# Serological Phenotyping Analysis Uncovers a Unique Metabolomic Pattern Associated With Early Onset of Type 2 Diabetes Mellitus

Linmin Zhu[1,2†], Qianyang Huang[3†], Xiao Li[4,5†], Bo Jin[4†], Yun Ding[3], C. James Chou[3], Kuo-Jung Su[3], Yani Zhang[4], Xingguo Chen[2], Kuo Yuan Hwa[3], Sheeno Thyparambil[3], Weili Liao[3], Zhi Han[3], Richard Mortensen[3], Yi Jin[4], Zhen Li[6], James Schilling[3,5], Zhen Li[4,5], Karl G. Sylvester[7], Xuguo Sun[1]* and Xuefeng B. Ling[7]*

[1]School of Laboratory Medicine, Tianjin Medical University, Tianjin, China, [2]Tianjin Teda Hospital, Tianjin, China, [3]mProbe Inc, Mountain View, CA, United States, [4]Tianjin Yunjian Medical Laboratory Institute Co., Ltd, Tianjin, China, [5]Binhai Industrial Technology Research Institute, Zhejiang University, Tianjin, China, [6]Shanghai Yunxiang Medical Technology Co., Ltd., Shanghai, China, [7]Department of Surgery, Stanford University, School of Medicine, Stanford, CA, United States

**Background:** Type 2 diabetes mellitus (T2DM) is a multifaceted disorder affecting epidemic proportion at global scope. Defective insulin secretion by pancreatic $\beta$-cells and the inability of insulin-sensitive tissues to respond effectively to insulin are the underlying biology of T2DM. However, circulating biomarkers indicative of early diabetic onset at the asymptomatic stage have not been well described. We hypothesized that global and targeted mass spectrometry (MS) based metabolomic discovery can identify novel serological metabolic biomarkers specifically associated with T2DM. We further hypothesized that these markers can have a unique pattern associated with latent or early asymptomatic stage, promising an effective liquid biopsy approach for population T2DM risk stratification and screening.

**Methods:** Four independent cohorts were assembled for the study. The T2DM cohort included sera from 25 patients with T2DM and 25 healthy individuals for the biomarker discovery and sera from 15 patients with T2DM and 15 healthy controls for the testing. The Pre-T2DM cohort included sera from 76 with prediabetes and 62 healthy controls for the model training and sera from 35 patients with prediabetes and 27 healthy controls for the model testing. Both global and targeted (amino acid, acylcarnitine, and fatty acid) approaches were used to deep phenotype the serological metabolome by high performance liquid chromatography-high resolution mass spectrometry. Different machine learning approaches (Random Forest, XGBoost, and ElasticNet) were applied to model the unique T2DM/Pre-T2DM metabolic patterns and contrasted with their effectiveness to differentiate T2DM/Pre-T2DM from controls.

**Results:** The univariate analysis identified unique panel of metabolites ($n = 22$) significantly associated with T2DM. Global metabolomics and subsequent structure determination led to the identification of 8 T2DM biomarkers while targeted LCMS profiling discovered 14 T2DM biomarkers. Our panel can effectively differentiate T2DM (ROC AUC = 1.00) or Pre-T2DM (ROC AUC = 0.84) from the controls in the respective testing cohort.

**Conclusion:** Our serological metabolite panel can be utilized to identifiy asymptomatic population at risk of T2DM, which may provide utility in identifying population at risk at an early stage of diabetic development to allow for clinical intervention. This early detection would guide ehanced levels of care and accelerate development of clinical strategies to prevent T2DM.

Keywords: metabolomics, type 2 diabetes mellitus, biomarker, early detection, serum

# 1 INTRODUCTION

Type 2 diabetes mellitus (T2DM) is a metabolic disease affecting a significant population worldwide. The prevalence of T2DM is expected to expand into half a billion people by the year 2030 (Whiting et al., 2011). It is manifested as chronic hyperglycemia caused by defective insulin secretion, progressive development of insulin resistance, and an inadequate compensatory insulin secretory response (Stumvoll et al., 2005; Whiting et al., 2011). In the presence of homeostatic imbalance of glucose, the overload of advanced glycation end products induces enhanced oxidative stress and systemic inflammation, which irreversibly causes functional loss of pancreatic islet $\beta$ cells (DeFronzo, 1988; Rossetti et al., 1990; Robertson et al., 2003), peripheral insulin-targeting tissues (Garvey et al., 1987; Rossetti et al., 1987), micro- and macro-vascular cells (Brownlee, 2001; Brownlee, 2005), leading to pathological complications such as cardiomyopathy, nephropathy, retinopathy, neuropathy, and atherosclerosis.

The early detection of T2DM remains a significant challenge in clinics today. As the onset of T2DM is asymptomatic for a large portion of the high-risk population, there is usually a latent phase before the confirmative diagnosis of T2DM during which risk factors for diabetic micro- and macro-vascular complications are markedly elevated (Harris and Eastman, 2000; Colagiuri and Davies, 2009). The diagnostic markers, such as fasting plasma glucose (FPG), oral glucose tolerance (OGT), and glycated hemoglobin (HbA1c), are the current gold standards for diagnosing T2DM in patients with hyperglycemia but their ability to detect T2DM is limited by the long asymptomatic phase of early T2DM (Lindström and Tuomilehto, 2003; Carson et al., 2010). As T2DM is a chronic condition that progresses over a long period and interventions with a proven beneficial effect are available. Early detection of the disease onset might allow immediate interventions to delay or prevent the disease progression (Colagiuri and Davies, 2009; Pan et al., 1997; Harris and Eastman, 2000; Knowler et al., 2002; investigators, 2006; Tuomilehto et al., 2001). We previously defined and characterized the critical transition state prior to the type 2 diabetes disease (Jin et al., 2017). Our analysis of a US state patients' pre-disease clinical history identified a dynamic driver network (DDN) and an associated critical transition state 6 months prior to their first confirmative T2DM state. A meta-analysis reported that 3-months lifestyle interventions decreased the risk for diabetes from the end of intervention up to 10 years later (He et al., 2012). Therefore, the identification of pre-diabetic biomarkers in the circulation of asymptomatic patients could be highly beneficial. as a promising liquid biopsy utility for population health management.

Metabolomics is an emerging technology that allows the comprehensive characterization of metabolites in biological systems. It presents as a most recent addition to the omics family and provides complementary information to genomics and proteomics. Owing to the recent advances in analytical instrumentation and bioinformatic approach, metabolomic analysis based on liquid chromatography-high resolution mass spectrometry (LC/HRMS) has given rise to the in-depth profiling of diverse metabolites in a given biological system with superior sensitivity, thus becoming a powerful platform to discover novel biomarkers in close association with disease phenotypes for clinical applications (Xia et al., 2013). As metabolites are downstream end-products of genetic and proteomic regulations, the characterization of metabolome allows the effects of a plethora of pathological factors from vastly different origins to be determined in a single measurement, enabling the comprehensive understanding of molecular mechanisms underlying the disease phenotypes (Xia et al., 2013).

In this study, we aim to characterize and identify unique metabolic signatures using both targeted and global LC/HRMS for the early T2DM detection.

# 2 MATERIALS AND METHODS

## 2.1 Study Population, Blood Collection, and Clinical Characteristic

Four independent cohorts were assembled for the study. The T2DM cohort included sera from 25 patients with T2DM and 25 healthy individuals for the biomarker discovery and sera from 15 patients with T2DM and 15 healthy controls for the testing. The Pre-T2DM cohort included sera from 76 patients with prediabetes and 62 healthy controls for the model training and sera from 35 patients with prediabetes and 27 healthy controls for the model testing. The study was approved by the Institutional Review Board of Teda Hospital and conducted in accordance with the Declaration of Helsinki. All experiments were performed in compliance with the requirements of the Human Ethics Procedures and Guidelines of the local government. Written informed consent was obtained from all participants. Patients with T2DM were diagnosed based on the American Diabetes Association criteria using FPG (FPG less than 100 mg/dL as non-diabetes, FPG between 100 and 125 mg/dL as prediabetes, and FPG greater than 125 mg/dL as diabetes). The serum samples were obtained by room temperature 30 min clotting of collected blood followed by 5 min centrifuge at 3,000 r/min. All serum samples were aliquoted and stored at −80°C prior to use. The

clinical characteristics, including TC, TG, HDL-C, and LDL-C, were determined by Cobas C 311 blood analyzer from Roche (South San Francisco, CA, United States) by following the protocols from the manufacturer.

## 2.2 Global Metabolomics

### 2.2.1 Metabolite Extraction

The metabolite extraction was conducted as previously described with slight modifications (Nemkov et al., 2017). Briefly, once thawed on ice, 10 μl of serum was extracted by 240 μl of prechilled methanol/acetonitrile/water (5:3:2, v/v) containing 3 stable isotope labeled amino acids as internal standards. Afterwards, the sample was vortexed rigorously for 30 min and centrifuged at 12,000 g for 10 min at 4°C. Thereafter, 200 μl of the supernatant was transferred into auto-sampler vial with micro-insert for LC/MS analysis.

### 2.2.2 Global Metabolomic Profiling

The global metabolomic profiling was performed as previously described with slight modifications (Nemkov et al., 2017). Briefly, 10 μl of serum extract was injected onto a Kinetex C18 column (1.7 μm, 2.1 × 150 mm; Phenomenex, Torrance, CA) via a Vanquish UHPLC system (Thermo Fisher, San Jose, CA, United States). The mobile phase A was water with 0.1% formic acid, and mobile phase B was acetonitrile with 0.1% formic acid. The separation was carried out using isocratic elution with 5% B at a flow rate of 0.25 ml/min for a total run time of 3 min. The eluted metabolites were detected by a Q Exactive Plus mass spectrometer (Thermo Fisher) operated in full scan setup using both electrospray positive and negative modes, operating separately as two independent runs. The conditions of ionization source were set at 3.4 kV for spray voltage, 15 for sheath gas, 5 for aux gas, 325°C for capillary temperature, 55 for S-lens, and 250°C for vaporizer temperature. The MS spectra were acquired with 2 μscans using an AGC target of $1e^6$ and a resolution of 70,000 (FWHM at 200 m/z) from 60 to 900 m/z. The column oven was maintained at 25°C throughout the analysis. Representative chromatograms, with either positive or negative mode, were shown in **Supplementary Figure S1**.

### 2.2.3 Data Pre-Processing

Following the analysis, the raw files were first converted into mzXML files by msconvert software from ProteoWizard Tools (http://proteowizard.sourceforge.net/tools.shtml). Subsequently, metabolic features with unique mass/charge ratio and retention time were extracted, aligned, quantified, grouped, and annotated using the XCMS online package (Smith et al., 2006). Afterwards, the obtained features underwent isotopic removal, blank subtraction, and missing value filtering and the QC-based signal drifting correction to normalize the undesired variations. The QC-based signal drifting correction based on locally estimated scatterplot smoothing was performed using StatTarget package (Luan et al., 2018).

### 2.2.4 Metabolite Identification and Metabolic Network and Pathway Analysis

Identification of significant metabolite was conducted by matching the experimental retention time and accurate mass of the metabolic features against an in-house library containing 600 + authentic standards. The metabolic networking analysis and pathway enrichment analysis were implemented using the built-in function available on MetaboAnalyst 4.0 (Chong et al., 2019).

### 2.2.5 QA/QC

The raw signals of internal standards were used to evaluate the extraction efficiency and instrumentation performance. The LC-MS system was conditioned with 10 consecutive injections of QC samples prior to the analysis of unknown samples. The batchwise systematic and random variations were evaluated by repeatedly injecting the QC samples that were spaced evenly. Signal drifting along the batch was corrected using a QC-based approach (Dunn et al., 2011). Between batches, the system was maintained and calibrated to meet the requirements for technical specifications.

### 2.2.6 Metabolic Features

The data acquisitions were implemented in both electrospray ionization positive (ESI+) and negative (ESI−) modes to improve the coverage of serum metabolome. Following the data acquisition, the metabolic features were extracted, aligned, quantified, grouped, and annotated to generate a sample-feature intensity matrix with annotations for subsequent qualification. In total, 1958 and 2,892 metabolic features were detected in all serum samples from ESI+ and ESI−, respectively. After isotope removal ([M+1]/[M+2] ions), blank subtracted (signal-to-noise ≥5), and missing value filtered (≤ 20% missing values in QC samples), 751 and 1,003 metabolic features were qualified for ESI+ and ESI−, respectively. Upon normalized by QC-based signal drifting correction based on locally estimated scatterplot smoothing (LOESS), 323 and 295 metabolic features were reproducibly detected in QC replicates (coefficient variation ≤30%) for ESI+ and ESI-, respectively, and thus selected for the downstream data interpretation.

### 2.2.7 Structure Determination

Metabolite biomarker identification was first performed as a Tier 1 or 2 identification with chemical standards according to MS1 (Schymanski et al., 2014). With tandem mass spectrometry (MS/MS, Thermo Q Exactive Plus) data of blood samples and manual review confirmation, the generated MS1/MS2 pairs were searched in the public databases: HMDB (http://www.hmdb.ca/), MoNA (http://mona.fiehnlab.ucdavis.edu/), MassBank (http://www.massbank.jp/), METLIN (https://metlin.scripps.edu), and NIST (https://www.nist.gov/). The metabolites of interest were procured and subjected to a Tier 1 identification comparing the retention time, MS1 and MS2 patterns with the biomarker candidates, using the same LCMS/MS protocol with the blood sample analysis.

## 2.3 Targeted Metabolomics

### 2.3.1 Metabolite Extraction

For MS/MS analysis, 10 μl of serum, 10 μl of internal standard solution (AA: amino acids; AC: acylcarnitines; FA: fatty acids), 90 μl of extraction buffer, and 200 μl hexane were added for extraction. The sample was vortexed vigorously for 1 min and

centrifuged at 12,000 g for 5 min 80 μl of lower layer was transferred into an auto-sampler vial for the analysis of amino acids and acylcarnitines. 180 μl of upper layer was transferred into another 1.5 ml centrifugal tube and was dried under nitrogen stream. The residue was reconstituted with 100 μl of derivatization buffer and incubated at 95°C for 15 min. After derivatization, 100 μl of neutralization buffer was added into each sample. The neutralized sample was vortexed vigorously for 1 min and centrifuged at 15,000 g for 3 min 100 μl of each reconstituted sample was transferred into an auto-sampler vial for the analysis of fatty acids.

### 2.3.2 Targeted Metabolomic Profiling

#### 2.3.2.1 AA and AC

2 μl of serum extract was injected into a Vanquish UHPLC system (Thermo Fisher, San Jose, CA, United States). The mobile phase A was water with 0.5% formic acid and 10 mM Ammonium Formate, mobile phase B was methanol with 0.5% formic acid and 10 mM Ammonium Formate. The separation was carried out using isocratic elution with 50% B at a flow rate of 0.10 ml/min. The eluted metabolites were detected by an Altis mass spectrometer (Thermo Fisher) operated in SRM setup using electrospray positive mode. The conditions of ionization source were set at 3.5 kV for spray voltage, 20 for sheath gas, 5 for aux gas, 300°C for ion transfer tube temperature, 200°C for vaporizer temperature. The MS spectra were acquired using an cycle time of 0.8 and Q1 resolution (FWHM) of 0.7, Q3 resolution (FWHM) of 0.7, CID gas of 1.5, chromatographic peak width of 12. The column oven was maintained at 30°C throughout the analysis.

#### 2.3.2.2 FA

5 μL of serum extract was injected into a Vanquish UHPLC system (Thermo Fisher, San Jose, CA, United States). The mobile phase was methanol with 0.5% formic acid and 10 mM Ammonium Formate. The separation was carried out using isocratic elution with at a flow rate of 0.10 ml/min. The eluted metabolites were detected by an Altis mass spectrometer (Thermo Fisher) operated in SRM setup using APCI positive mode. The conditions of ionization source were set at 6 μA for ion discharge current, 20 for sheath gas, 5 for aux gas, 300°C for ion transfer tube temperature, 300°C for vaporizer temperature. The MS spectra were acquired using an cycle time of 1 and Q1 resolution (FWHM) of 0.7, Q3 resolution (FWHM) of 0.7, CID gas of 1.5, chromatographic peak width of 12. The column oven was maintained at 30°C throughout the analysis.

Multiple reaction monitoring (MRM) is the most common targeted method for quantitation of analytes by LC/MS/MS. In MRM, ions are selected to make it through the first quadrupole and into the collision cell. In this study, the transition from precursor/parent ion to product/daughter ions is referred to as an ion transition (**Supplementary Table S1**).

### 2.3.3 Statistical Analysis

The normalized intensity of metabolic features from various samples were log-transformed and auto scaled prior to the statistical analysis. Multivariate and univariate statistical analysis were performed using MetaboAnalyst 4.0 (https://www.metaboanalyst.ca/) (Chong et al., 2019). A global false discovery rate of 5% was applied to correct for the errors from multiple hypothesis testing. All statistical analyses were preformed using R packages (Tibshirani, 2006; Team, 2008). We have applied power analysis (Tibshirani, 2006) to determine the sample size for the testing of the T2DM and Pre-T2DM models (**Supplementary Figures 2A,B**).

### 2.3.4 Machine Learning Approach Evaluation

The learning procedure was empirically compared against a number of standard multivariate algorithms, including Random Forest (Breiman, 2001), Elastic Net (Zou and Hastie, 2005) and XGboost (Chen and Guestrin, 2016). All algorithms were evaluated using the same two-layer leave-one-subject-out cross validation strategy.

## 3 RESULTS

### 3.1 Study Design

T2DM is a highly prevalent chronic metabolic disorder characterized by hyperglycemia. We implemented both the global and targeted mass spectrometry (MS)-based metabolomics to advance T2DM research. The time associated with traditional chromatographic methods for resolving metabolites prior to mass analysis has limited the potential to perform large-scale, highly powered metabolomics studies, and clinical applications. Therefore, we applied a 3-min method (Nemkov et al., 2017) for the rapid profiling of central metabolic pathways through UHPLC/MS. Moreover, we applied targeted metabolomics to profile amino acids, acylcarnitines, and fatty acids as several studies suggest a central role for oxidative stress in the pathogenesis of the disease. The short-, medium-, and long-chain acylcarnitines are a family of metabolites known to be dysregulated in T2DM (Bene et al., 2018), linked to peripheral insulin resistance. These acylcarnitines have an indispensable role in lipid metabolism via their involvement in the less severe disruptions in β-oxidation of long-chain fatty acids in T2DM. The dysregulated fatty acid metabolism along with tissue lipid accumulation is generally assumed to be associated in the development of insulin resistance and T2DM (Sears and Perry, 2015). Increased levels of total plasma free fatty acid and branched-chain amino acids (BCAAs) are associated with T2DM pathogenesis (Sjögren et al., 2021).

Shown in **Figure 1**, four independent cohorts were assembled for the study. The T2DM cohort included sera from 25 patients with T2DM and 25 healthy individuals for the biomarker discovery and sera from 15 patients with T2DM and 15 healthy controls for the testing. The Pre-T2DM cohort included sera from 76 patients with prediabetes and 62 healthy controls for the model training and sera from 35 patients with prediabetes and 27 healthy controls for the model testing.
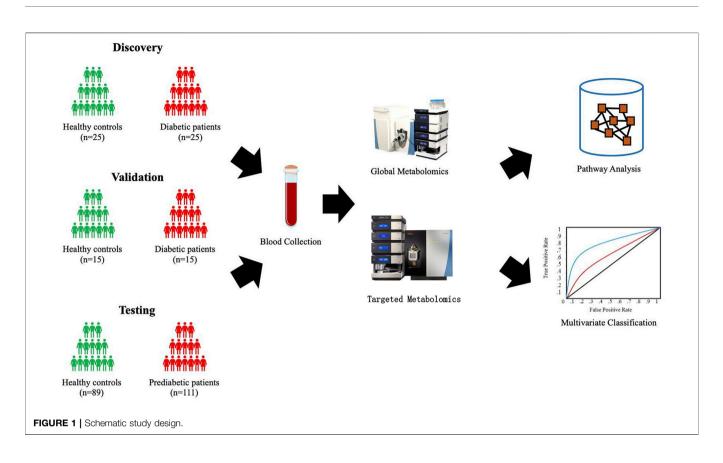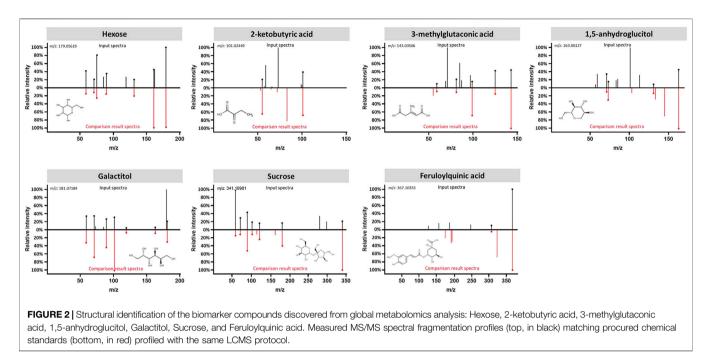
**FIGURE 1 |** Schematic study design.

**TABLE 1 |** Demographic table.

| Characteristic | Diabetic Cohort | | | | | | Pre-diabetic Cohort | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Training | | | Testing | | | Training | | | Testing | | |
| | Non-Diabetes (n = 25) | Diabetes (n = 25) | p value | Non-Diabetes (n = 15) | Diabetes (n = 15) | p value | Non-Diabetes (n = 62) | Pre-Diabetes (n = 76) | p value | Non-Diabetes (n = 27) | Pre-Diabetes (n = 35) | p value |
| Age (year) | 45.9 (1.1) | 52.2 (4.7) | 0.045 | 69.5 (3.2) | 67.2 (5.9) | 0.192 | 54.5 (5.2) | 53.8 (6.2) | 0.441 | 54.3 (4.2) | 55.5 (6.8) | 0.404 |
| Gender | | | | | | | | | | | | |
| Male | 10 (40.0%) | 15 (60.0%) | | 5 (33.3%) | 10 (66.7%) | | 29 (46.8) | 53 (69.7) | | 7 (25.9) | 21 (60) | |
| Female | 15 (60.0%) | 10 (40.0%) | | 10 (66.7%) | 5 (33.3%) | | 33 (53.2) | 23 (30.3) | | 20 (74.1) | 14 (40) | |
| FPG (mM) | 5.1 (1.8) | 8.9 (3.0)*** | $7.1 \times 10^{-10}$ | 5.1 (0.3) | 9.6 (2.2)*** | $1.7 \times 10^{-6}$ | 5.2 (0.3) | 6 (0.3)*** | $3.7 \times 10^{-5}$ | 5.2 (0.6) | 6.1 (0.4)*** | $2.9 \times 10^{-6}$ |
| TC (mM) | 4.9 (0.8) | 5.1 (1.2) | 0.72 | 9.6 (2.2) | 4.7 (1.1) | 0.426 | 4.7 (0.9) | 4.8 (1) | 0.319 | 5.2 (0.9) | 4.9 (1) | 0.231 |
| Creatinine (mM) | 65.9 (13.7) | 66.8 (19.9) | 0.98 | 68.2 (11.0) | 68 (14.1) | 0.485 | 68 (14.1) | 71.2 (13.5) | 0.175 | 64.1 (10.8) | 69.9 (13.1) | 0.069 |
| TG (mM) | 1.3 (0.6) | 2.3 (1.0)*** | $4.5 \times 10^{-4}$ | 1.4 (0.5) | 1.4 (0.6) | $8.4 \times 10^{-5}$ | 1.4 (0.6) | 1.8 (1.1)** | 0.03 | 1.3 (0.5) | 1.8 (0.9)** | 0.009 |
| HDL-C (mM) | 1.4 (0.6) | 1.1 (0.4) | 0.049 | 1.4 (0.3) | 1.3 (0.4) | 0.0236 | 1.3 (0.3) | 1.2 (0.4)*** | 0.249 | 1.5 (0.4) | 1.1 (0.3)*** | $2.8 \times 10^{-4}$ |
| LDL-C (mM) | 3.2 (0.8) | 3.1 (1.1) | 0.791 | 3.0 (0.7) | 3.0 (1.0) | 0.86 | 3 (0.8) | 3 (1) | 0.93 | 3.4 (0.9) | 3.2 (1) | 0.299 |

*All values are presented as mean (SD) except for gender where percentage is applied. The p values were determined by Mann-Whitney U test and classified into several categories based on following criteria by comparing against the control group: \*: p < 0.05; \*\*: p < 0.01; \*\*\*: p < 0.001. FPG, fast plasma glucose; TC, total cholesterol; HDL-C, high density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol; TG, triglycerides.*

**FIGURE 2 |** Structural identification of the biomarker compounds discovered from global metabolomics analysis: Hexose, 2-ketobutyric acid, 3-methylglutaconic acid, 1,5-anhydroglucitol, Galactitol, Sucrose, and Feruloylquinic acid. Measured MS/MS spectral fragmentation profiles (top, in black) matching procured chemical standards (bottom, in red) profiled with the same LCMS protocol.

## 3.2 Sample Demographics and Clinical Characteristics

The sample demographics and clinical characteristics are summarized in **Table 1**. T2DM associated clinical characteristics were analyzed: fasting plasma glucose (FPG), high-density lipoprotein cholesterol, triglyceride, and creatinine were found to be statistically significant ($p < 0.05$) between case and control groups, whereas the other clinical characteristics, including low density lipoprotein (LDL) cholesterol, total cholesterol, and serum creatinine, were insignificant ($p > 0.05$) between case and control groups among all studied cohorts.

## 3.3 Identification of Metabolic Signature for T2DM

The data interpretation was performed in a stepwise manner using a suite of bioinformatic approaches. The Mann-Whitney U test with a 5% false discovery rate was used to determine the statistical significance of individual metabolic features by comparing the case to control groups from the T2DM training cohort. From the global metabolomic analysis, 12 validated features were putatively identified, with MS1 Tier 1 identification, by matching the retention times and accurate masses against an in-house library containing 600+ authentic standards. 8/12 were structurally determined (**Figure 2**, with Tier 2 identification) were determined: hexose $[M + Na]^+$, hexose $[M-H]^-$, 1,5-anhydroglucitol, feruloylquinic acid, galactitol, 2-ketobutyric acid, 3-methylglutaconic acid, and sucrose. From the targeted metabolomic analysis, 14 were identified: pyroglutamic acid, ornithine, hexose, valine, leucine/isoleucine, tyrosine, phenylalanine, tryptophan, C16-carnitine, C14-carnitine, C5DC-carnitine/C6OH-carnitine, C5OH-carnitine,

tritriacontanoic acid (33:0) butyl ester, and dotriacontanoic acid (32:0) butyl ester. These 22 metabolites constitute the T2DM biomarker panel for the following analyses. The univariate statistics of the T2DM panel biomarkers are illustrated in **Figure 3**.

With unsupervised clustering, the heatmap revealed a distinct T2DM pattern of these 22 compound biomarkers (**Figure 4**), separating T2DM subjects from the non-diabetic controls in both training and testing cohorts.

## 3.4 Exploration of the T2DM Metabolic Panel to Assess Prediabetic Patients in the General Population

We set to apply the T2DM panel to assess those who are at high risk of developing T2DM. Not as a pronounced pattern of the T2DM panel seen in the T2DM cohort, the pattern of these 22 compound biomarkers (**Figure 4**) persists in the Pre-T2DM cohort, separating prediabetic subjects from the non-diabetic controls in both training and testing cohorts. We applied different machine learning algorithms, including Random Forest, XGBoost and Elastic Net, to these datasets for the T2DM assessment. For the T2DM dataset, we developed the model with the T2DM training set which achieved similar superb performance for the testing set: Random Forest, ROC AUC 0.98; XGBoost, ROC AUC 0.99; Elastic Net, ROC AUC 1.000 (**Figure 5A**). For the Pre-T2DM cohort, we developed the model with the Pre-T2DM training set and found XGBoost performed the best for the Pre-T2DM testing set: Random Forest, ROC AUC 0.78; XGBoost, ROC AUC 0.84; Elastic Net, ROC AUC 0.78 (**Figure 5B**). Therefore, XGBoost Pre-T2DM model would be the model for the analysis of the general population at risk for the T2DM at early prediabetic stage.
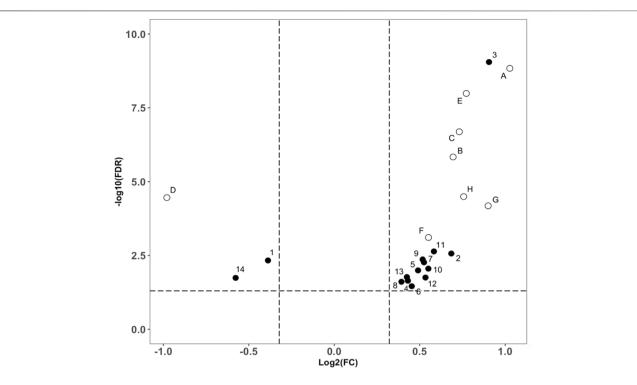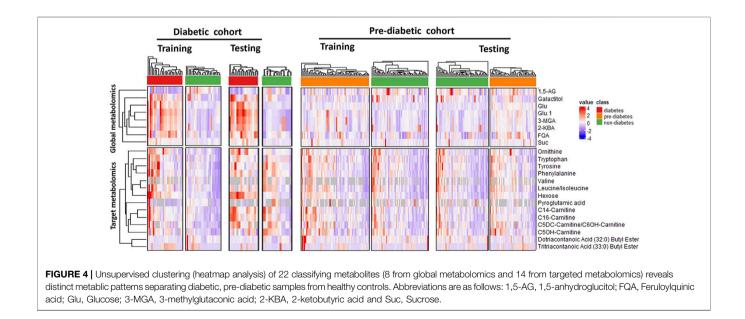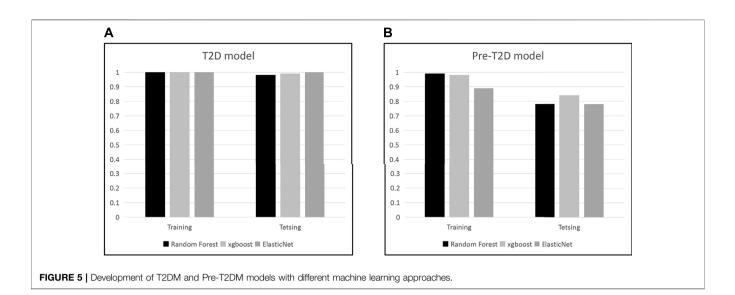
**FIGURE 3 |** Volcano plot analysis of the biomarker compounds revealed in our discovery analysis with the T2DM training cohort. Filled circles representing compounds discovered from the targeted metabolomics analysis. **(A)**: Glucose ([M + Na]+), **(B)**: 2-Ketobutyric acid, **(C)**: 3-Methylglutaconic acid, **(D)**: 1,5-anhydroglucitol, **(E)**: Glucose ([M-H]-), **(F)**: Galactitol, **(G)**: Sucrose, **(H)**: Feruloylquinic acid. Open circles representing compounds discovered from the global metabolomics analysis, 1: Pyroglutamic acid, 2: Ornithine, 3: Hexose, 4: Valine, 5: Leucine/Isoleucine, 6: Tyrosine, 7: Phenylalanine, 8: Tryptophan, 9: C16-Carnitine, 10: C14-Carnitine, 11: C5DC-Carnitine/C6OH-Carnitine, 12: C5OH-Carnitine, 13: Tritriacontanoic Acid (33:0) Butyl Ester, and 14: Dotriacontanoic Acid (32:0) Butyl Ester.



**FIGURE 4 |** Unsupervised clustering (heatmap analysis) of 22 classifying metabolites (8 from global metabolomics and 14 from targeted metabolomics) reveals distinct metablic patterns separating diabetic, pre-diabetic samples from healthy controls. Abbreviations are as follows: 1,5-AG, 1,5-anhydroglucitol; FQA, Feruloylquinic acid; Glu, Glucose; 3-MGA, 3-methylglutaconic acid; 2-KBA, 2-ketobutyric acid and Suc, Sucrose.

Cross-sectional retrospective cohort analysis (Le et al., 2019) of the 2003–2014 National Health and Nutrition Examination Survey (NHANES) reported the prevalence of prediabetes (34.8%). Shown in **Figure 6**, the prevalence adjusted positive predictive values (PPV), a measure of clinical risk, is shown as a function of XGBoost Pre-T2DM model predictor score. Stratification of subjects with increasing predictor scores occurs as positive predictive values (PPV) from a background value (population rate of 34.8%) to
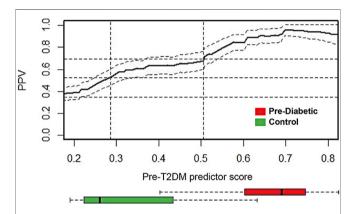
**FIGURE 5 |** Development of T2DM and Pre-T2DM models with different machine learning approaches.



**FIGURE 6 |** Prevalence-corrected positive predictive values (PPV) was plotted as a function of Pre-T2DM predictor score for the Pre-T2DM cohort samples. Horizontal dashed lines identify the average population risk of 34.8%, and relative risks of 1.5X (52.2%) and 2X (69.6%). Vertical dashed lines identify corresponding predictor scores. The confidence interval about the PPV curve (gray shaded area) was estimated using all Pre-T2DM subjects. Confidence intervals about the PPV were calculated with the normal approximation of the error for binomial proportions. Box plots at the foot of the figure correspond to the distributions of predictor scores for prediabetic and control subjects. The PPV curve and the box plots share the same predictor score axis.

relative risks of 1.5X (52.2%) and 2X (69.6%) (dashed lines) and higher. The distribution of Pre-T2DM predictor score values for subjects color-coded (green, controls; red, Pre-T2DM) are shown in box plots in **Figure 6**. Using the risk curve in **Figure 6**, Pre-T2DM subjects were identified as high or low risk according to a predictor score cutoff corresponding to 2X relative risk (PPV of 69.6%, predictor score = 0.505).

# 4 DISCUSSION

Early detection of T2DM remains a challenge in current clinical practice. Diagnostic markers including FPG, OGT,

and Hb1Ac are gold standards in identifying impaired glucose disposition and poor glycose management in patients with T2DM. However, a long latent phase of T2DM onset occurs asymptomatically before hyperglycemic symptoms are manifested (Lindström and Tuomilehto, 2003; Carson et al., 2010). The current screening test utilizes several risk factors such as metabolic syndrome and family history for early assessment of T2DM, which are short in both sensitivity and specificity. Moreover, T2DM is a chronic condition that progresses over a long time, and early detection of T2DM might allow immediate interventions such as dietary control, weight loss, and glucose-lowering medications for effective management of the disease (Pan et al., 1997; Knowler et al., 2002; investigators, 2006; Tuomilehto et al., 2001). Aiming for the early detection of general population at high risk of T2DM, we employed an omics-based approach integrating global metabolomic profiling and bioinformatic analysis to identify potential pre-diabetic biomarkers in blood. The efforts from the T2DM metabolic panel discovery, testing, and application to prediabetic cohorts to case find subjects at high risk for T2DM led to the identification of 22 metabolites in circulation with a unique metabolic pattern associated with T2DM risk. The multivariate analysis based on those metabolites indicated the presence of significantly dysregulated metabolism in patients with early progressing T2DM, which was differentiative between pre-diabetic and non-diabetic subjects. Longitudinal measurement of the metabolic panel thus might offer an opportunity for assessing diabetic onset before the development of relevant clinical signs and symptoms.

Our study also identified 22 metabolites that were significantly altered in T2DM, and models were developed in risk stratification to case find the general population at risk of the onset of prediabetes. Hexose represents monosaccharides with a six-carbon scaffold, and glucose is known to constitute over 95% of hexose content in humans. Sucrose is a nonreducing disaccharide composed of glucose and fructose linked via their

anomeric carbons. Increased dietary intake of sucrose is considered a risk factor for diabetic pathogenesis, and a recent study has discovered that the dietary replacement of sucrose by fructose can significantly improve glucose disposition and insulin sensitivity (Evans et al., 2017). Our study revealed significantly increased circulating hexose and sucrose induced by defective insulin secretion and impaired insulin sensitivity in T2DM, consistent with disease pathophysiology (Gonzalez-Franquesa et al., 2016; Klein and Shearer, 2016; Arneth et al., 2019; Sun et al., 2019). 2–KBA is produced by amino acid catabolism (threonine and methionine) and glutathione anabolism (cysteine formation pathway) and is metabolized to propionyl-CoA and carbon dioxide (Gall et al., 2010; Syed Ikmal et al., 2013). During the formation of 2–KBA, 2-hydroxybutyrate (2–HBA) is formed as a by-product, and mounting evidence has demonstrated that circulating 2-HBA is inversely correlated with insulin sensitivity and elevated 2-HBA is an early marker for both insulin resistance and impaired glucose regulation independent of sex, age, and BMI (Gall et al., 2010; Syed Ikmal et al., 2013). 2–KBA can also be produced in the conversion of cystathionine to cysteine. In the presence of enhanced oxidative stress in T2DM, a metabolic shift occurs by stimulating homocysteine production from transmethylation of methionine to transsulfuration of homocysteine to produce cystathionine and cysteine for meeting the increased demand of glutathione, the primary cytoplasmic antioxidant (Ratnam et al., 2002; Schalinske, 2003). Cystathionine is a metabolic intermediate formed by the condensation of serine and homocysteine via the catalysis of cystathionine $\beta$-synthase (CBS) for homocysteine transsulfuration. Markedly elevated transcriptional expression and enzymatic activity of CBS were observed in a streptozotocin-induced diabetic rat model, and the administration of regulatory hormones such as insulin was able to revert the elevation of CBS to result in the attenuation of diabetic phenotype (Ratnam et al., 2002). Our study discovered the upregulation of 2-KBA and cystathionine in T2DM, further confirming their mechanistic roles in glycemic regulation. 1,5-AG is a metabolically inert polyol that competes with glucose for reabsorption in the kidneys and has been validated as a marker of short-term glycemic control. Its level is inversely correlated with glucose level in blood circulation, and reduced plasma 1,5-AG is a sensitive marker for increased urinary excretion of glucose (Dungan, 2008; Kim and Park, 2013). Several studies have proposed 1,5-AG as a short-term retrospective marker for monitoring glucose excursions (Dungan, 2008; Kim and Park, 2013). Our results revealed the downregulation of 1,5-AG in T2DM, again validating its utility as a biomarker for monitoring short-term glycemic management. FQA is a subclass of chlorogenic acids (CGAs) that are esters formed between caffeic and quinic acids. CGA represents an abundant group of plant polyphenols present in the human diet. *In-vitro* and *in-vivo* studies have indicated their potent anti-oxidative and anti-inflammation activities by prolonging the lifetime of the phenoxyl radical and upregulating the pro-inflammatory cytokines at transcriptional level (Liang and Kitts, 2016). Results from epidemiological studies have suggested that the

consumption of beverages containing CGA is associated with reduced risks of developing chronic diseases such as T2DM (Tajik et al., 2017). Our results illustrated a significant upregulation of FQA in T2DM, suggesting a surging demand for antioxidants to counteract increased oxidative stress in diabetes. Our targeted analysis revealed branched chain amino acid (BCAA) valine and leucine/isoleucine, and aromatic amino acid (AAA) phenylalanine, tyrosine and tryptophan as biomarkers for T2DM. Elevated plasma concentrations of BCAA and AAA circulate up to 10 years prior to a diagnosis of T2DM, hence interest in their role as biomarkers for insulin resistance and T2DM (Cheng et al., 2012; Saleem et al., 2019). Our results of C16/C14/C5DC/C6OH/C5OH carnitine alterations in serum concentrations in both T2DM and prediabetic states are in line with previous findings about the role of mitochondrial function in the complex pathogenesis of type 2 diabetes. The mitochondrial function seems to play a significant role in the complex pathogenesis of insulin resistance (Kelley et al., 2002; Koves et al., 2008; Mihalik et al., 2010). Recent studies suggest that concentrations of various acylcarnitines are associated with insulin resistance and T2D (Adams et al., 2009; Huffman et al., 2009; Tai et al., 2010; Ha et al., 2012) and incomplete fatty acid oxidation results in elevated acylcarnitine concentrations (Koves et al., 2005). In addition, our study also identified several novel markers, including 3-MGA, galactitol, ornithine, pyroglutamic acid, and dotria/tritriacontanoic acid (32:0/33:0) butyl esters with unknown implications in T2DM and further investigations on their pathological implications might uncover novel molecular mechanisms to provide more comprehensive understanding of T2DM pathophysiology and guide the development of new therapeutic with better efficacy for treating T2DM.

Despite the strengths of our study results, these findings need to be evaluated in the context of the study limitations. A prospective longitudinal high resolution sampling cohort assembly with the emerging T2DM as a clinical outcome can validate the clinical utility of our T2DM metabolic panel to allow the early detection of the disease onset for proactive intervention.

In conclusion, we presented a metabolomics-based discovery study with T2DM and prediabetic subjects. Using sera from four independent cohorts, a panel of metabolic biomarker candidates was discovered from T2DM cohort and tested with Pre-T2DM cohort to risk stratify the general population to case find the high risk subjects of emerging T2DM. Application of this predictor should enable early and sensitive detection of generation population at risk of T2DM. This early detection may improve clinical outcomes through increased clinical surveillance as well as accelerate the development of clinical interventions for T2DM prevention.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Teda Hospital. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

XS, XL, and XC contributed to concept development and design. LZ, YZ, XL, ST, WL, YD, and ZL contributed to the acquisition of data. LZ, QH, BJ, ZL (16th author), ZL (18th author), RM, YJ, KS, and XC contributed to the analysis and interpretation of data. XS, XBL, QH, KH, and XC drafted and critically revised the manuscript. XBL, CJC, ZH, JS, and KS revised the manuscript. All the authors gave final approval of the version to be submitted and agreed to be accountable for all aspects of the work.

## FUNDING

## ACKNOWLEDGMENTS

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.841209/full#supplementary-material

## REFERENCES

Adams, S. H., Hoppel, C. L., Lok, K. H., Zhao, L., Wong, S. W., Minkler, P. E., et al. (2009). Plasma Acylcarnitine Profiles Suggest Incomplete Long-Chain Fatty Acid β-Oxidation and Altered Tricarboxylic Acid Cycle Activity in Type 2 Diabetic African-American Women. *J. Nutr.* 139 (6), 1073–1081. doi:10.3945/jn.108.103754

Arneth, B., Arneth, R., and Shams, M. (2019). Metabolomics of Type 1 and Type 2 Diabetes. *Ijms* 20 (10), 2467. doi:10.3390/ijms20102467

Bene, J., Hadzsiev, K., and Melegh, B. (2018). Role of Carnitine and its Derivatives in the Development and Management of Type 2 Diabetes. *Nutr. Diabetes* 8 (1), 8. doi:10.1038/s41387-018-0017-1

Breiman, L. (2001). Random Forests. *Machine Learn.* 45 (1), 5–32. doi:10.1023/A:1010933404324

Brownlee, M. (2001). Biochemistry and Molecular Cell Biology of Diabetic Complications. *Nature* 414 (6865), 813–820. doi:10.1038/414813a

Brownlee, M. (2005). The Pathobiology of Diabetic Complications. *A Unifying Mechanism* 54 (6), 1615–1625. doi:10.2337/diabetes.54.6.1615

Carson, A. P., Reynolds, K., Fonseca, V. A., and Muntner, P. (2010). Comparison of A1C and Fasting Glucose Criteria to Diagnose Diabetes Among U.S. Adults. *Diabetes Care* 33 (1), 95–97. doi:10.2337/dc09-1227

Chen, T., and Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discov. Data Mining (Kdd'16). San Francisco, CA, 785-794.

Cheng, S., Rhee, E. P., Larson, M. G., Lewis, G. D., McCabe, E. L., Shen, D., et al. (2012). Metabolite Profiling Identifies Pathways Associated with Metabolic Risk in Humans. *Circulation* 125 (18), 2222–2231. doi:10.1161/CIRCULATIONAHA.111.067827

Chong, J., Wishart, D. S., and Xia, J. (2019). Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr. Protoc. Bioinformatics* 68 (1), e86. doi:10.1002/cpbi.86

Colagiuri, S., and Davies, D. (2009). The Value of Early Detection of Type 2 Diabetes. *Curr. Opin. Endocrinol. Diabetes Obes.* 16 (2), 95–99. doi:10.1097/MED.0b013e328329302f

DeFronzo, R. A. (1988). The Triumvirate: β-Cell, Muscle, Liver: A Collusion Responsible for NIDDM. *Diabetes* 37 (6), 667–687. doi:10.2337/diab.37.6.667

Dungan, K. M. (2008). 1,5-anhydroglucitol (GlycoMark) as a Marker of Short-Term Glycemic Control and Glycemic Excursions. *Expert Rev. Mol. Diagn.* 8 (1), 9–19. doi:10.1586/14737159.8.1.9

Dunn, W. B., Broadhurst, D., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., et al. (2011). Procedures for Large-Scale Metabolic Profiling of Serum and Plasma Using Gas Chromatography and Liquid Chromatography Coupled to Mass Spectrometry. *Nat. Protoc.* 6 (7), 1060–1083. doi:10.1038/nprot.2011.335

Evans, R. A., Frese, M., Romero, J., Cunningham, J. H., and Mills, K. E. (2017). Fructose Replacement of Glucose or Sucrose in Food or Beverages Lowers Postprandial Glucose and Insulin without Raising Triglycerides: a Systematic Review and Meta-Analysis. *Am. J. Clin. Nutr.* 106 (2), 506–518. doi:10.3945/ajcn.116.145151

Gall, W. E., Beebe, K., Lawton, K. A., Adam, K.-P., Mitchell, M. W., Nakhle, P. J., et al. (2010). α-Hydroxybutyrate Is an Early Biomarker of Insulin Resistance and Glucose Intolerance in a Nondiabetic Population. *PLOS ONE* 5 (5), e10883. doi:10.1371/journal.pone.0010883

Garvey, W. T., Olefsky, J. M., Matthaei, S., and Marshall, S. (1987). Glucose and Insulin Co-regulate the Glucose Transport System in Primary Cultured Adipocytes. A New Mechanism of Insulin Resistance. *J. Biol. Chem.* 262 (1), 189–197. doi:10.1016/S0021-9258(19)75908-0

Gonzalez-Franquesa, A., Burkart, A. M., Isganaitis, E., and Patti, M.-E. (2016). What Have Metabolomics Approaches Taught Us about Type 2 Diabetes? *Curr. Diab Rep.* 16 (8), 74. doi:10.1007/s11892-016-0763-1

Ha, C. Y., Kim, J. Y., Paik, J. K., Kim, O. Y., Paik, Y.-H., Lee, E. J., et al. (2012). The Association of Specific Metabolites of Lipid Metabolism with Markers of Oxidative Stress, Inflammation and Arterial Stiffness in Men with Newly Diagnosed Type 2 Diabetes. *Clin. Endocrinol.* 76 (5), 674–682. doi:10.1111/j.1365-2265.2011.04244.x

Harris, M. I., and Eastman, R. C. (2000), Early Detection of Undiagnosed Diabetes Mellitus: a US Perspective *Diabetes Metab. Res. Rev.* 16. 2, 230–236. doi:10.1002/1520-756010.1002/1520-7560(2000)9999:9999<::aid-dmrr122>3.0.co;2-w

He, D., Liu, Z. P., Honda, M., Kaneko, S., and Chen, L. (2012). Coexpression Network Analysis in Chronic Hepatitis B and C Hepatic Lesions Reveals Distinct Patterns of Disease Progression to Hepatocellular Carcinoma. *J Mol Cell Biol.* 4, 140–152. doi:10.1093/jmcb/mjs011

Huffman, K. M., Shah, S. H., Stevens, R. D., Bain, J. R., Muehlbauer, M., Slentz, C. A., et al. (2009). Relationships between Circulating Metabolic Intermediates and Insulin Action in Overweight to Obese, Inactive Men and Women. *Diabetes Care* 32 (9), 1678–1683. doi:10.2337/dc08-2075

investigators, D. t., Gerstein, H. C., Yusuf, S., Bosch, J., Pogue, J., Sheridan, P., et al. (2006). Effect of Rosiglitazone on the Frequency of Diabetes in Patients with Impaired Glucose Tolerance or Impaired Fasting Glucose: a Randomised Controlled Trial. *Lancet* 368 (9541), 1096–1105. doi:10.1016/s0140-6736(06)69420-8

Jin, B., Liu, R., Hao, S., Li, Z., Zhu, C., Zhou, X., et al. (2017). Defining and Characterizing the Critical Transition State Prior to the Type 2 Diabetes Disease. *PLOS ONE* 12 (7), e0180937. doi:10.1371/journal.pone.0180937

Kelley, D. E., He, J., Menshikova, E. V., and Ritov, V. B. (2002). Dysfunction of Mitochondria in Human Skeletal Muscle in Type 2 Diabetes. *Diabetes* 51 (10), 2944–2950. doi:10.2337/diabetes.51.10.2944

Kim, W. J., and Park, C.-Y. (2013). 1,5-Anhydroglucitol in Diabetes Mellitus. *Endocrine* 43 (1), 33–40. doi:10.1007/s12020-012-9760-6

Klein, M. S., and Shearer, J. (2016). Metabolomics and Type 2 Diabetes: Translating Basic Research into Clinical Application. *J. Diabetes Res.* 2016, 1–10. doi:10.1155/2016/3898502

Knowler, W. C., Barrett-Connor, E., Fowler, S. E., Hamman, R. F., Lachin, J. M., Walker, E. A., et al. (2002). Reduction in the Incidence of Type 2 Diabetes with Lifestyle Intervention or Metformin. N. Engl. J. Med. 346 (6), 393–403. doi:10.1056/NEJMoa012512

Koves, T. R., Li, P., An, J., Akimoto, T., Slentz, D., Ilkayeva, O., et al. (2005). Peroxisome Proliferator-Activated Receptor-γ Co-activator 1α-Mediated Metabolic Remodeling of Skeletal Myocytes Mimics Exercise Training and Reverses Lipid-Induced Mitochondrial Inefficiency *. J. Biol. Chem. 280 (39), 33588–33598. doi:10.1074/jbc.M507621200

Koves, T. R., Ussher, J. R., Noland, R. C., Slentz, D., Mosedale, M., Ilkayeva, O., et al. (2008). Mitochondrial Overload and Incomplete Fatty Acid Oxidation Contribute to Skeletal Muscle Insulin Resistance. Cel Metab. 7 (1), 45–56. doi:10.1016/j.cmet.2007.10.013

Le, P., Chaitoff, A., Rothberg, M. B., Alkhouri, N., and McCullough, A. (2019). Trends in Prevalence of Nonalcoholic Fatty Liver Disease in US Adults with Prediabetes. J. Gen. Intern. Med. 34 (11), 2336–2338. doi:10.1007/s11606-019-05163-8

Liang, N., and Kitts, D. D. (2016). Role of Chlorogenic Acids in Controlling Oxidative and Inflammatory Stress Conditions. Nutrients 8 (1), 16. doi:10.3390/nu8010016

Lindström, J., and Tuomilehto, J. (2003). The Diabetes Risk Score. A Pract. tool predict type 2 Diabetes Risk 26 (3), 725–731. doi:10.2337/diacare.26.3.725

Luan, H., Ji, F., Chen, Y., and Cai, Z. (2018). statTarget: A Streamlined Tool for Signal Drift Correction and Interpretations of Quantitative Mass Spectrometry-Based Omics Data. Analytica Chim. Acta 1036, 66–72. doi:10.1016/j.aca.2018.08.002

Mihalik, S. J., Goodpaster, B. H., Kelley, D. E., Chace, D. H., Vockley, J., Toledo, F. G. S., et al. (2010). Increased Levels of Plasma Acylcarnitines in Obesity and Type 2 Diabetes and Identification of a Marker of Glucolipotoxicity. Obesity 18 (9), 1695–1700. doi:10.1038/oby.2009.510

Nemkov, T., Hansen, K. C., and D'Alessandro, A. (2017). A Three-Minute Method for High-Throughput Quantitative Metabolomics and Quantitative Tracing Experiments of central Carbon and Nitrogen Pathways. Rapid Commun. Mass. Spectrom. 31 (8), 663–673. doi:10.1002/rcm.7834

Pan, X.-R., Li, G.-W., Hu, Y.-H., Wang, J.-X., Yang, W.-Y., An, Z.-X., et al. (1997). Effects of Diet and Exercise in Preventing NIDDM in People with Impaired Glucose Tolerance: The Da Qing IGT and Diabetes Study. Diabetes Care 20 (4), 537–544. doi:10.2337/diacare.20.4.537

Ratnam, S., Maclean, K. N., Jacobs, R. L., Brosnan, M. E., Kraus, J. P., and Brosnan, J. T. (2002). Hormonal Regulation of Cystathionine β-Synthase Expression in Liver *. J. Biol. Chem. 277 (45), 42912–42918. doi:10.1074/jbc.M206588200

Robertson, R. P., Harmon, J., Tran, P. O., Tanaka, Y., and Takahashi, H. (2003). Glucose Toxicity in β-Cells: Type 2 Diabetes, Good Radicals Gone Bad, and the Glutathione Connection. Diabetes 52 (3), 581–587. doi:10.2337/diabetes.52.3.581

Rossetti, L., Giaccari, A., and DeFronzo, R. A. (1990). Glucose Toxicity. Diabetes Care 13 (6), 610–630. doi:10.2337/diacare.13.6.610

Rossetti, L., Smith, D., Shulman, G. I., Papachristou, D., and DeFronzo, R. A. (1987). Correction of Hyperglycemia with Phlorizin Normalizes Tissue Sensitivity to Insulin in Diabetic Rats. J. Clin. Invest. 79 (5), 1510–1515. doi:10.1172/JCI112981

Saleem, T., Dahpy, M., Ezzat, G., Abdelrahman, G., Abdel-Aziz, E., and Farghaly, R. (2019). The Profile of Plasma Free Amino Acids in Type 2 Diabetes Mellitus with Insulin Resistance: Association with Microalbuminuria and Macroalbuminuria. Appl. Biochem. Biotechnol. 188 (3), 854–867. doi:10.1007/s12010-019-02956-9

Schalinske, K. L. (2003). Interrelationship between Diabetes and Homocysteine Metabolism: Hormonal Regulation of Cystathionine β-Synthase. Nutr. Rev. 61 (4), 136–138. doi:10.1301/nr.2003.apr.136-138

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., et al. (2014). Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. Environ. Sci. Technol. 48 (4), 2097–2098. doi:10.1021/es5002105

Sears, B., and Perry, M. (2015). The Role of Fatty Acids in Insulin Resistance. Lipids Health Dis. 14 (1), 121. doi:10.1186/s12944-015-0123-1

Sjögren, R. J. O., Rizo-Roca, D., Chibalin, A. V., Chorell, E., Furrer, R., Katayama, S., et al. (2021). Branched-chain Amino Acid Metabolism Is Regulated by ERRα in Primary Human Myotubes and Is Further Impaired by Glucose Loading in Type 2 Diabetes. Diabetologia 64 (9), 2077–2091. doi:10.1007/s00125-021-05481-9

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. Anal. Chem. 78 (3), 779–787. doi:10.1021/ac051437y

Stumvoll, M., Goldstein, B. J., and van Haeften, T. W. (2005). Type 2 Diabetes: Principles of Pathogenesis and Therapy. The Lancet 365 (9467), 1333–1346. doi:10.1016/S0140-6736(05)61032-X

Sun, Y., Gao, H.-Y., Fan, Z.-Y., He, Y., and Yan, Y.-X. (2019). Metabolomics Signatures in Type 2 Diabetes: A Systematic Review and Integrative Analysis. J. Clin. Endocrinol. Metab. 105 (4), 1000–1008. doi:10.1210/clinem/dgz240

Syed Ikmal, S. I. Q., Zaman Huri, H., Vethakkan, S. R., and Wan Ahmad, W. A. (2013). Potential Biomarkers of Insulin Resistance and Atherosclerosis in Type 2 Diabetes Mellitus Patients with Coronary Artery Disease. Int. J. Endocrinol. 2013, 1–11. doi:10.1155/2013/698567

Tai, E. S., Tan, M. L. S., Stevens, R. D., Low, Y. L., Muehlbauer, M. J., Goh, D. L. M., et al. (2010). Insulin Resistance Is Associated with a Metabolic Profile of Altered Protein Metabolism in Chinese and Asian-Indian Men. Diabetologia 53 (4), 757–767. doi:10.1007/s00125-009-1637-8

Tajik, N., Tajik, M., Mack, I., and Enck, P. (2017). The Potential Effects of Chlorogenic Acid, the Main Phenolic Components in Coffee, on Health: a Comprehensive Review of the Literature. Eur. J. Nutr. 56 (7), 2215–2244. doi:10.1007/s00394-017-1379-1

Team, R. (2008). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Tibshirani, R. (2006). A Simple Method for Assessing Sample Sizes in Microarray Experiments. BMC Bioinformatics 7 (1), 106. doi:10.1186/1471-2105-7-106

Tuomilehto, J., Lindström, J., Eriksson, J. G., Valle, T. T., Hämäläinen, H., Ilanne-Parikka, P., et al. (2001). Prevention of Type 2 Diabetes Mellitus by Changes in Lifestyle Among Subjects with Impaired Glucose Tolerance. N. Engl. J. Med. 344 (18), 1343–1350. doi:10.1056/nejm200105033441801

Whiting, D. R., Guariguata, L., Weil, C., and Shaw, J. (2011). IDF Diabetes Atlas: Global Estimates of the Prevalence of Diabetes for 2011 and 2030. Diabetes Res. Clin. Pract. 94 (3), 311–321. doi:10.1016/j.diabres.2011.10.029

Xia, J., Broadhurst, D. I., Wilson, M., and Wishart, D. S. (2013). Translational Biomarker Discovery in Clinical Metabolomics: an Introductory Tutorial. Metabolomics 9 (2), 280–299. doi:10.1007/s11306-012-0482-9

Zou, H., and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. J. R. Stat. Soc B 67 (2), 301–320. doi:10.1111/j.1467-9868.2005.00503.x

**Conflict of Interest:** QH, YD, CJC, KJS, KH, ST, WL, ZH, RM, and JS were employed by the company mProbe Inc. XL, BJ, YZ, YJ, and ZL were employed by the company Tianjin Yunjian Medical Laboratory Institute Co., Ltd. ZL was employed by Shanghai Yunxiang Medical Technology Co., Ltd., Shanghai, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.