imply a metabolic profile more accurately capture these associations between diet and health outcomes compared to self-reported dietary data, which are more prone to confounding from misreporting.

Mechanisms driving these nutritional and metabolic pathways are complex and multi-factorial. A metabolite profile score may capture metabolites, such as acylcarnitines, glycerophospholipids, and sphingolipids which are not captured in HDI score. Such metabolites have been associated with increased risks of fatty acid oxidation, insulin resistance and T2D (31, 32). Furthermore, the T_{pred} score may also capture other non-lifestyle factors such as interactions with an individual's microbiome (33, 34). Gut microbiota modified by diet may also play a role in the relationships between acylcarnitines, phosphatidylcholine and cardiometabolic traits (35, 36). Numerous metabolites such as acylcarnitines and phosphatidylcholine, are involved in gut microbe–dependent pathways that contribute to the formation of hepatic production of trimethylamine-N-oxide from choline and sequentially into trimethylamine in the microbiota, which may increase risk of atherosclerosis (37)and glucose metabolism (36, 38). A study that characterized faecal microbiome and serum metabolome of non-diabetic individuals found serum levels of branched-chain amino acids (BCAA) were higher in insulin-resistant individuals and also correlated with specific strains of faecal microbiomes with higher biosynthetic potential for BCAAs (34). These predictions were further validated in an animal study by showing that same microbiome strains could induce insulin resistance while increasing circulating BCAA levels when introduced into high fat diet *(34)*. Further investigations are needed to elucidate the interplay between specific dietary components and these metabolic and biological pathways.

The baseline T_{pred} score also showed favourable relationships with BMI and waist measurement in both cohorts at 18 months and significant lower fasting glucose and insulin in cohort 2 at 18 months. It is of interest that the Prevención con Dieta Mediterránea (PREDIMED) study showed that baseline metabolite profiles characterized by elevated concentrations of 28 acylcarnitines are independently associated with risks of CVD (serum lipid profile and adiposity markers) after 1 year and that these effects were lessened in participants consuming healthier diet patterns (Mediterranean diet compared to a control diet) (39). The T_{pred} metabolic profile score also included 18 of these acylcarnitines, which could suggest that the relationships observed in our study between T_{pred} score and serum lipid profile (HDL-cholesterol and triglycerides) and lower adiposity measurements were partially driving by the differences in acylcarnitines captured by T_{pred} score and other beneficial interaction effects of healthy dietary patterns. The Framingham study found association between 217 metabolites (amino acids, acylcarnitines, glycerophospholipids and sphingolipids) and body compositions, glycaemic and cardiometabolic traits (32). Their discoveries defined three distinct factors; dyslipidaemia, insulin resistance and adiposity at baseline. However, similar to our findings, their baseline metabolite profiles associations with longitudinal changes in these traits were distinct from the cross-sectional findings. They showed body compositions remained unchanged, whereas glycaemic traits varied across BMI strata, modifying metabolic signatures of insulin resistance. BMI was also associated with broad alteration in other multiple biochemical pathways (32).

Strengths and limitations of study

This study builds on previous metabolomics studies, but contributes a larger sample size, more detailed phenotyping and comprehensive dietary data analysis. Most epidemiological studies rely on dietary data from food frequency questionnaires, which may be limited by the specific nutrients and food groups studied. The IMI-DIRECT comprehensive dietary data analysis permitted us to study a range of single nutrients and food groups and overall dietary patterns and their relationships. This gives us a better understanding of what is driving a person's diet in relation to the metabolic impact and its association with important phenotypic traits. Furthermore, the data are derived from two well-designed and rigorously conducted cohort studies. These longitudinal data allow us to study the longer-term effects of the metabolite profile score on phenotypic traits in both participants with normal and impaired glucose regulation.
Lastly, the integration of metabolic and dietary profiling strengthens our findings as the metabolic profiling serves as an objective and unbiased approach, that does not rely on arbitrary cut-offs, which can strengthen accuracy in dietary data and nutritional epidemiology.

An important limitation is that our study design does not allow for determination of the temporality of the observed associations, which is important when considering the longitudinal dimension of T2D development and the complexity of nutrient and metabolomics. We performed metabolic profiling at one single time point, precluding the study of glycaemic and cardiometabolic effects on longitudinal changes in metabolite profile.

It also remains unclear whether metabolites represent biomarkers or actual mediators of metabolic disease or are acting as surrogate markers, thus, and causal inferences cannot be drawn from our observational study without further investigation. Another important limitation of this study is the use of a short-term dietary assessment method. A single 24-hour recall is unable to account for day-to-day variation, two or more non-consecutive recalls are required to estimate usual dietary intake distributions on an individual level. Multiple recalls are also recommended when used to examine diet and health. Though, the main focus of our study was to apply the metabolic model T_{pred} on a large free-living population. The T_{pred} was also derived from a single time point sample. Hence, to confirm these finding replication studies are needed using data collected over multiple time points to gain better accuracy and understanding on how the two methods compare in relation to each other and health outcomes.

This study demonstrates that application of dietary metabolic profiling at the population level provides an objective measurement of dietary patterns and is associated with glycaemic and cardiometabolic risk profiles.

Acknowledgement
We thank the participants across all IMI DIRECT Study centres for their contributions to the study. We also thank all the staff for their contribution to the planning, implementation or conduct of the study: http://www.direct-diabetes.org/

Funding
This work was supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115317 (DIRECT), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies in kind contribution. G.F. funded by Imperial College NIHR BRC and is an NIHR senior investigator. I.G.-P. is supported by a National Institute for Health Research (NIHR) Fellowship (NIHR-CDF-2017-10-032). J.M.P. is supported by a Rutherford Fund Fellowship at Health Data Research (HDR) UK (MR/S004033/1). RWK was funded by a STAR Award Novo Nordisk co-financed PhD fellowship. MMcC was a Wellcome Investigator (090532, 098381, 106130, 203141, 2122590 and an NIHR Senior Investigator (Niddk. U01-DK105535). This work was supported in part by ERC-2015-CoG_NASCENT_681742 and the Swedish Research Council; strategic funding for Lund University Diabetes Centre, where some of the work described herein was performed, was provided by the Swedish Research Council, Strategic Research Area Exodiab, (Dnr 2009-1039), the Swedish Foundation for Strategic Research (IRC15-0067), the Swedish Research Council, Linnaeus grant (Dnr 349-2006-237). EP holds a Wellcome Trust Investigator award (grant reference 102820/Z/13/Z). Contributions to this work by SBru. were co-financed by the Novo Nordisk Foundation (grants NNF17OC0027594 and NNF14CC0001).
The funders did not have any role in this study design, data collection, data analysis, interpretation or writing of the manuscript.

Declarations of interests
The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. MMcC has served on advisory panels for Pfizer, NovoNordisk and Zoe Global, has received honoraria from Merck, Pfizer, Novo Nordisk and Eli Lilly, and research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier, and Takeda. As of June 2019, MMcC is an employee of Genentech, and a holder of Roche stock.

Author contributions
GF, RE, IGP, PWF and JMP formulated the research questions and methodological design; RE, IGP and JMP were responsible for data analysis and drafting of the manuscript. GF, RE, IGP, EH, JA, JMP and PWF contributed to the interpretation of results and final manuscript. RE contributed to the development of the food diary coding protocol, coder training and audit checking. RE with the dietary data extraction and cleaning. EH, IGP and JMP were responsible for the validation of the metabolomic data extracts used in the analyses and constructing the T_{pred} score. MH, SS and JA were responsible for the metabolomics measurements and analysis. AV and JF were responsible for data analysis and quality checking of the metabolomic dataset. RB and AM modelled the mathematical progression slopes for fasting plasma glucose and HbA1c. Additionally, study design and coordination were contributed by RWK, JA, JBel, JBeu., SBru, GF, TH, AH, ML, AMar, TJMcD, OP, JMS, HJAT, AMah, MIMcC, HR, MW, EP, MH and IP. PWF, RWK, GNG, TW, JBel, JBeu, SBra, FDe M, IMF, GF, THH, TK, AK, AMar, TJMcD, FR, ELT, AV and AMah contributed to sample assaying, data analysis/processing and/or data quality control procedures. RWK, GNG, IMF, THH, TH, AH, TK, ML, AMar, TJMcD, OP, FR, JBeu, MW, EP and PWF contributed to quality control and data collection at study centres. All

authors contributed to drafting the article and/or revising it critically for important intellectual content. All authors approved the final version of the manuscript. All authors accept responsibility for all aspects of the work insofar as ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.

References

1. International Diabetes Federation: Diabetes atlas. 2018. Assess date: 15.07.2019. Available from: https://www.diabetesatlas.org/.
2. World Health Organisation. Diet, nutrition and the prevention of chronic diseases. Technical report series; 2003. Contract No.: ISSN 0512-3054.
3. Committee on Medical Aspects of Food Policy. Nutritional Aspects of Cardiovascular Disease. 1994(No.46).
4. Esposito K KC, Panagiotakos DB, Giugliano D. Mediterranean diet and weight loss: meta-analysis of randomized controlled trials. 2011;9(1).
5. Kant AK. Dietary patterns and health outcomes. Journal of the American Dietetic Association. 2004;104(4):615-35.
6. Waijers PM, Feskens EJ, Ocké MC. A critical review of predefined diet quality scores. British journal of nutrition. 2007;97(02):219-31.
7. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Current opinion in lipidology. 2002;13(1):3-9.
8. Levy L, Tedstone A, editors. UK Dietary Policy for the Prevention of Cardiovascular Disease. Healthcare; 2017: Multidisciplinary Digital Publishing Institute.
9. Jankovic N, Geelen A, Streppel MT, de Groot LC, Orfanos P, van den Hooven EH, et al. Adherence to a healthy diet according to the World Health Organization guidelines and all-cause mortality in elderly adults from Europe and the United States. American journal of epidemiology. 2014;180(10):978-88.
10. Reidlinger DP, Darzi J, Hall WL, Seed PT, Chowienczyk PJ, Sanders TA. How effective are current dietary guidelines for cardiovascular disease prevention in healthy middle-aged and older men and women? A randomized controlled trial. The American journal of clinical nutrition. 2015;101(5):922-30.
11. Rennie KL, Coward A, Jebb SA. Estimating under-reporting of energy intake in dietary surveys using an individualised method. British Journal of Nutrition. 2007;97(6):1169-76.
12. Stubbs, J. Detecting and Modelling Mis-reporting of Food Intake with Special Reference to Under-reporting in the Obese. 2003.
13. Coulston AM, Boushey C. Nutrition in the Prevention and Treatment of Disease: Academic Press; 2008.
14. Garriguet D. Impact of identifying plausible respondents on the under-reporting of energy intake in the Canadian Community Health Survey. Health Reports. 2008;19(4):47.
15. Playdon MC, Moore SC, Derkach A, Reedy J, Subar AF, Sampson JN, et al. Identifying biomarkers of dietary patterns by using metabolomics. The American journal of clinical nutrition. 2016;105(2):450-65.
16. Hanhineva K, Lankinen MA, Pedret A, Schwab U, Kolehmainen M, Paananen J, et al. Nontargeted metabolite profiling discriminates diet-specific biomarkers for consumption of whole grains, fatty fish, and bilberries in a randomized controlled trial. The Journal of nutrition. 2014;145(1):7-17.
17. Garcia-Perez I, Posma JM, Gibson R, Chambers ES, Hansen TH, Vestergaard H, et al. Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. The Lancet Diabetes & Endocrinology. 2017;5(3):184-95.
18. Koivula RW, Heggie A, Barnett A, Cederberg H, Hansen TH, Koopman AD, et al. Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: rationale and design of the epidemiological studies within the IMI DIRECT Consortium. Diabetologia. 2014;57(6):1132-42.
19. Koivula RW, Forgie IM, Kurbasic A, Viñuela A, Heggie A, Giordano GN, et al. Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: descriptive characteristics of the epidemiological studies within the IMI DIRECT Consortium. Diabetologia. 2019:1-15.
20. Thomas EL, Fitzpatrick J, Malik S, Taylor-Robinson SD, Bell JD. Whole body fat: content and distribution. Progress in nuclear magnetic resonance spectroscopy. 2013;73:56-80.
21. Wood D, Kotseva K, Connolly S, Jennings C, Mead A, Jones J, et al. Nurse-coordinated multidisciplinary, family-based cardiovascular disease prevention programme (EUROACTION) for patients with coronary heart disease and asymptomatic individuals at high risk of cardiovascular disease: a paired, cluster-randomised controlled trial. The Lancet. 2008;371(9629):1999-2012.

22. Gibson R, Frost G, Elliott P, Chan Q, Eriksen R, Vergnaud A, et al. Dietary assessment of British police force employees: a description of diet record coding procedures and cross-sectional evaluation of dietary energy intake reporting (the airwave health monitoring study). BMJ.7(4).
23. Biocrates. Biocrates life Sciences AG . The Standard in Targeted Metabolomics. AbsoluteIDQ p150 Kit. No. 35 025, V02-2019. Access date: 19.05.2019. https://www.biocrates.com/images/p150_KitFolder.pdf.
24. Posma JM, Garcia-Perez I, Ebbels TM, Lindon JC, Stamler J, Elliott P, et al. Optimized phenotypic biomarker discovery and confounder elimination via covariate-adjusted projection to latent structures from metabolic spectroscopy data. Journal of proteome research. 2018;17(4):1586-95.
25. Bizzotto R, Jennison C, Jones A, Walker M, Pearson E, Mari A, editors. Glucose sensitivity, insulin sensitivity and their longitudinal changes are strong independent determinants of type 2 diabetes progression: an IMI DIRECT study. Diabetologia; 2019: SPRINGER 233 SPRING ST, NEW YORK, NY 10013 USA.
26. Rebholz CM, Lichtenstein AH, Zheng Z, Appel LJ, Coresh J. Serum untargeted metabolomic profile of the Dietary Approaches to Stop Hypertension (DASH) dietary pattern. The American journal of clinical nutrition. 2018;108(2):243-55.
27. Menni C, Zhai G, MacGregor A, Prehn C, Römisch-Margl W, Suhre K, et al. Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. Metabolomics. 2013;9(2):506-14.
28. Akbaraly T, Würtz P, Singh-Manoux A, Shipley MJ, Haapakoski R, Lehto M, et al. Association of circulating metabolites with healthy diet and risk of cardiovascular disease: analysis of two cohort studies. Scientific reports. 2018;8(1):8620.
29. Suissa K, Benedetti A, Henderson M, Gray-Donald K, Paradis G. The Cardiometabolic Risk Profile of Underreporters of Energy Intake Differs from That of Adequate Reporters among Children at Risk of Obesity. The Journal of nutrition. 2019;149(1):123-30.
30. Gottschald M, Knüppel S, Boeing H, Buijsse B. The influence of adjustment for energy misreporting on relations of cake and cookie intake with cardiometabolic disease risk factors. European journal of clinical nutrition. 2016;70(11):1318.
31. Schooneman MG, Vaz FM, Houten SM, Soeters MR. Acylcarnitines: reflecting or inflicting insulin resistance? Diabetes. 2013;62(1):1-8.
32. Ho JE, Larson MG, Ghorbani A, Cheng S, Chen M-H, Keyes M, et al. Metabolomic profiles of body mass index in the Framingham Heart Study reveal distinct cardiometabolic phenotypes. PLoS One. 2016;11(2):e0148361.
33. Shaffer M, Armstrong AJ, Phelan VV, Reisdorph N, Lozupone CA. Microbiome and metabolome data integration provides insight into health and disease. Translational Research. 2017;189:51-64.
34. Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyotylainen T, Nielsen T, Jensen BA, et al. Human gut microbes impact host serum metabolome and insulin sensitivity. Nature. 2016;535(7612):376.
35. Tang WW, Hazen SL. The contributory role of gut microbiota in cardiovascular disease. The Journal of clinical investigation. 2014;124(10):4204-11.
36. Hoyles L, Jiménez-Pranteda ML, Chilloux J, Brial F, Myridakis A, Aranias T, et al. Metabolic retroconversion of trimethylamine N-oxide and the gut microbiota. Microbiome. 2018;6(1):73.
37. Gessner A, di Giuseppe R, Koch M, Fromm MF, Lieb W, Maas R. Trimethylamine-N-oxide (TMAO) determined by LC-MS/MS: distribution and correlates in the population-based PopGen cohort. Clinical Chemistry and Laboratory Medicine (CCLM). 2020.
38. Ussher JR, Lopaschuk GD, Arduini A. Gut microbiota metabolism of L-carnitine and cardiovascular risk. Atherosclerosis. 2013;231(2):456-61.
39. Guasch-Ferré M, Zheng Y, Ruiz-Canela M, Hruby A, Martínez-González MA, Clish CB, et al. Plasma acylcarnitines and risk of cardiovascular disease: effect of Mediterranean diet interventions. The American journal of clinical nutrition. 2016;103(6):1408-16.

Table 1: Study inclusion and exclusion criteria for the IMI DIRECT cohorts

Cohort 1	Cohort 2
Inclusion criteria	**Inclusion criteria**
• No treatment with insulin-sensitising, glucose-lowering or other antidiabetic drugs	• Patients diagnosed with T2D not <6 months and not >24 months before baseline
• Fasting capillary blood glucose <10 mmol/l at baseline	• Management by lifestyle with or without metformin therapy
• White European	• White European
• Age ≥35 and <75 years	• Age ≥35 and <75
	• Estimated GFR >50 ml/min
	• All HbA1c <60 mmol/mol within previous 3 months
Exclusion criteria	**Exclusion criteria**
• Diagnosed diabetes of any type, HbA1c ≥48 mmol/mol or fasting plasma glucose ≥7.0 mmol/l or 2 h plasma glucose	• Type 1 diabetes
>11.0 mmol/l previously	• A previous HbA1c >75 mmol/mol
• For women, pregnancy, lactation or plans to conceive within the study period	• Prior treatment with insulin or an oral hypoglycaemic agent other than metformin
• Use of a pacemaker	• BMI <20 or >50 kg/m2
• Any other significant medical reason for exclusion as determined by the investigator	• For women, pregnancy, lactation or plans to conceive within the study period • Any other significant medical reason for exclusion as determined by the investigator

Table 2: Baseline characteristics of participants in, the IMI DIRECT cohorts

	Cohort I (n=403)		Cohort II (n=458)		
	Mean or n	**SD or %**	**Mean or n**	**SD or %**	***P* value [b]**
Sex (female) %	234	59.1	183	41.8	<.0001
Age (years)	64	8	62	7.9	<.0001
Diet quality					
T_{pred} metabolic score [range -3.5, 3.5]	-0.6	0.9	-0.6	0.7	0.80
HDI diet score [range 0, 12]	4.6	2.7	4.8	2.6	0.48
Daily energy (kcal) intake	1753.5	610.7	1802.5	618.9	0.18
Alcohol %					0.09
No alcohol	315	69	380	73	
Within UK guidelines	76	17	64	12	
Above UK guidelines	65	14	77	15	
Cigarette smoking %					
Never	149	37	190	43	0.21
Former	192	48	197	45	
Current	57	14	51	12	
Body mass index (kg/m^2)	28.9	4.2	30.8	5.1	<.0001
Waist circumference (cm)	102.1	11	102.6	13.8	0.58
Liver fat (%)	5.2	4.6	8.6	7.8	<.0001
Glycaemic & cardio-metabolic traits					

Glucose (mmol/L)	5.6	0.9	6.9	1.4	<.0001
Insulin (pmol/mol)	11.9	16.3	103.9	70.4	<.0001
HbA1c (mmol/mol)	38.1	3.2	46.5	5.9	<.0001
Triglycerides (mmol/L)	1.3	0.6	1.4	0.8	<.0001
LDL-cholesterol (mmol/L)	3.4	0.9	2.3	0.9	<.0001
HDL-cholesterol (mmol/L)	1.2	0.3	1.2	0.4	0.09

Abbreviations: Cohort 1; participants with normal or impaired glucose regulation, cohort 2; participants with diabetes type 2, T_{pred}, metabolic profile score; HDI, Healthy Diet Indicator WHO diet score

[a] Values are unadjusted means (standard deviation) or n (%)

[b] *P* value linear model (continuous variable) Mantel-Haenzel chi-square test (categorical variables)

Table 3: Metabolic profile score T_{pred} association with mean dietary intake adjusted for age, gender, smoking, alcohol consumption, BMI and study centre, the IMI DIRECT cohorts

	Cohort I n=403			Cohort II n=458		
	β [a]	95% CI	*P* value	β [a]	95% CI	*P* value
Fibre (NSP) per 1000 kcal	0.02	-0.004, 0.05	0.09	0.03	0.001, 0.05	0.004
Fruit/vegetables per 1000 kcal	0.0004	0.0001, 0.0009	0.23	0.0007	0.0003, 0.001	0.002
Wholegrains per 1000 kcal	0.004	0.0005, 0.007	0.02	0.003	0.003, 0.006	0.03
Fish per 1000 kcal	0.004	0.002, 0.005	0.0002	0.004	0.002, 0.005	<.0001
Carbohydrate %TEI	-0.005	-0.01, 0.004	0.34	-0.007	-0.01, -0.0004	0.04
Protein %TEI	0.02	0.001, 0.03	0.04	0.03	0.02, 0.04	<.0001
Fat %TEI	-0.0006	-0.01, 0.008	0.86	-0.006	-0.01, 0.002	0.23
Saturated fat %TEI	-0.03	-0.05, -0.02	<.0001	-0.03	-0.05, -0.02	<.0001
Added sugar %TEI	-0.01	-0.02, -0.006	0.31	-0.02	-0.03, -0.01	<.0001
Mean kcal	-0.0002	-0.0003, -0.0001	0.04	-0.0002	-0.0003, -0.0001	0.003
HDI score	0.03	-0.004, 0.06	0.09	0.05	0.02, 0.07	0.0002

Abbreviation: Cohort 1; participants with normal or impaired glucose regulation, cohort 2; participants with diabetes type 2, NSP, non-starch polysaccharides, %TEI, percentage of total energy intake, 95% CI, confidential interval, HDI, Healthy Diet Indicator. a, Generalised linear regression model coefficient represents the mean change in the nutritional variable for one unit change (increase) in T_{pred} score fully adjusted models.

Table 4: Metabolic profile score T_{pred} association with phenotypic traits adjusted for age, gender, smoking, alcohol consumption and study centre at baseline, the IMI DIRECT cohorts

	Cohort I n=403			Cohort II n=458		
	β [a]	95% CI	*P* value	β [a]	95% CI	*P* value
Waist circumference (cm)	-0.9	-2.1, 0.3	0.09	-1.5	-2.9, 0.08	0.06
Weight (kg)	-1.8	-3.2, -0.4	0.01	-2.1	-3.8, -0.2	0.03
Body mass index (kg/m2)	-0.5	-1.0, -0.1	0.02	-0.5	-1.1, 0.1	0.09
Liver fat (%)	-0.97	-0.82, 1.11	0.63	-0.74	-0.67, 0.81	<.0001
Fasting HbA1c (mmol/mol)†	-0.4	-0.7, -0.01	0.04	-0.9	-1.5, -0.1	0.02
Fasting glucose (mmol/L)†	0.04	-0.04, 0.1	0.32	-0.2	-0.4, -0.05	0.01
Fasting insulin (pmol/mol)†	0.2	-0.7, 0.7	0.88	-11.0	-19.5, -2.6	0.01
Fasting triglycerides (mmol/L)	-0.1	-0.2, -0.03	0.003	-0.2	-0.3, -0.09	0.0002
Fasting LDL-cholesterol (mmol/L)	0.001	-0.1, 0.1	0.86	-0.06	-0.2, 0.04	0.24
Fasting HDL-cholesterol (mmol/L)	0.07	0.03, 0.1	<.0001	0.08	0.04, 0.1	0.0002

Abbreviations: Cohort I; participants with normal or impaired glucose regulation, Cohort II; participants with type 2 diabetes, HDL; high density lipoprotein cholesterol, LDL; low density lipoprotein cholesterol, HbA1c; glycated haemoglobin. a: Generalised linear regression model coefficient represents the mean change in the phenotypic trait for one unit change (increase) in T_{pred} score.
† Cohort II is additionally adjusted for usage of glucose lowering medication

Table 5: Metabolic profile score T_{pred} effect on phenotypic traits changes adjusted for gender, age, smoking, alcohol consumption and study centre at 18 months follow up, the IMI DIRECT cohorts

	Cohort I n=403			Cohort II n=458		
	β [a]	95% CI	*P* value	β [a]	95% CI	*P* value
Fasting insulin (pmol/mol) †	-0.2	-0.9, 0.5	0.61	-9.2	-17.9, -0.4	0.04
Fasting glucose (mmol/L) †	0.03	-0.03, 0.1	0.28	-0.2	-0.3, -0.01	0.03
Fasting HbA1c (mmol/mol) †	-0.4	-0.7, 0.01	0.06	-0.6	-1.5, 0.1	0.09
Body mass index (kg/m^2)	-0.5	-0.9, -0.06	0.03	-0.7	-1.6, 0.1	0.09
Waist circumference (cm)	-1.1	-2.2, 0.03	0.05	-1.6	-3.0, -0.1	0.04
Weight (kg)	-1.6	-3.05, -0.2	0.02	-1.7	-3.5, -0.1	0.06

Abbreviation: cohort I participants with normal or impaired glucose regulation, cohort II participants with type 2 diabetes. a, Generalised estimating equation regression coefficient represent the mean change in phenotypic trait for one unit change (increase) in T_{pred} score adjusted for covariates. † Cohort II is additionally adjusted for usage of glucose lowering medication

Supplementary material

Table S1: Discordance analysis of metabolic profile score T_{pred} and Healthy Diet Indicator score (t-test) unadjusted, the IMI DIRECT cohorts

	Discordant [a] n=291		Concordant [b] n=70		
	Mean†	**95% CI**	**Mean†**	**95% CI**	***P*-value**
Weight (kg)	88.2	86.4, 89.9	79.2	75.9, 82.5	<.0001
Body mass index (kg/m^2)	30.3	29.8, 30.9	27.8	26.8, 28.7	<.0001
Liver fat (%)	16.8	15.4, 18.2	11.5	9.3, 13.7	0.08
Fasting HbA1c (mmol/mol)	42.3	41.6, 43.1	41.6	40.3, 42.8	0.32
Fasting glucose (mmol/L)	6.2	6.0, 6.4	6.0	5.7, 6.3	0.27
Fasting insulin (pmol/mol)	58.9	49.9, 67.9	42.1	33.1, 50.9	0.009
Fasting triglycerides (mmol/L)	1.5	1.4, 1.6	1.1	0.9, 1.2	<.0001
Fasting HDL-cholesterol (mmol/L)	1.1	1.1, 1.2	1.3	1.2, 1.4	0.003
Fasting LDL-cholesterol (mmol/L)	2.9	2.8, 3.1	2.7	2.4, 2.9	0.05

a, discordant; highest HDI diet score quartile (healthiest diet) and lowest metabolite score T_{pred} quartile (least favourable metabolic profile) b, concordant; highest HDI diet score quartile (healthiest diet) and highest metabolite score T_{pred} quartile (most favourable metabolic profile). †Means and 95%CI are unadjusted

Table S2: Healthy Diet Indicator score association with phenotypic traits adjusted for age, gender, smoking, alcohol consumption, BMI and study centre at baseline, the IMI DIRECT cohorts

	Cohort I n=403			Cohort II n=485		
	β [a]	**95% CI**	***P* value**	**β [a]**	**95% CI**	***P* value**
Waist circumference (cm) †	-0.4	-0.8, 0.01	0.05	-0.7	-1.2, -0.3	0.002
Weight (kg) †	-0.5	-1.01, -0.07	0.02	-1.0	-1.5, -0.5	0.0002
Body mass index (kg/m^2) †	-0.2	-0.3, 0.004	0.06	-0.3	-0.5, -0.2	0.0002
Liver fat (%)	0.06	-0.4, 0.5	0.83	-0.7	-1.0, -0.3	0.0001
Fasting HbA1c (mmol/mol) ††	0.05	-0.07, 0.2	0.41	-0.1	-0.3, 0.1	0.33
Fasting glucose (mmol/L) ††	-0.04	-0.07, -0.02	0.002	-0.06	-0.1, -0.01	0.03
Fasting insulin (pmol/mol) ††	-0.2	-0.4, -0.003	0.04	-1.9	-4.2, 0.3	0.09
Fasting triglycerides (mmol/L)	-0.01	-0.03, 0.01	0.29	-0.01	-0.04, 0.01	0.26
Fasting LDL-cholesterol (mmol/L)	-0.004	-0.04, 0.03	0.77	0.01	-0.02, 0.04	0.53
Fasting HDL-cholesterol (mmol/L)	0.00004	-0.01, 0.01	0.92	-0.005	-0.02, 0.01	0.48

Abbreviations: Cohort 1; participants with normal or impaired glucose regulation, Cohort 2; participants with type 2 diabetes, HDL; high density lipoprotein cholesterol, LDL; low density lipoprotein cholesterol, HbA1c; glycated haemoglobin. a, Generalised linear regression model coefficient represents the mean change in the phenotypic trait for one unit change (increase) in the Healthy Diet Indicator score. † variables not adjusted for body mass index

†† Cohort 2 additionally adjusted for usage of glucose lowering medication

Table S3 Metabolic profile score T_{pred} effect on phenotypic traits changes adjusted for gender, age,

smoking, alcohol consumption, BMI and study centre at 18 months follow up, the IMI DIRECT cohorts

	Cohort I (n=403)			Cohort II (n=458)		
	β [a]	95% CI	*P* value	β [a]	95% CI	*P* value
Fasting insulin (pmol/mol)†	0.2	-0.4, 0.8	0.48	-6.3	-14, 1.5	0.11
Fasting glucose (mmol/L)†	0.05	-0.02, 0.1	0.21	-0.2	-0.3, 0.003	0.05
Fasting HbA1c (mmol/mol)†	-0.3	-0.7, 0.04	0.08	-0.6	-1.4, 0.2	0.09

Abbreviation: cohort I participants with normal or impaired glucose regulation, cohort II participants with type 2 diabetes. a, Generalised estimating equation regressions coefficient represent effect per one unit increase T_{pred} on dependent variable adjusted for covariates. † Cohort II is additionally adjusted for usage of glucose lowering medication

Table S4 Baseline metabolic profile score T_{pred} effect on glycaemic traits changes at 36 months follow up adjusted for gender, age, BMI and study centre, the IMI DIRECT cohort II

	Cohort II (n=458)		
	β [a]	95% CI	*P* value
Fasting glucose (mmol/L)	-0.01	-0.04, 0.02	0.62
Fasting HbA1c (mmol/mol)	-0.07	-0.21, 0.07	0.28

Abbreviations: Cohort 2; participants with type 2 diabetes, HbA1c; glycated haemoglobin. a: Generalised linear regression model coefficient represents the mean change in the phenotypic trait for one unit change (increase) in T_{pred} score using mathematically modelled glycaemic progression slopes for fasting plasma glucose and HbA1c concentrations from 36-month data adjusted for changes in BMI, age, gender, study centre and usage of glucose-lowering medication.

Table S5 Targeted metabolites groups included in the metabolic profile score T_{pred}, the IMI DIRECT cohorts

Metabolite class	Biological relevance (selected examples)
Amino Acids	Amino acid metabolism, urea cycle, activity of gluconeogenesis and glycolysis, insulin sensitivity/resistance, neurotransmitter metabolism, oxidative stress
Acylcarnitines	Energy metabolism, fatty acid transport and mitochondrial fatty acid oxidation (e.g., ketosis, oxidative stress, mitochondrial membrane damage (apoptosis)
Hexoses	Carbohydrate metabolism
Phosphatidylcholines	Dyslipidaemia, membrane composition and damage, fatty acid profile, activity of desaturases
Lysophosphatidylcholines	Degradation of phospholipids (phospholipase activity), membrane damage, signalling cascades, fatty acid profile
Sphingomyelins	Signalling cascades, membrane damage (e.g., neurodegeneration)

Targeted metabolomics Biocrates Absolute*IDQ*™ p150 Kit quality control

***Sapna Sharma*[1,2], *Cornelia Prehn*[5], *Mark Haid*[5], *Harald Grallert*[1,2,3,4], *Jerzy Adamski*[5,6,7] *Ana Viñuela*[8], *Juan Fernandez*[9]**

[1] German Center for Diabetes Research (DZD), Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

[2] Research Unit of Molecular Epidemiology, Institute of Epidemiology II, Helmholtz Zentrum München, Germany

[3] Clinical Cooperation Group Type 2 Diabetes. Helmholtz Zentrum München and Ludwig-Maximilians Universität München, Germany

[4] Clinical Cooperation Group Nutrigenomics and Type 2 Diabetes, Helmholtz Zentrum München and Technische Universität München, Germany

[5] Research Unit Molecular Endocrinology and Metabolism, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[6] Chair for Experimental Genetics, Technical University of Munich, Freising-Weihenstephan, Germany

[7] Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597, Singapore.

[8] Department of Genetic Medicine and Development, University of Geneva Medical School, CH-1211 Geneva, Switzerland.

[9] Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

Biocrates Absolute*IDQ*™ p150

Plasma concentrations of 163 metabolites were determined using a FIA-ESI-MS/MS-based targeted metabolomics approach with the Absolute*IDQ*™ p150 kit (BIOCRATES Life Sciences AG, Innsbruck, Austria). Analytical specifications for LOD and evaluated quantification ranges, further LOD for semi-quantitative measurements, identities of quantitative and semiquantitative metabolites, specificity, potential interferences, linearity, precision and accuracy, reproducibility, and stability were described in Biocrates manual AS-P150.

The assay allows simultaneous quantification of 163 metabolites out of 10 µL plasma, and includes free carnitine, 40 acylcarnitines (Cx:y), 15 amino acids (Leu and Ile are measured together as xLeu), hexoses (sum of hexoses – about 90-95 % glucose), 91 glycerophospholipids (15 lysophosphatidylcholines (lysoPC.Cx:y) and 76 phosphatidylcholines (PC.Cx:y)), and 15 sphingolipids (SM.Cx:yc). The abbreviations Cx:y are used to describe the total number of carbons and double bonds of all chains, respectively.

The LODs were set to three times the values of the zero samples (PBS). The LLOQ and ULOQ were determined experimentally by Biocrates. The assay procedures of the Absolute*IDQ*™ p150 Kit as well as the metabolite nomenclature have been described in detail previously (1,2).

Analytical specifications for LOD and evaluated quantification ranges, further LOD for semi-quantitative measurements, identities of quantitative and semi-quantitative metabolites, specificity, potential interferences, linearity, precision and accuracy, reproducibility, and stability were described in Biocrates manual AS-P150.

Sample handling was performed by a Hamilton Microlab STAR™ robot (Hamilton Bonaduz AG, Bonaduz, Switzerland) and a Ultravap nitrogen evaporator (Porvair Sciences, Leatherhead, U.K.), beside standard laboratory equipment. Mass spectrometric analyses were done on an API 4000 triple quadrupole system (Sciex Deutschland GmbH, Darmstadt, Germany) equipped with a 1200 Series HPLC (Agilent Technologies Deutschland GmbH, Böblingen, Germany) and a HTC PAL auto sampler (CTC Analytics, Zwingen, Switzerland) controlled by the software Analyst 1.6.2. Data evaluation for quantification of metabolite concentrations and quality assessment was performed with the software MultiQuant 3.0.1 (Sciex) and the Met*IDQ*™ software package, which is an integral part of the Absolute*IDQ*™ Kit. Metabolite concentrations were calculated using internal standards and reported in µM.

In addition to the investigated samples, five aliquots of a pooled reference plasma were analyzed on each kit plate. These reference plasma samples were used for normalization purposes and for calculation of coefficient of variance (CV) for each metabolite.

Targeted metabolomics – quality control

After data export from Met*IDQ*™, a first technical QC comprising analysis of peak shapes, retention times, and compound identity was performed. In subsequence QC steps, we join metabolites measurements from both cohorts and evaluated the number of samples with zero values in the metabolites concentration and remove any individual with more than 50% of zeros. No sample was removed in this step. We then evaluate the CV per metabolite and removed 33 with CV > 0.25 relative to the reference samples. Metabolites with concentration below the LOD were discarded, removing an additional 14 metabolites. Of the 163 metabolites, 116 passed all quality controls in 3,029 individuals.

Reference

1. Römisch-Margl W, Prehn C, Bogumil R, Röhring C, Suhre K, Adamski J. Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. Metabolomics. 2012; 8: 133-142; doi:10.1007/s11306-011-0293-4.

3. Zukunft S, Sorgenfrei M, Prehn C, Möller G, Adamski J. (2013) Targeted Metabolomics of Dried Blood Spot Extracts. Chromatographia, DOI 10.1007/s10337-013-2429-3

Figure S1: Metabolic profile score T_{pred} distribution, the IMI DIRECT cohorts