ENDOCRINOLOGY RESEARCH AND PRACTICE

SOCIETY OF ENDOCRINOLOGY AND METABOLISM OF TURKEY

ORIGINAL ARTICLE
Endocrinol Res Pract. 2023;27(3):135-147

Metabolite Biomarkers and Predictive Model Analysis for Patients with Type 2 Diabetes Mellitus With and Without Complications

ABSTRACT

Objective: Understanding the pathogenesis of type 2 diabetes mellitus including the interaction between the inherent susceptibility, lifestyles, and environment is believed to cast hope to predict, prevent, and personalize cure for type 2 diabetes mellitus and its complications. To identify the differentially expressed metabolites as potential diabetes-associated metabolite biomarkers that identify individuals with and without diabetes.

Methods: Sixty-four subjects were recruited to identify the systemic metabolic changes and biomarkers related to type 2 diabetes mellitus, and the related complications (ischemic heart disease and chronic kidney disease) using quadrupole time-of-flight liquid chromatography coupled to mass spectrometry. The top 5 biomarkers were identified, and the prediction accuracies for models developed by 4 algorithms were compared.

Result: Tyrosine, tryptophan, glycerophospholipid, porphyrin and chlorophyll, sphingolipid metabolism, and glycosylphosphatidylinositol-anchor biosynthesis were the lipids and amino acid-related pathways differentially regulated in the type 2 diabetes mellitus patients compared to normal subjects and patients with complications. Hydroxyprolyl-leucine and *N*-palmitoyl threonine were higher in patients; 4,4′-Thiobis-2-butanone, geranyl-hydroxybenzoate, and Sesamex were higher in patients with chronic kidney disease complications; Asp Glu Trp, Trp Met Met were higher in patients with type 2 diabetes mellitus and ischemic heart disease compared to those normal subjects without risk. Random forest produced a consistently higher accuracy of more than 70% in the prediction for all the comparison groups. Pathways perturbated and biomarkers differentially regulated in individuals with risks or with the existing conditions of type 2 diabetes mellitus and its complications of ischemic heart disease and chronic kidney disease were identified using time-of-flight liquid chromatography coupled to mass spectrometry.

Conclusion: Metabolomics is a new emerging field that provides comprehensive phenotypic information on the disease and drug response of a patient. It serves as a potential comprehensive therapeutic drug monitoring approach to be adopted in the near future for pharmaceutical care.

Keywords: Type 2 diabetes mellitus (T2DM), LCMS-QTOF, metabolomics, ischemic heart diseases (IHD), chronic kidney diseases (CKD)

Introduction

The occurrence of diabetes mellitus (DM) among the adult population is rising globally.[1,2] According to the Diabetes Atlas from the International Diabetes Federation, a total of 415 million adults had diabetes globally in 2014, and there will be approximately 780 million people suffering from diabetes, by 2045.[3] Type 2 diabetes mellitus (T2DM) has become a pandemic globally and the complications and mortality rates will continue to rise due to a lack of precise strategies for diagnosis and treatment.

The number of patients with T2DM that had progressed toward different complications had increased. The complications include both macrovascular and microvascular disorders which include diabetic nephropathy, diabetic peripheral neuropathy, diabetic retinopathy, and ischemic heart disease (IHD).[2] Nephropathy leading to chronic kidney failure (CKD) and IHD are the most common diabetic-related complications among the adult population.[4]

As the morbidity and mortality due to diabetes are projected to increase due to the increasing elderly population and sedentary lifestyles, identifying patients with a high risk of T2DM and the complications at an early stage are important strategies based on precision health. Conventional clinical and biochemical markers, such as body mass index (BMI), fasting plasma glucose, oral glucose tolerance test, and glycated hemoglobin, are well-established predictors to monitor the glycemic status in diabetic patients but remain imperfect in providing

Teh Lay Kek[1,4]
Mohd Salleh Rofiee[1,3]
Rohana Abdul Ghani[2]
Nurul Aqmar Mohd Nor Hazalin[1,4]
Mohd Zaki Salleh[1,4]

[1]*Integrative Pharmacogenomics Institute (iPROMISE), Universiti Teknologi MARA Selangor Branch, Selangor, Malaysia*
[2]*Universiti Teknologi MARA Selangor Branch, Faculty of Medicine, Selangor, Malaysia*
[3]*Universiti Teknologi MARA Selangor Branch, Faculty of Health Sciences, Selangor, Malaysia*
[4]*Universiti Teknologi MARA Selangor Branch, Faculty of Pharmacy, Selangor, Malaysia*

Corresponding authors:
Mohd Zaki Salleh
✉ zakisalleh@uitm.edu.my;
Rohana Abdul Ghani
✉ rohana1773@uitm.edu.my

Received: February 23, 2023
Revision Requested: March 29, 2023
Last Revision Received: April 9, 2023
Accepted: April 17, 2023
Publication Date: July 4, 2023

Cite this article as: Lay Kek T, Salleh Rofiee M, Abdul Ghani R, Aqmar Mohd Nor Hazalin N, Zaki Salleh M. Metabolite biomarkers and predictive model analysis for patients with type 2 diabetes mellitus with and without complications. *Endocrinol Res Pract.* 2023;27(3):135-147.

DOI: 10.5152/erp.2023.23224

clues with respect to pre-diabetes and the development of complications due to diabetes. Novel approaches are important to advance the understanding of the mechanisms of diabetes development and to identify more precise and early biomarkers.

Metabolomics is a global approach in studying biological systems that has been used for the identification and quantification of metabolites in biological samples. This approach offers a new alternative way of identifying novel biomarkers by evaluating the large numbers of metabolites that are substrates and products in metabolic pathways.[4,5] Recent metabolomics studies have suggested that certain metabolites and metabolite classes may be associated with the risk of obesity, insulin resistance, and type 2 diabetes.[6]

In this study, we used an untargeted metabolomics platform with quadrupole time-of-flight liquid chromatography coupled to mass spectrometry (QTOF-LC/MS) to profile the metabolite compounds in the serum samples and investigate the differences in the metabolome between patients with T2DM complicated with either IHD or CKD vs. those without T2DM, IHD, CKD, and non-diabetic. We aimed to identify the differentially expressed metabolites as potential diabetes-associated metabolite biomarkers that identify individuals with and without diabetes. In addition, we aimed to identify the metabolic pathways associated with T2DM and its complications. These findings will help us to better understand the development of diabetes and could assist in identifying new molecular targets for the treatment of the disease.

Material and Methods

Study Subject

The protocol of the study was reviewed and approved by the Human Research Ethics Committee of of Universiti Teknologi MARA (UiTM) (Protocol Number: REC/377/16; Date: 22 December 2016). The protocol followed good clinical practice and the Declaration of Helsinki strictly. All the subjects were explained about the study objectives and procedures before participation. Blood samples were obtained after written informed consent was obtained.

A total of 64 subjects composed of 16 subjects without T2DM, 16 patients with T2DM, 16 patients with T2DM and IHD, and 16 patients with T2DM and CKD were recruited from the UiTM Specialist Centre, Sungai Buloh Selangor. The total number of subjects was estimated based on the assumptions of a medium effect size of 0.5, SD of 1.5, aim for a statistical power of 80%, and significant level at .05.

MAIN POINTS

- Lipids and amino acid pathways were differentially regulated in type 2 diabetes mellitus (T2DM) patients compared to normal subjects and patients with complications.
- Hydroxyprolyl-leucine and *N*-palmitoyl threonine were higher in patients.
- 4,4′-Thiobis-2-butanone, geranyl-hydroxybenzoate, and Sesamex were higher in patients with chronic kidney disease complications.
- Asp Glu Trp, Trp Met Met were higher in patients with T2DM and ischemic heart disease compared to those normal subjects without risk.
- Random forest produced a consistently higher accuracy of more than 70% in the prediction of T2DM and its complication.

The subjects defined as normal comprised of 2 groups. One group is healthy subjects (n = 8) without family risk and have no symptoms of T2DM, age > 30 years with BMI < 23 kg/m^2. Another group is healthy, non-T2DM subjects with a family history of T2DM; history of gestational DM; metabolic syndrome (BMI > 23 kg/m^2) (n = 8). The healthy subjects were considered normal and were excluded if they were diagnosed with T2DM and underlying malignancy.

Patients diagnosed with T2DM were on anti-diabetic therapy within the past 3 months. Patients with a previous history of ischemic heart disease, including any cardiac or coronary intervention, acute coronary syndrome, or myocardial infarction were classified as patients with T2DM and IHD. Patients with proteinuria or abnormal estimated glomerular filtration rate of between 30 and 60 mL/min/1.73 m^2 were labeled as T2DM and CKD. The patients were excluded if they had acute illness, recent hospital admission within 6 weeks, and underlying malignancy.

Sample Preparation

The samples were frozen immediately after collection and thawed just prior to preparations to minimize metabolite degradation. Sample purification was carried out on ice using a modified protein precipitation protocol by Wang et al.[7] Deproteinization was performed by adding 450 μL cold methanol and 150 μL cold deionized water into 150 μL of each serum sample. The samples were vortexed at maximum speed for 30 seconds and centrifuged at 14700 × *g* for 10 minutes at 4°C. Subsequently, 650 μL of the supernatant was transferred into a new microcentrifuge tube, and the deproteinization steps were repeated once. A total of 1000 μL of the supernatant from the 2 deproteinization steps were dried using a vacuum concentrator (5301, Brinkmann, Eppendorf). The dried samples were stored at −80°C until analysis.

Untargeted Metabolomics Using Quadrupole Time-of-Flight Liquid Chromatography Coupled to Mass Spectrometry

The vacuum-dried samples were reconstituted with 30 μL of the mobile phase (A 50% dH_2O: B 50% ACN). Two microliters of the samples were injected and analyzed by LC/MS-QTOF (6520 Agilent Technologies, Santa Clara, Calif, USA) using a ZORBAX Eclipse Plus C18 column (100 mm × 2.1 mm ×1.8 μm, Agilent Technologies) maintained at 40°C. The system was operated with a flow rate of 0.25 mL/min with mobile phase A (water with 0.1% formic acid) and mobile phase B (acetonitrile with 0.1% formic acid). A linear gradient over 18 minutes from 5% to 95% of mobile phase B was used, with 95% of mobile phase B maintained over 12 minutes of post-run. Electrospray ionization (ESI) source settings were set as follows: V Cap 4000 V, skimmer voltage 68 V, and fragment 215 V. The nebulizer was set at 20 psi, and the flow rate and the temperature of the nitrogen drying gas were maintained at 12 L/min and 350°C, respectively. Data were collected by a full scan positive ESI mode from 50 to 1000 *m/z*. During the analysis, 2 reference masses of 121.0509 *m/z* (C5H4N4) and 922.0098 *m/z* (C18H18O6N3P3F24) were continuously injected to allow consistent, accurate mass correction.

The accuracy and reproducibility of the analytical method were measured by injecting quality control (QC) samples for each batch of the sample analyses. The QC samples were prepared by pooling the aliquots of all the samples analyzed. One QC sample was analyzed for each batch of serum samples. The QC samples were injected at the beginning, middle, and end of the run to ensure the system performance and assay reproducibility. Evaluation of the QC was done

by calculating the distribution of relative standard deviation (RSD) of metabolites that are consistently present in 80% of the pooled samples. Metabolites within the range of *m/z* and retention time covered in the analysis were chosen to represent the QC. The data of QC samples were compared against the data of subject's samples by treating them as a separate group and processing them using the same parameters selected for processing the whole sample set. In addition, a Principal Component Analysis (PCA plot) was generated using data from QC samples and subject samples

Data Processing

Agilent MassHunter Qualitative Software from Agilent Technologies was used for metabolite extraction and data mining. For the positive ionization mode, the adducts used were H+ and Na+. Several parameters were set to select the molecular features. The chromatogram and spectra were observed to determine the reproducibility of the results. Metabolites detected within 0.100-18.000 minutes of the analysis, within 50-1000 *m/z* were identified. The metabolites data were then processed using "Find Compound" by "Search Database" algorithm parameters. All the data from (.d) files were converted to the (.cef) file using DA Reprocessor (Agilent Technologies) software and then further analyzed using Mass Profiler Professional (MPP) (Agilent Technologies) software. The data were subjected to normalization, filtration, and recursion analysis.

Statistical Analysis

The data were filtered using "Filter by Frequency" and "Filter by Flags" analysis, which was set to 50% to ensure the identified compounds were detected in at least 50% of all the technical replicates of the biological samples. Filtering was further done by analysis of variance to select compounds that were significantly differentiated between the subjects without T2DM and patients with T2DM only, and patients with T2DM and complications. The compounds with *P*-value and fold change (FC) cut-off scores of .05 and 2.0, respectively, were filtered to determine the differentially expressed metabolites. All the metabolites were identified using the ID browser packed with the Metlin database of the MPP software. Kyoto Encyclopedia of Genes and Genomes and Human Metabolome Database and PubChem were used to confirm the identities of the metabolites.[8,9] The metabolites were then transferred into the entity list and exported to visualize using MetaboAnalyst web-based tools. The data were re-examined by recursion analysis. The identified compounds from recursion analysis were subjected again for filtering using "Filter by Frequency" and "Filter by Flags" analysis. The differential analysis was done using MPP software. All the statistical analyses were done using MetaboAnalyst 5.0 (https://www.metaboanalyst.ca/) (McGill University, Quebec, Canada).

Pathways Analyses

The PCA and orthogonal partial least square discriminant analysis (OPLS-DA) were performed to illustrate the different metabolic profiles among the 4 groups of subjects. Additional comparisons were conducted between normal subjects without family risk to other patients with T2DM and other complication, normal subjects with family risks and other patients' groups.

The metabolic pathways were determined using MetaboAnalyst 5.0 web-based tool (https://www.metaboanalyst.ca/). MetaboAnalyst was also used for potential biomarker identification. Metabolites from the biological pathways were assessed for their potential as biomarkers for T2DM and its complications.

Biomarker Analysis

Biomarker analyses were performed using MetaboAnalyst 5.0 (https://www.metaboanalyst.ca/). Five steps were conducted which included data uploading, data processing, biomarker selection, performance evaluation, and model creation. The data were first subjected to receiver operation characteristic (ROC) curve analysis for individual biomarkers, followed by manually selecting a subset of features/samples for ROC analysis and the third part of the analysis was done using Multivariate Exploratory ROC Analysis.

Receiver Operation Characteristic Curve Analysis for Individual Biomarkers

Biomarkers were determined using the Classical univariate ROC curve analyses. Features were ranked based on area under ROC curve, T-statistics or Log2 FC. The 95% confidence interval was calculated using 500 bootstrapping.

Manually Select a Subset of Features/Samples for Receiver Operation Characteristic Analysis

Biomarkers with the area under the curve (AUC) more than 0.8 and top 5 were manually selected to create biomarker models using 4 algorithms which were linear support vector machine (SVM), partial least squares discriminant analysis (PLS-DA), random forest, and logistic regression. Twenty-five percent of the samples for each group were held out as a subset of samples for extra validation purposes. In order to produce a smooth ROC curve, 100 cross-validations (CV) were performed and the results were averaged to generate the plot. The models were predicted as class probabilities of each sample across the 100 cross-validations.

Multivariate Exploratory Receiver Operation Characteristic Analysis

The ROC curves were generated by Monte-Carlo cross-validation (MCCV) using balanced sub-sampling. In each MCCV, two-thirds of the samples were used to evaluate the feature importance. The top 2, 3, 5, 10 ...100 (max) important features were then used to build classification models which were validated on one-third of the samples that were left out. The procedures were repeated multiple times to calculate the performance and confidence interval of each model. The algorithms used included linear SVM, PLS-DA algorithm, and random forest.

Results

Demographic and Characteristics of Participants

The age ranges for the subjects were from 30 to 68 years old at the time of sample collection. There were 33 males (51.6%) and 31 females (48.4%). The subjects comprised of 55 Malay (85.9%), 2 Indian (3.2%), and 7 Chinese (10.9%). The demographic data of the subjects are presented in Table 1.

Metabolomics Differences Between Studied Groups

Metabolite Profiling: Total ion chromatograms for the analyzed serum samples were assessed to ensure the reproducibility of the replicates. The metabolites were resolved over 18 minutes of run time. A good QC sample was visualized as the QC clustered within the patients' samples with an acceptable RSD of 20% for mass and 2% for retention time. This indicates that the analysis was stable and reliable (Supplementary Table S1).

Table 1. Demographic Data for the Subjects

Variables	Normal, Without T2DM	T2DM Patients	T2DM Patients with IHD	T2DM Patients with CKD
N	16 (25%)	16 (25%)	16 (25%)	16 (25%)
Ethnicity				
Malay	16 (100%)	16 (100%)	10 (62.5%)	14 (87.5%)
Chinese	0 (0%)	0 (0%)	1 (6.25%)	2 (12.5%)
Indian	0 (0%)	0 (0%)	5 (31.25%)	0 (0%)
Sex				
Female	13 (81.25%)	6 (37.5%)	4 (25%)	8 (50%)
Male	3 (18.75%)	10 (62.5%)	12 (75%)	8 (50%)
Age (years) Mean ± SD	34.00 ± 4.00	53.00 ± 10.00	57.00 ± 8.00	62.00 ± 6.00
HbA1C (%) Mean ± SD	5.00 ± 0.30	7.00 ± 1.00	7.00 ± 1.00	7.00 ± 1.00
BMI (kg/m^2)				
Mean ± SD	26.00 ± 5.00	31.00 ± 5.00	28.00 ± 6.00	31.00 ± 4.00

BMI, body mass index; CKD, chronic kidney disease; HbA1C, glycated hemoglobin; IHD, ischemic heart disease; T2DM, type 2 diabetes mellitus.

Identification of Differentially Expressed Metabolites

The flow of analysis and the numbers of metabolites identified for each comparison group are shown in Supplementary Table S2, respectively.

Multivariate Analysis

Multivariate exploratory analyses (unsupervised PCA and supervised Orthogonal PLS-DA) were performed to demonstrate the separation between the normal (non-T2DM) and patient groups. Ortho PLS-DA plots show the clusters and separation of metabolites between the normal and patients with T2DM only, IHD, and CKD groups (Figure 1).

Pathway Analyses

The differentially expressed metabolites identified were analyzed using MetaboAnalyst 5.0 to determine the metabolic pathways differentially altered between the normal and T2DM with and without complications and patients with T2DM and those with complications (Figure 2 and Supplementary Table S3).

Eleven metabolism pathways were identified to be differentially regulated between the 5 groups of subjects compared to the other 2 groups do not show any pathways which were differentiated. Four pathways (caffeine metabolism, glycerophospholipid metabolism, glycosylph osphatidylinositol (GPI)-anchor biosynthesis, and porphyrin and chlorophyll metabolism) were significantly differentiated only between patients with T2DM and patients with T2DM and complications.

Drug metabolism—cytochrome P450 was identified in all 5 patients' groups but only reach significant differences between patients with T2DM and patients with T2DM and complications.

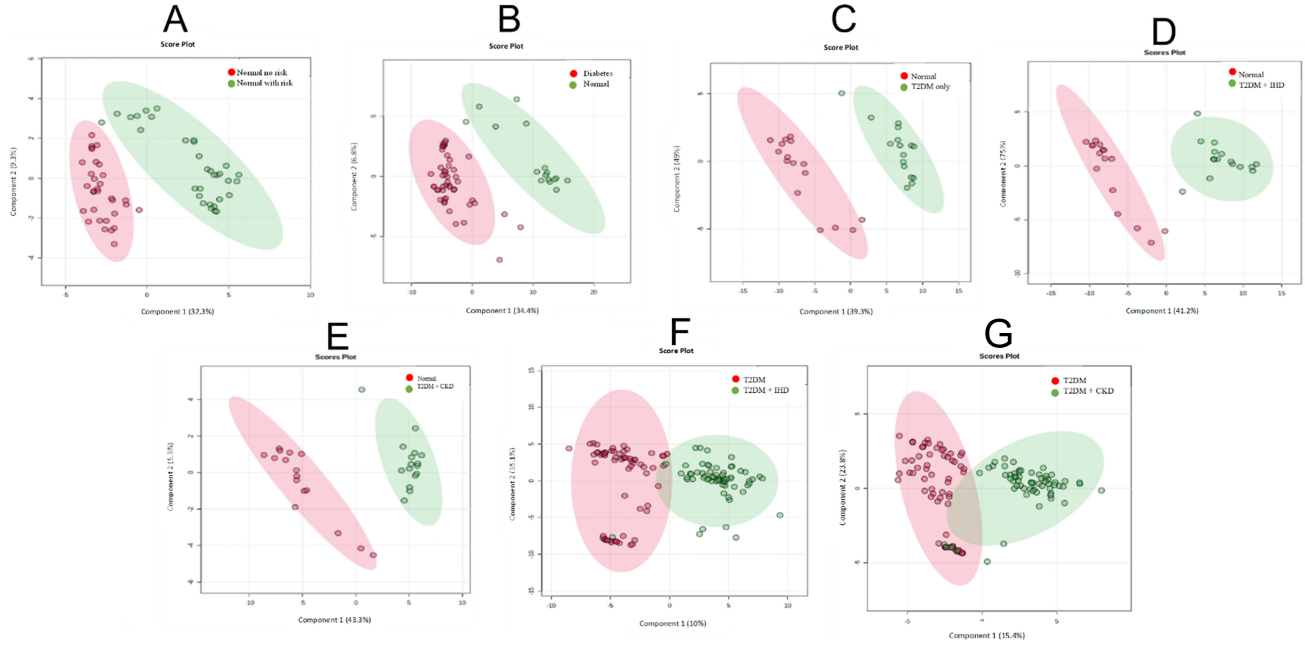
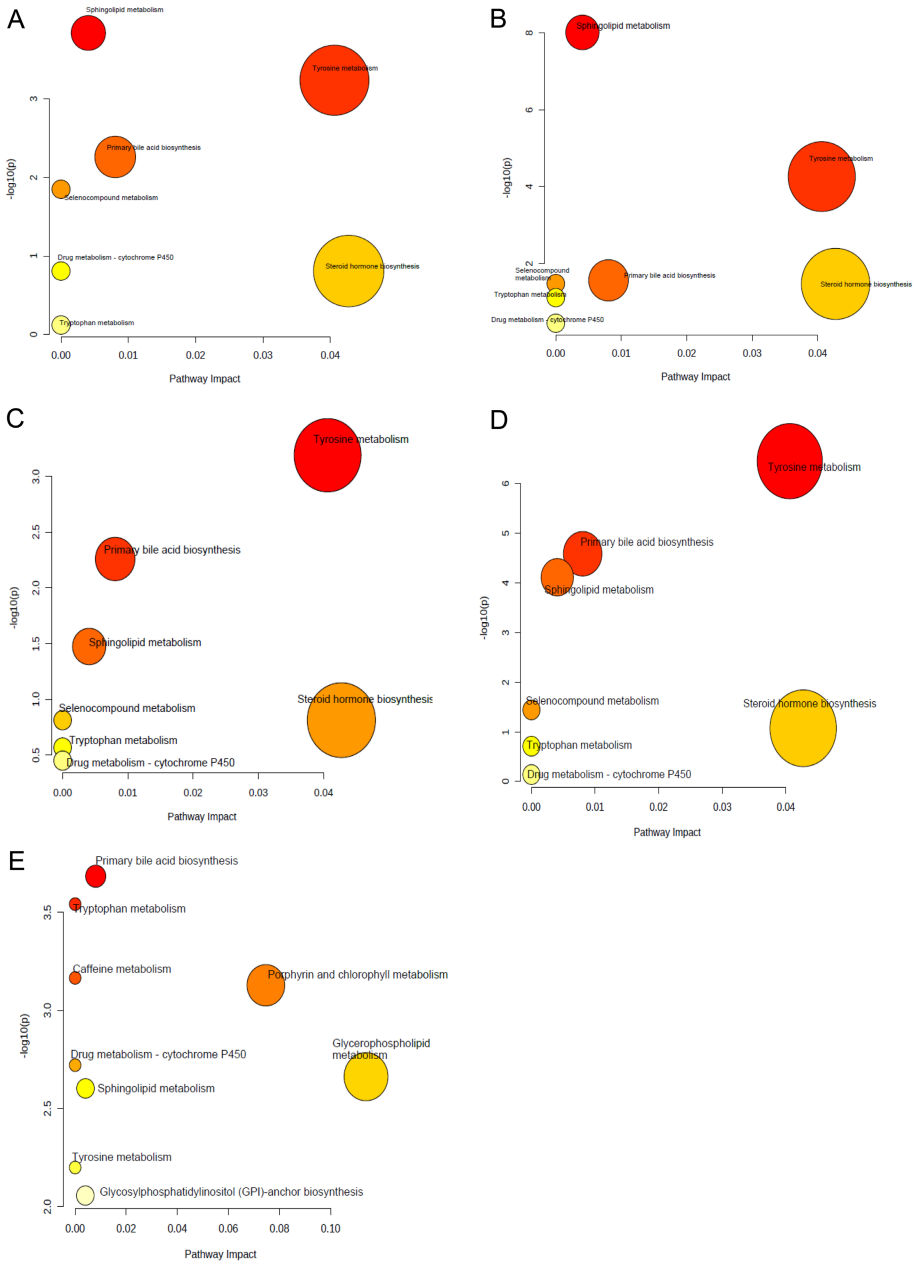


Figure 1. Orthogonal PLSDA (OPLS-DA) analysis between the normal and patient groups. (A) Normal without risk vs. with risk; (B) normal vs. T2DM and T2DM with complications of IHD and CKD; (C) normal vs. T2DM only; (D) normal vs. T2DM with IHD; (E) normal vs. T2DM with CKD; (F) patients with T2DM vs. patients with T2DM with IHD; and (G) patients with T2DM vs. patients with T2DM with CKD. CKD, chronic kidney disease; IHD, ischemic heart disease; OPLS-DA, orthogonal partial least square discriminant analysis; T2DM, type 2 diabetes mellitus.



Pathways	Comparison groups (*significant or # insignificant)
Caffeine metabolism	*5
Drug metabolism - cytochrome P450	#1, #2, #3, #4, *5
Glycerophospholipid metabolism	*5
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	*5
Porphyrin and chlorophyll metabolism	*5
Primary bile acid biosynthesis	*1, *2, *3, *4, *5
Selenocompound metabolism	*1, *2, #3, *4
Sphingolipid metabolism	*1, *2, *3, *4, *5
Steroid hormone biosynthesis	#1, *3, #4,
Tryptophan metabolism	#1, #2, #3, #4, *5
Tyrosine metabolism	*1, *2, *3, *4, *5

Figure 2. Pathways differentially expressed in different comparison groups. *Significant; #Not significant. Groups: 1. Normal vs T2DM; 2. Normal vs T2DM + IHD; 3. Normal vs T2DM + CKD; 4. Normal vs Patients T2DM + IHD + CKD; 5. Patients T2DM vs Patients T2DM +IHD + CKD

Primary bile acid biosynthesis, sphingolipid metabolism, and tyrosine metabolism were significantly differentially regulated between all 5 groups of subjects.

Selenocompound metabolism was significantly differentiated between 3 groups of subjects which are normal and patients with T2DM; normal and patients with T2DM and IHD; and normal and patients with T2DM and IHD and CKD. It was detected in a group of subjects which were normal and patients with T2DM and CKD but did not reach statistical significance.

Steroid hormone biosynthesis was detected in 3 comparison groups which are (i) normal and patients with T2DM; (ii) normal and patients with T2DM and CKD; and (iii) normal and patients with T2DM and IHD and CKD but was significant only in normal and patients with T2DM and CKD.

Receiver Operation Characteristic Curve Analysis for Individual Biomarkers

Five metabolites with the highest AUC scores were identified, differentiating normal with and without risk of T2DM, patients with T2DM, patients with T2DM and IHD, and patients with T2DM and CKD (Table 2).

The metabolites which were higher in the normal subjects with family risks are PE (17:1 (9Z)/0:0), LysoPE (0:0/20:0), and 9Z,12Z,15E-octa decatrienoic acid. Two other metabolites that were higher in the normal subject without risk for T2DM are (6S)-dehydrovomifoliol and 5,5-Dimethyl-1-pyrroline N-oxide (DMPO) (Table 2).

Comparing normal subjects with patients with T2DM and complications, L-beta-aspartyl-L-phenylalanine, 6-keto-PGF1, 2-methyl-3 -phenyl-2-propenal, Bn-NCC-1, and 6,9-heptadecadiynoic acid were higher among the normal subjects (Table 2).

All the 5 top listed metabolites which include L-beta-aspartyl-L-ph enylalanine, Bn-NCC-1, 2-methyl-3-phenyl-2-propenal, 6-keto-PGF1, and 6,9-heptadecadiynoic acid were more abundant among the healthy subjects compared to patients with T2DM without complications (Table 2).

C16 sphinganine, eplerenone, and phytosphingosine were higher among the patients with T2DM and IHD compared to the healthy normal; while Bn-NCC-1 and 2-methyl-3-phenyl-2-propenal were higher among the healthy subjects (Table 2).

All the 5 metabolites including the L-beta-aspartyl-L-phenylalanine, 6-keto-PGF1, PI [14:1 (9Z)/18:4 (6Z, 9Z,12Z, 15Z)], 2-methyl-3-phen yl-2-propenal, and Gly Val Asn were higher among the normal and lower than the patients with T2DM and CKD (Table 2).

The metabolites that differentiated patients with T2DM and patients with T2DM and IHD included 5-chola-7,9 (11)-dien-24-oic acid, *N*-acryloylglycine, 4,8 dimethylnonanoyl carnitine, cis-caryophyllene, and 9,10-epoxy-18-hydroxystearate. These metabolites have an AUC of more than 0.7 to 0.8. Patients with T2DM were differentiated from patients with T2DM and CKD by metabolites which were hydroxyprolyl-leucine, 4,4′-Thiobis-2-butanone, geranyl-hydroxybenzoate, *N*-palmitoyl threonine, and Sesamex.

Manually Select a Subset of Features/Samples for Receiver Operation Characteristic Analysis

Prediction models were developed for each group using these top 5 biomarkers listed in Table 2. All 4 algorithms were used to build and evaluate the AUC and accuracy of the prediction models.

Area under the curve models developed using different algorithms achieved more than 0.8 for all comparison groups. However, the differences in AUC were not remarkable when the algorithms used were linear SVM, PLS-DA, and random forest. Random Forest produced the highest values of AUC for all the comparison groups with 0.88 as the lowest value achieved in comparing patients with T2DM and patients with T2DM and CKD. The AUC values calculated by logistic regression saw the most variabilities, especially between the patients with T2DM and patients with T2DM and CKD. The coefficient variation calculated for the AUC for each model was less than 10%; with 0.82% for the model differentiating normal without risk from normal with risk; the highest CV was for the model predicting patients with T2DM vs. T2DM + IHD (6.87%). The accuracy of the prediction models was high too. The lowest percentage for accuracy was the prediction model for patients with T2DM and T2DM and IHD, 69% using logistic regression and 10.91% of the CV. The AUC and accuracy for predicting hold-out data were the lowest at 0.59 and 0.63%, respectively for the model predicting patients with T2DM and patients with T2DM and CKD. The model predicting patients with T2DM and patients with T2DM and IHD achieved AUC and accuracy of 0.77 and 0.78%, respectively using logistic regression (Table 3).

Multivariate Exploratory Receiver Operation Characteristic Analysis

Three different algorithms which are linear SVM, PLS-DA, and random forest were used to model the variables to predict different groups. The metabolites used by each algorithm were different for various groups.

The AUC and accuracies for the 7 groups were compared. The modeling for groups comparing the normal without risk and normal with risk and group comparing patients with T2DM and patients with T2DM and CKD achieved good AUC and accuracies with 2 and 3 metabolites as variables. While for the other 5 groups, 5 metabolites were used to produce a prediction model with good AUC and accuracy (Table 4). The metabolites used for the development of each model are made available in Appendix A.

Discussion

Globally, DM is on the increasing trend in causing morbidity and mortality. About half a million kidney disease deaths and 20% of cardiovascular deaths were caused by increased blood glucose levels.[10] Besides, DM is associated with macro and microvascular complications, including nephropathy, retinopathy, IHDs,[11,12] and stroke.[13]

In parallel with one of the aims of precision health, which focuses on the rapid and accurate detection of pathologies, efforts to search for ideal biomarkers and therapy continue and remain in the mainstream of medical science. An ideal biomarker helps to predict the risk of disease to monitor the progression of the disease and the patient's response to treatment. The biomarker is a useful clinical prediction tool which provides insight into the biological processes that result in the onset of DM or its complications and might also be a surrogate biomarker of the underlying disease process. In addition to understanding the disease process, we aimed to identify potential biomarkers that can detect subjects at risk of developing DM despite the presence or absence of other known conventional risk factors.

We opted for metabolomics to profile and identify various metabolites as potential biomarkers simultaneously using LCMS-QTOF. The

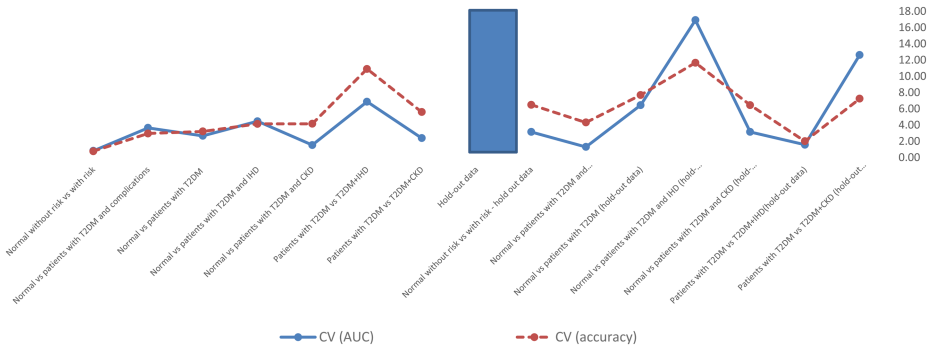
Table 2. Five Top Biomarkers Differentiated Among Normal Subjects and Patients T2DM With and Without Complications

Metabolite	AUC	*T*-Test	Log2 FC	Levels of Metabolites	
Normal without risk vs. normal with risk				**Normal (−) risk**	**Normal (+) risk**
PE (17:1(9Z)/0:0)	0.908 (0.819-0.978)	2.196E-11	−1.4376	Low	High
LysoPE (0:0/20:0)	0.823 (0.696-0.922)	1.6404E-4	−0.75239	Low	High
(6S)-dehydrovomifoliol	0.813 (0.676-0.911)	6.8818E-7	−0.903	Low	High
9Z,12Z,15E-octadecatrienoic acid	0.812 (0.710-0.897)	2.3387E-7	−1.2827	Low	High
DMPO	0.800 (0.698-0.888)	1.0411E-6	1.4271	High	Low
Normal vs. patients with T2DM and IHD and CKD				T2DM + IHD + CKD	Normal
L-Beta-aspartyl-L-phenylalanine	0.926 (0.827-0.991)	9.3164E-6	−0.98459	Low	High
6-Keto-PGF1	0.924 (0.829-0.980)	2.21E-11	−1.6784	Low	High
2-Methyl-3-phenyl-2-propenal	0.913 (0.832-0.975)	3.3723E-4	−0.64269	Low	High
Bn-NCC-1	0.914 (0.809-1.000)	4.2186E-13	−1.9428	Low	High
6,9-Heptadecadiynoic acid	0.905 (0.816-0.981)	3.2131E-9	−1.5034	Low	High
Normal vs. patients with T2DM only				Normal	T2DM only
L-Beta-aspartyl-L-phenylalanine	0.961 (0.873-1.000)	2.5874E-7	1.516	High	Low
Bn-NCC-1	0.938 (0.839-1.000)	4.8004E-9	2.072	High	Low
2-Methyl-3-phenyl-2-propenal	0.938 (0.822-0.992)	2.4586E-4	0.75583	High	Low
6-Keto-PGF1	0.938 (0.832-1.000)	1.1155E-6	1.4391	High	Low
6,9-Heptadecadiynoic acid	0.928 (0.815-1.000)	6.8653E-7	1.4782	High	Low
Normal vs. patients with T2DM + IHD				Normal	Patient T2DM + IHD
C16 Sphinganine	0.938 (0.844-1.000)	7.3188E-11	−2.4962	Low	High
Eplerenone	0.938 (0.844-1.000)	5.3541E-11	−2.3836	Low	High
Bn-NCC-1	0.930 (0.828-1.000)	8.8457E-8	1.9038	High	Low
2-Methyl-3-phenyl-2-propenal	0.914 (0.781-0.984)	0.0079152	0.36111	High	Low
Phytosphingosine	0.906 (0.812-1.000)	9.6428E-9	−2.3996	Low	High
Normal vs. patients with T2DM + CKD				Normal	patients with T2DM + CKD
L-Beta-aspartyl-L-phenylalanine	0.948 (0.849-1.000)	1.0794E-4	1.0015	High	Low
6-Keto-PGF1	0.951 (0.855-1.000)	5.0385E-9	1.8359	High	Low
PI [14:1(9Z)/18:4 (6Z, 9Z,12Z,15Z)]	0.906 (0.812-1.000)	7.002E-9	2.2378	High	Low
2-Methyl-3-phenyl-2-propenal	0.914 (0.752-1.000)	1.1724E-4	0.84287	High	Low
Gly Val Asn	0.902 (0.780-0.992)	4.17E-6	1.523	High	Low
T2DM vs. T2DM + IHD				Patients with T2DM	Patients with T2DM and IHD
5-Chola-7,9(11)-dien-24-oic Acid	0.798 (0.719-0.866)	1.2596E-9	−0.7803	Low	High
N-Acryloylglycine	0.793 (0.711-0.861)	5.061E-6	−0.38269	Low	High
4,8 dimethylnonanoyl carnitine	0.779 (0.697-0.851)	1.3863E-4	−0.38768	Low	High
cis-Caryophyllene	0.772 (0.691-0.846)	2.4993E-6	−0.68556	Low	High
9,10-Epoxy-18-hydroxystearate	0.756 (0.688-0.840)	1.8658E-4	−0.40239	Low	High
T2DM vs. T2DM+CKD				Patients with T2DM	Patients with T2DM and IHD
Hydroxyprolyl-leucine	0.810 (0.728-0.873)	6.8133E-8	−0.77378	Low	High
4,4′-Thiobis-2-butanone	0.792 (0.707-0.870)	2.3626E-6	−0.61298	Low	High
Geranyl-hydroxybenzoate	0.771 (0.688-0.846)	1.1051E-8	−1.0001	Low	High
N-palmitoyl threonine	0.770 (0.683-0.843)	0.0015492	−0.33026	Low	High
Sesamex	0.768 (0.692-0.832)	5.583E-8	−1.0698	Low	High

AUC, area under the curve; CKD, chronic kidney disease; FC, fold change; IHD, ischemic heart disease; T2DM, type 2 diabetes mellitus.

use of the metabolomics approach was justified as DM is a systemic disorder influenced by lifestyles and diet which affect the metabolism of cells. Metabolites are the end products of cellular activity as an interaction between biological and external factors, which include diet and the environment.[14] As observed in our study, tyrosine and tryptophan metabolism were differentially regulated when we compared metabolism between (i) normal subjects vs. patients with T2DM; (ii) normal subjects vs. patients with T2DM and IHD; (iii) normal subjects vs. patients with T2DM and CKD; (iv) normal subjects vs. patients with T2DM and IHD and CKD; and (v) patients with T2DM vs. patients with T2DM and IHD and CKD. Primary bile acid biosynthesis, sphingolipid metabolism, and tyrosine metabolism were significantly differentially regulated between all 5 groups of subjects, while glycerophospholipid metabolism and GPI-anchor biosynthesis

Table 3. Efficacy of Prediction Model Based on 5 Top Biomarkers

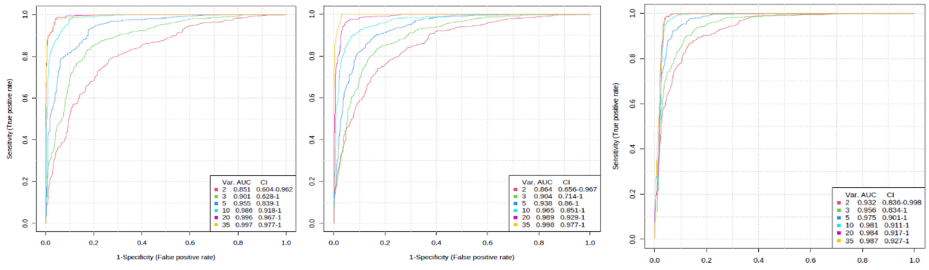
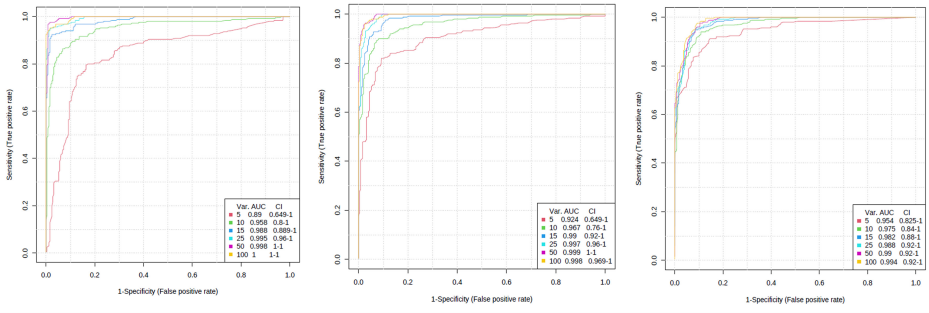
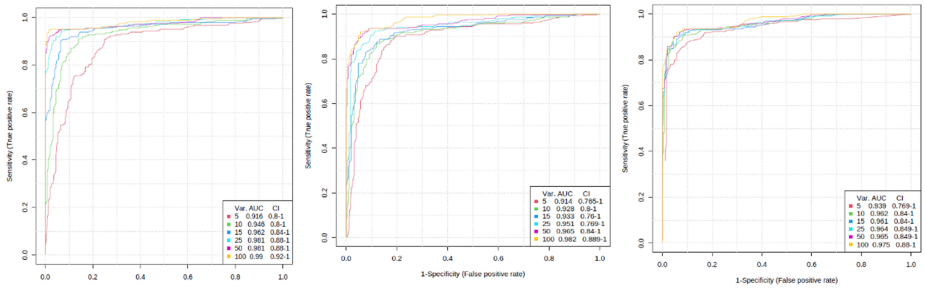


	AUC					Accuracy Predictive Model				
Groups	Linear SVM	PLS-DA	Random forest	Logistic regression	CV (AUC)	Linear SVM	PLS-DA	Random forest	Logistic regression	CV (accuracy)
Normal without risk vs. with risk	0.98	0.99	0.99	0.97	0.82	0.96	0.95	0.94	0.94	0.78
Normal vs. patients with T2DM and IHD + CKD	0.94	0.94	0.96	0.88	3.66	0.89	0.89	0.90	0.84	2.97
Normal vs. patients with T2DM only	0.97	0.96	0.97	0.92	2.67	0.89	0.91	0.90	0.85	3.21
Normal vs. patients with T2DM and IHD	0.99	0.98	1.00	0.90	4.46	0.87	0.88	0.94	0.86	4.14
Normal vs. patients with T2DM and CKD	1.00	1.00	1.00	0.97	1.54	0.93	0.95	0.93	0.86	4.15
Patients with T2DM vs. T2DM + IHD	0.82	0.82	0.93	0.81	6.87	0.71	0.70	0.86	0.69	10.91
Patients with T2DM vs. T2DM + CKD	0.83	0.85	0.88	0.84	2.38	0.77	0.76	0.85	0.75	5.60
Hold-out data										
Normal without risk vs. with risk - hold out data	1.00	1.00	1.00	0.94	3.15	0.94	1.00	0.88	0.88	6.49
Normal vs. patients with T2DM + IHD + CKD (hold-out data)	0.92	0.92	0.94	0.94	1.31	0.81	0.88	0.88	0.81	4.31
Normal vs. patients with T2DM only (hold-out data)	1.00	1.00	1.00	0.88	6.45	0.88	1.00	1.00	0.88	7.70
Normal vs. patients with T2DM and IHD (hold-out data)	1.00	1.00	1.00	0.69	16.92	0.88	0.88	1.00	0.75	11.66
Normal vs. patients with T2DM and CKD (hold-out data)	1.00	1.00	0.94	1.00	3.15	1.00	1.00	0.88	1.00	6.45
Patients with T2DM vs. T2DM + IHD (hold-out data)	0.80	0.78	0.79	0.77	1.58	0.78	0.78	0.75	0.78	2.00
Patients with T2DM vs. T2DM + CKD (hold-out data)	0.59	0.77	0.77	0.79	12.63	0.63	0.63	0.72	0.63	7.25

AUC, area under the curve; CKD, chronic kidney disease; CV, cross-validation; IHD, ischemic heart disease; PLS-DA, partial least square discriminant analysis; SVM, support vector machine; T2DM, type 2 diabetes mellitus.

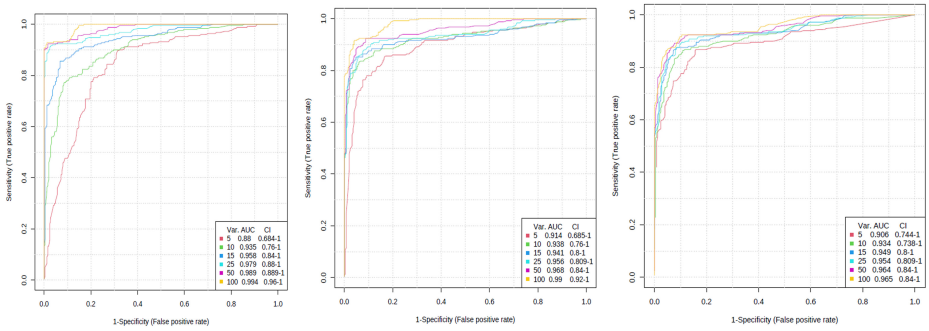
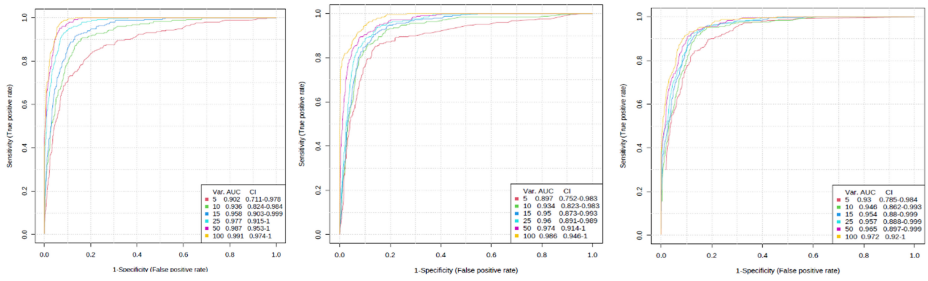
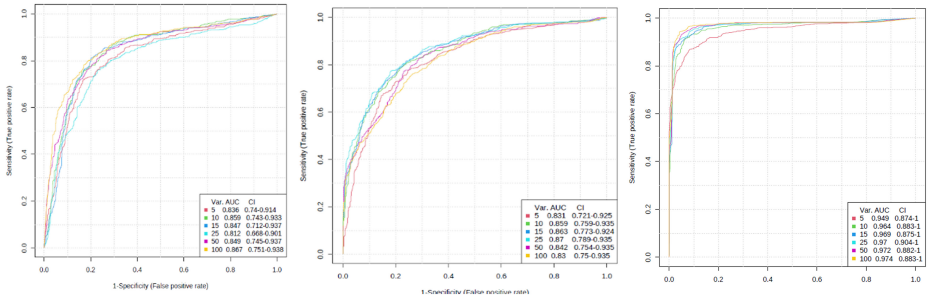
and porphyrin and chlorophyll metabolism were significantly differentiated only between patients with T2DM and patients with T2DM and complications. This seemed to be linked to similar findings by Lin et al, which showed that GPI-anchor biosynthesis was one of the discriminating metabolites between non-alcoholic fatty pancreas disease and T2DM.[15] Fatty liver is currently gaining much interest as both a complication as well as a precursor for T2DM. Nonetheless, the differentially expressed metabolism pathway should be studied in

Table 4. Comparison of AUC and Accuracies of Models Based on Different Number of Variables

Number of Variables	SVM	Predictive Accuracies (%)	PLS-DA	Predictive Accuracies (%)	Random Forest	Predictive Accuracies (%)
Normal without risk vs. normal with risk						
2	0.851 (0.604-0.962)	74.2	0.864 (0.656-0.967)	77.5	**0.932 (0.836-0.998)**	77.5
3	0.901 (0.628-1)	82.8	0.904 (0.714-1)	82.6	0.956 (0.834-1)	82.6
5	0.955 (0.839-1)	86	0.938 (0.86-1)	86.8	0.975 (0.901-1)	86.8
10	0.986 (0.918-1)	93.2	0.965 (0.851-1)	91	0.981 (0.911-1)	91
20	0.996 (0.967-1)	96.9	0.989 (0.929-1)	95.6	0.984 (0.917-1)	95.6
35	0.997 (0.977-1)	96.6	0.998 (0.977-1)	97.8	0.987 (0.927-1)	97.8
						
Normal vs. T2DM only						
5	0.916 (0.8-1)	80.8	0.914 (0.765-1)	84.2	0.939 (0.769-1)	87.2
10	0.946 (0.8-1)	88	0.928 (0.8-1)	85.6	0.962 (0.84-1)	89
15	0.962 (0.84-1)	90.6	0.933 (0.76-1)	86.4	0.961 (0.84-1)	91.2
25	0.981 (0.88-1)	93.2	0.951 (0.769-1)	90	0.964 (0.849-1)	91
50	0.981 (0.88-1)	93.4	0.965 (0.84-1)	92.2	0.965 (0.849-1)	92.2
100	0.99 (0.92-1)	93.8	0.982 (0.889-1)	91.8	0.975 (0.88-1)	92.4
						
Normal vs. patients with T2DM and IHD						
5	0.89 (0.649-1)	81.8	0.924 (0.649-1)	80.6	**0.954 (0.825-1)**	**85.4**
10	0.958 (0.8-1)	86.2	0.967 (0.76-1)	89	0.975 (0.84-1)	90.2
15	0.988 (0.889-1)	93.8	0.99 (0.92-1)	93.6	0.982 (0.88-1)	92.8
25	0.995 (0.96-1)	96.4	0.997 (0.96-1)	96.4	0.988 (0.92-1)	95
50	0.999 (1-1)	96.6	0.999 (1-1)	97.8	0.99 (0.92-1)	95.8
100	1 (1-1)	96.2	0.998 (0.969-1)	96.4	0.994 (0.92-1)	96.2
						
Normal vs. patients with T2DM and CKD						
5	0.88 (0.6-1)	80	0.914 (0.685-1)	83.2	**0.906 (0.744-1)**	**85.4**
10	0.935 (0.76-1)	87.2	0.938 (0.76-1)	86.4	0.934 (0.738-1)	86.8

(*Continued*)

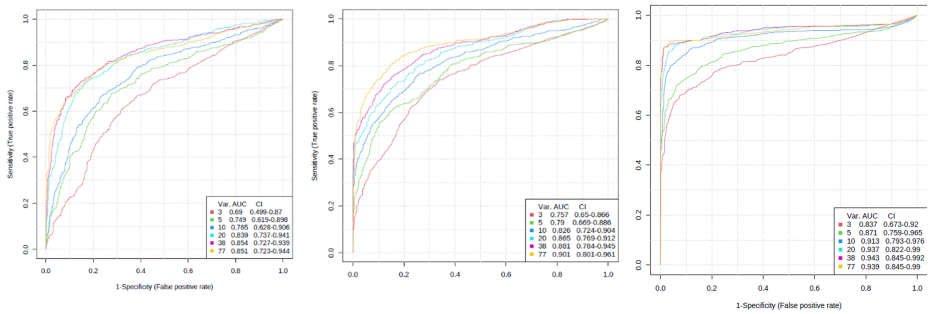
Table 4. Comparison of AUC and Accuracies of Models Based on Different Number of Variables (*Continued*)

Number of Variables	SVM	Predictive Accuracies (%)	PLS-DA	Predictive Accuracies (%)	Random Forest	Predictive Accuracies (%)
15	0.958 (0.84-1)	90.2	0.941 (0.8-1)	88.2	0.949 (0.8-1)	87.4
25	0.979 (0.88-1)	94.6	0.956 (0.809-1)	89.6	0.954 (0.809-1)	89.4
50	0.989 (0.889-1)	95.8	0.968 (0.84-1)	91.6	0.964 (0.84-1)	90.2
100	0.994 (0.96-1)	96.8	0.99 (0.92-1)	92.8	0.965 (0.84-1)	90.8
						
Normal vs. patients with T2DM and IHD and CKD						
5	0.902 (0.711-0.978)	82.8	0.897 (0.752-0.983)	88.7	**0.93 (0.785-0.984)**	87
10	0.936 (0.824-0.984)	87.2	0.934 (0.823-0.983)	90	0.946 (0.862-0.993)	88.3
15	0.958 (0.903-0.999)	89.4	0.95 (0.873-0.993)	90.1	0.954 (0.88-0.999)	89.1
25	0.977 (0.915-1)	92	0.96 (0.891-0.989)	90.9	0.957 (0.888-0.999)	89.1
50	0.987 (0.953-1)	94.1	0.974 (0.914-1)	92.1	0.965 (0.897-0.999)	90.3
100	0.991 (0.974-1)	94.6	0.986 (0.946-1)	93.8	0.972 (0.92-1)	91.3
						
Patients T2DM vs. patients with T2DM and IHD						
5	0.836 (0.74-0.914)	76.6	0.831 (0.721-0.925)	75.9	**0.949 (0.874-1)**	**89.2**
10	0.859 (0.743-0.933)	78.4	0.859 (0.759-0.935)	77.9	0.964 (0.883-1)	92.1
15	0.847 (0.712-0.937)	80.2	0.863 (0.773-0.924)	78.3	0.969 (0.875-1)	92.5
25	0.812 (0.668-0.901)	76.1	0.87 (0.789-0.935)	79.2	0.97 (0.904-1)	93.2
50	0.849 (0.745-0.937)	78.8	0.842 (0.754-0.935)	74.8	0.972 (0.882-1)	93.5
100	0.867 (0.751-0.938)	80.1	0.83 (0.75-0.935)	73.7	0.974 (0.883-1)	94.3
						
Patients T2DM vs. patients with T2DM and CKD						
3	0.7 (0.519-0.828)	63.8	0.757 (0.65-0.866)	69.4	**0.835 (0.736-0.96)**	**78.6**
5	0.749 (0.547-0.896)	68.3	0.79 (0.669-0.886)	70.5	0.884 (0.792-0.998)	83.9
10	0.777 (0.635-0.886)	70.9	0.826 (0.724-0.904)	74.7	0.916 (0.815-0.99)	87
20	0.815 (0.732-0.934)	74.4	0.865 (0.769-0.912)	77.8	0.939 (0.849-0.997)	89.8

(*Continued*)

Table 4. Comparison of AUC and Accuracies of Models Based on Different Number of Variables (*Continued*)

Number of Variables	SVM	Predictive Accuracies (%)	PLS-DA	Predictive Accuracies (%)	Random Forest	Predictive Accuracies (%)
38	0.83 (0.705-0.917)	75.6	0.881 (0.784-0.945)	80	0.943 (0.84-1)	91.3
77	0.831 (0.712-0.939)	76.1	0.901 (0.801-0.961)	82.4	0.939 (0.832-1)	92.4



AUC, area under the curve; CKD, chronic kidney disease; IHD, ischemic heart disease; PLS-DA, partial least square discriminant analysis; SVM, support vector machine; T2DM, type 2 diabetes mellitus.

more depth using in vitro or in vivo models to allow us to understand how the perturbated metabolism can be modified to help prevent getting diabetes and its complications.

Several metabolites were previously reported to be associated with the risk of T2DM, including levels of α-hydroxybutyrate[16] and branched-chain amino acids.[17-19] Amino acids such as phenylalanine, tryptophan, tyrosine, alanine, glycine, isoleucine, leucine, proline, and valine have been associated with insulin secretion and resistance leading to increased risk of type 2 diabetes.[20] Metabolites, such as amino acids, have been used to facilitate understanding, diagnosing, and predicting the occurrence of T2D[16,17,20] and glucose tolerance in pre-diabetes.[11,21,22] High leucine levels were found to increase the activity of the mTOR pathway which activates S6 kinase and results in the inhibition of insulin receptor substrates by serine phosphorylation. This causes beta-cells not to release insulin due to the inhibitory effect on S6 kinase. This eventually causes cellular insulin resistance and the development of T2DM. Not only leucine, the concentrations of other branched-chain amino acids such as isoleucine and valine were also found to be statistically significantly higher by 1.5- to 2-fold higher in T2DM patients than those in healthy subjects.[23,24] Therefore, an increased concentration of branched-chain amino acids is a reliable predictor of future insulin resistance among T2DM patients.[25,26] Our study found that hydroxyprolyl-leucine and *N*-palmitoyl threonine were higher in patients with T2DM complicated with CKD compared to patients without complication. These may be potential biomarkers for monitoring T2DM patients in an attempt to delay or prevent them from being inflicted with CKD. According to Human Metabolome Database (HMDB), hydroxyprolyl-leucine has not yet been identified in human tissues or biofluids. It is a dipeptide of hydroxyproline and leucine due to an incomplete breakdown product of protein digestion or protein catabolism and is likely to be a short-lived intermediate. It might have been profiled in our samples before they were degraded by proteolysis. However, it may carry physiological or cell-signaling effects that require further study (accessed on 21 November 2022. Human Metabolome Database: showing metabocard for hydroxyprolyl-leucine (HMDB0028867). On the other hand, *N*-palmitoyl threonine is an amino acid conjugate of a long-chain *N*-acylamide. The research on the roles of *N*-acyl amides is ongoing and more potential novel roles of *N*-acyl amides in health and disease will unwind in the future. Thus far, *N*-acyl amides have been reported to play various signaling functions in cardiovascular activity, metabolic homeostasis, memory, cognition, pain, and motor control.[27] In addition, *N*-acyl amides are implicated in cell migration, inflammation, and diseases such as diabetes, cancer, neurodegenerative disease, and obesity.[28,29] At the same time, we found that 4,4′-Thiobis-2-butanone, geranyl-hydroxybenzoate, and Sesamex were higher in patients with CKD complications.

Besides, Asp Glu Trp, Trp Met Met were higher in patients with T2DM and IHD compared to those normal subjects without risk. We detected L-beta-aspartyl-L-phenylalanine and phenyl-2-propenal which were higher in the normal subjects (Table 2). These amino acids have not been reported to be related to any disease thus far but may have roles in cell signaling or other physiological importance.

Changes in the plasma phospholipids, triglycerides, cholesterol esters, sphingolipids, glycerophospholipids, sphingomyelins, and fatty acids, such as dodecanoic and myristic acids, were reported in individuals suffering from T2DM.[30-32] Sphingomyelins and glycerophospholipids,[33] myristic, and stearic acid tend to be higher in individuals with type 1 diabetes mellitus (T1DM) than in individuals without diabetes.[34,35] Consistently with the earlier findings, C16 sphinganine and phytosphingosine were higher in patients with T2DM and IHD compared to those normal subjects. Lysophosphatidyl ethanolamine and lysophosphatidyl choline were also reported to be higher in those with T1DM than those without diabetes.[30] We detected PE (17:1(9Z)/0:0), LysoPE (0:0/20:0), and (6S)-dehydrovomifoliol and 9Z,12Z,15E-octadecatrienoic acid were higher in subjects with risk of T2DM compared to those without risk. 6,9-Heptadecadiynoic acid was profiled to be present at a lower abundance in patients with T2DM only and those with complications. The biological roles of PE [17:1 (9Z)/0:0), LysoPE (0:0/20:0), and (6S)-dehydrovomifoliol, 9Z,12 Z,15E-octadecatrienoic acid, 6, 9-heptadecadiynoic acid are not clear and will require further exploration.

Other metabolites that had been reported by other studies to be higher in T2DM patients are levels of glucose, deoxyhexose, mannose, and dihexose.[36] Organic acids, such as acetic acid, dimethyl ester, and maleic acid, arginine, citrulline, and ornithine have been associated with T2DM.[25] Suhre et al[36] showed that plasma levels of

1,5-anhydroglucitol were approximately 40% lower in people with T2DM than in healthy individuals.[36] In our study, (S)-alpha-terpinyl glucoside was identified as a marker differentially expressed between normal and patients with T2DM and CKD. We also detected that 5-chola-7, 9 (11)-dien-24-oic acid, *N*-acryloylglycine, 4,8 dimethylnonanoyl carnitine, cis-caryophyllene, and 9,10-epoxy-18-hydroxystearate to be higher in patients with T2DM and complicated with IHD compared to T2DM patients without complications.

We developed the predictive models using the 4 algorithms provided by Metaboanalyst 5.0. It was interesting to note that random forest provided a better predictive accuracy for all the comparison groups we studied here compared to linear SVM, PLS-DA, or logistic regression. This could be due to the advantage of Random Forest being able to handle large data sets and its capability to work with thousands of variables. There have been many studies completed which compared the efficiency of random forest and logistic regression. Random forest was found to perform better in the model for clinical risk scores for predicting clinical outcomes of atrial fibrillation[37], but the outcomes were similar to other machine learning approaches for reports.[38,39] Therefore, we would suggest that different algorithms should be evaluated to identify model with good predictive accuracies.

Conclusion

The LCMS-QTOF was used to profile the differentially expressed metabolites in the normal subjects and T2DM patients with and without complications of IHD or CKD. We report here amino acids as well as lipids which are potential biomarkers differentiating the different subjects and complications. The metabolism pathways that were dysregulated among the patients are the tyrosine, tryptophan, and glycerol metabolism pathways which are consistent with many reports. The top 5 biomarkers approach showed that random forest algorithm produced the prediction with the highest accuracies. However, these models will require further validation before they can be translated into clinical use. Metabolomics is a new emerging field that provides comprehensive phenotypic information on the disease and drug response of a patient. It serves as a potential comprehensive therapeutic drug monitoring approach to be adopted in the near future for pharmaceutical care.

Ethics Committee Approval: This study was approved by UiTM Research Committee of Universiti Teknologi MARA University (Date: December 22, 2016; Approval No: REC/377/16).

Informed Consent: Written consent was obtained from the participants.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – M.Z.S.; Design – R.A.G.; Supervision – M.S.R.; Resources – R.A.G.; Materials – T.L.K.; Data Collection and/or Processing – M.S.R.; Analysis and/or Interpretation – T.L.K.; Literature Search – N.A.M.N.H.; Writing Manuscript – T.L.K.; Critical Review – M.Z.S.

Acknowledgment: The authors thanked all the subjects who participated in the study and the technicians and nurses involved.

Declaration of Interests: The authors have no conflicts of interest to declare

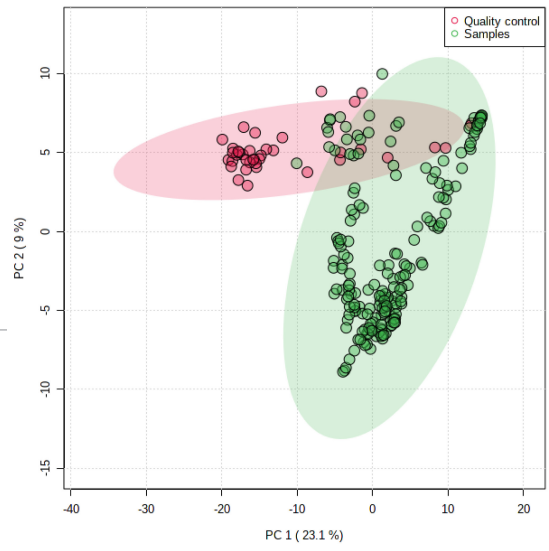
Funding: This study was funded by budget from Integrative Pharmacogenomics, UiTM for a project entitled Global Metabolomics for Patients with T2DM and Complication (241910/2019/RAG/11).

References

1. Lipscombe LL, Hux JE. Trends in diabetes prevalence, incidence, and mortality in Ontario, Canada 1995-2005: a population-based study. *Lancet*. 2007;369(9563):750-756. [CrossRef]
2. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care*. 2004;27(5):1047-1053. [CrossRef]
3. Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract*. 2022;183:109119. [CrossRef]
4. Chen SC, Tseng CH. Dyslipidemia, kidney disease, and cardiovascular disease in diabetic patients. *Rev Diabet Stud*. 2013;10(2-3):88-100. [CrossRef]
5. Lu Y, Wang Y, Ong CN, et al. Metabolic signatures and risk of type 2 diabetes in a Chinese population: an untargeted metabolomics study using both LC-MS and GC-MS. *Diabetologia*. 2016;59(11):2349-2359. [CrossRef]
6. Steinberger J, Daniels SR, American Heart Association Atherosclerosis, Hypertension, and Obesity in the Young Committee (Council on Cardiovascular Disease in the Young), American Heart Association Diabetes Committee (Council on Nutrition, Physical Activity, and Metabolism). Obesity, insulin resistance, diabetes, and cardiovascular risk in children: an American Heart Association scientific statement from the Atherosclerosis, Hypertension, and Obesity in the Young Committee (Council on Cardiovascular Disease in the Young) and the Diabetes Committee (Council on Nutrition, Physical Activity, and Metabolism). *Circulation*. 2003;107(10):1448-1453.
7. Wang W, Vignani R, Scali M, Cresti M. A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. *Electrophoresis*. 2006;27(13):2782-2786. [CrossRef]
8. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. [CrossRef]
9. Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res*. 2018;46(D1):D608-D617. [CrossRef]
10. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019. *Results. Institute for health metrics and evaluation*; 2020. Available at: https://vizhub.healthdata.org/gbd-results/.
11. Guasch-Ferré M, Hruby A, Toledo E, et al. Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis. *Diabetes Care*. 2016;39(5):833-846. [CrossRef]
12. Tai ES, Tan ML, Stevens RD, et al. Insulin resistance is associated with a metabolic profile of altered protein metabolism in Chinese and Asian-Indian men. *Diabetologia*. 2010;53(4):757-767. [CrossRef]
13. Chen R, Ovbiagele B, Feng W. Diabetes and stroke: epidemiology, pathophysiology, pharmaceuticals and outcomes. *Am J Med Sci*. 2016;351(4):380-386. [CrossRef]
14. Jin Q, Ma RCW. Metabolomics in diabetes and diabetic complications: insights from epidemiological studies. *Cells*. 2021;10(11):10(11):2832. [CrossRef]
15. Lin M, Weng SY, Chai KF, Mao ZJ. Lipidomics as a tool of predicting progression from non-alcoholic fatty pancreas disease to type 2 diabetes mellitus. *RSC Adv*. 2019; 16;9(71):41419-41430. [CrossRef]
16. Gall WE, Beebe K, Lawton KA, et al. Alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS One*. 2010;5(5):e10883. [CrossRef]
17. Newgard CB, An J, Bain JR, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab*. 2009;9(4):311-326. [CrossRef]
18. Lotta LA, Scott RA, Sharp SJ, et al. Genetic predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: a Mendelian randomisation analysis. *PLoS Med*. 2016;13(11):e1002179. [CrossRef]

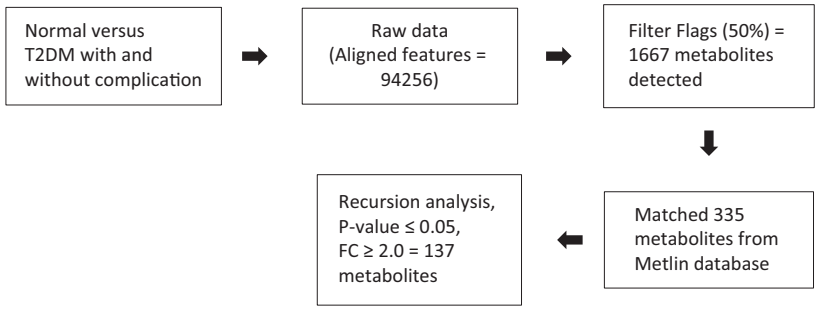
19. Vangipurapu J, Stancáková A, Smith U, Kuusisto J, Laakso M. Nine amino acids are associated with decreased insulin secretion and elevated glucose levels in a 7.4-year follow-up study of 5,181 Finnish men. *Diabetes*. 2019;68(6):1353-1358. [CrossRef]
20. Lucio M, Fekete A, Weigert C, et al. Insulin sensitivity is reflected by characteristic metabolic fingerprints–a Fourier transform mass spectrometric non-targeted metabolomics approach. *PLoS One*. 2010;5(10):e13317. [CrossRef]
21. Ho JE, Larson MG, Vasan RS, et al. Metabolite profiles during oral glucose challenge. *Diabetes*. 2013;62(8):2689-2698. [CrossRef]
22. Geidenstam N, Spégel P, Mulder H, Filipsson K, Ridderstråle M, Danielsson APH. Metabolite profile deviations in an oral glucose tolerance test-a comparison between lean and obese individuals. *Obesity (Silver Spring)*. 2014;22(11):2388-2395. [CrossRef]
23. Roberts LD, Koulman A, Griffin JL. Towards metabolic biomarkers of insulin resistance and type 2 diabetes: progress from the metabolome. *Lancet Diabetes Endocrinol*. 2014;2(1):65-75. [CrossRef]
24. Savolainen O, Fagerberg B, Vendelbo Lind M, Sandberg AS, Ross AB, Bergström G. Biomarkers for predicting type 2 diabetes development—can metabolomics improve on existing biomarkers? *PLoS One*. 2017;12(7):e0177738. [CrossRef]
25. Yoon MS. The emerging role of branched-chain amino acids in insulin resistance and metabolism. *Nutrients*. 2016;8(7):405. [CrossRef]
26. Tessari P, Cecchet D, Cosma A, et al. Insulin resistance of amino acid and protein metabolism in type 2 diabetes. *Clin Nutr*. 2011;30(3):267-272. [CrossRef]
27. Bradshaw HB, Walker JM. The expanding field of cannabimimetic and related lipid mediators. *Br J Pharmacol*. 2005;144(4):459-465. [CrossRef]
28. Grapov D, Adams SH, Pedersen TL, Garvey WT, Newman JW. Type 2 diabetes associated changes in the plasma non-esterified fatty acids, oxylipins and endocannabinoids. *PLoS One*. 2012;7(11):e48852. [CrossRef]
29. Raboune S, Stuart JM, Leishman E, et al. Novel endogenous N-acyl amides activate TRPV1-4 receptors, BV-2 microglia, and are regulated in brain in an acute model of inflammation. *Front Cell Neurosci*. 2014;8(8):195. [CrossRef]
30. Han LD, Xia JF, Liang QL, et al. Plasma esterified and non-esterified fatty acids metabolic profiling using gas chromatography-mass spectrometry and its application in the study of diabetic mellitus and diabetic nephropathy. *Anal Chim Acta*. 2011;689(1):85-91. [CrossRef]
31. Ha CY, Kim JY, Paik JK, et al. 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 (Oxf)*. 2012;76(5):674-682. [CrossRef]
32. Meikle PJ, Wong G, Barlow CK, et al. Plasma lipid profiling shows similar associations with pre-diabetes and type 2 diabetes. *PLoS One*. 2013;8(9):e74341. [CrossRef]
33. Mahendran Y, Cederberg H, Vangipurapu J, et al. Glycerol and fatty acids in serum predict the development of hyperglycemia and type 2 diabetes in Finnish men. *Diabetes Care*. 2013;36(11):3732-3738. [CrossRef]
34. Kaur P, Rizk N, Ibrahim S, et al. Quantitative metabolomic and lipidomic profiling reveals aberrant amino acid metabolism in type 2 diabetes. *Mol Biosyst*. 2013;9(2):307-317. [CrossRef]
35. Wang-Sattler R, Yu Z, Herder C, et al. Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol*. 2012;8:615. [CrossRef]
36. Suhre K, Meisinger C, Döring A, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One*. 2010;5(11):e13953. [CrossRef]
37. Watanabe E, Noyama S, Kiyono K, et al. Comparison among random forest, logistic regression, and existing clinical risk scores for predicting outcomes in patients with atrial fibrillation: a report from the J-RHYTHM registry. *Clin Cardiol*. 2021;44(9):1305-1315. [CrossRef]
38. Hajipour F, Jozani MJ, Moussavi Z. A comparison of regularized logistic regression and random forest machine learning models for daytime diagnosis of obstructive sleep apnea. *Med Biol Eng Comput*. 2020;58(10):2517-2529. [CrossRef]
39. Lu Z, Zhang J, Hong J, et al. Development of a nomogram to predict 28-day mortality of patients with sepsis-induced coagulopathy: an analysis of the MIMIC-III database. *Front Med (Lausanne)*. 2021;8:661710. [CrossRef]

Supplementary Table S1. The distribution and analyses for quality controls data points



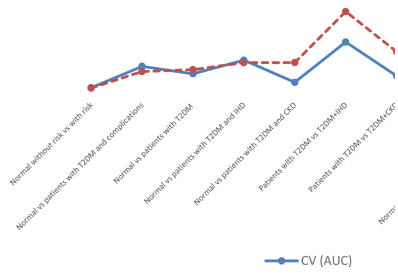
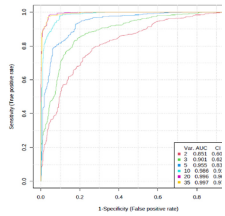
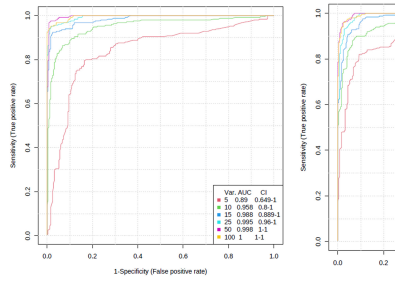
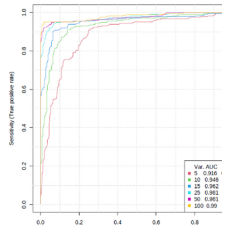
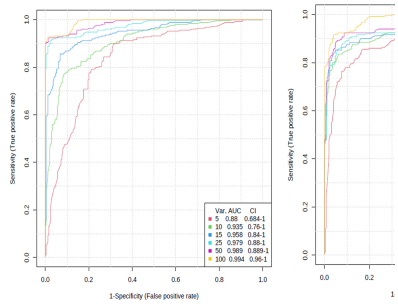
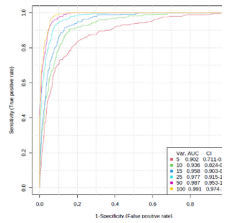
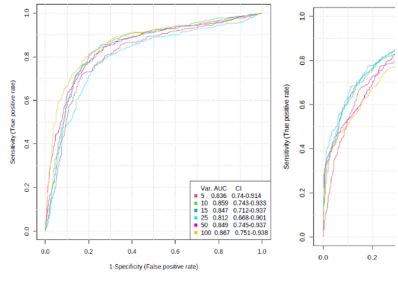
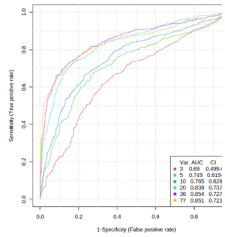
Metabolites	m/z			Retention time		
	Mean	SD	RSD (%)	Mean	SD	RSD (%)
D-Proline	116.0706	0.0008	0.0007	0.7815	0.0022	0.2777
2-Amino-3-methyl-1-butanol	104.4903	0.2422	0.2318	0.8599	0.0018	0.2038
3beta-Hydroxy-16-phosphonopregn-5-en-20-one monoethyl ester	447.2227	0.0188	0.0042	5.8948	0.0682	1.1563
Octylamine	130.1572	0.0014	0.0011	6.2328	0.0283	0.4540
Isodiospyrin	413.042	0.0021	0.0005	6.7254	0.0128	0.1901
Fumaricine	408.1231	0.0022	0.0005	8.8150	0.0091	0.1035
Casimiroedine	435.2517	0.0106	0.0024	9.0615	0.0591	0.6518
9-oxo-2E-decenoic acid	185.1164	0.0034	0.0018	10.4688	0.0239	0.2285
Met His Lys	437.2120	0.1414	0.0323	15.4281	0.0301	0.1949
3α,9α,11β-Trihydroxy-5β-cholan-24-oic Acid	426.3637	0.0378	0.0089	16.7094	0.0136	0.0814

Supplementary Table S2. The flow of analysis and the number of metabolites identified through statistical analysis for each comparison groups



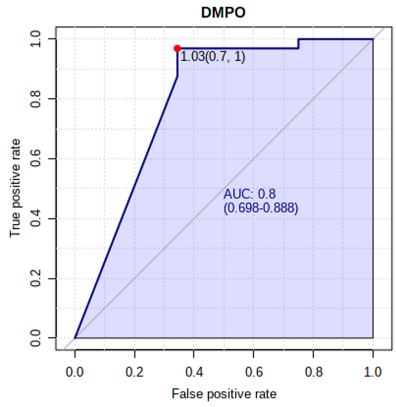
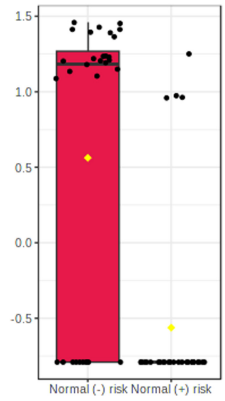
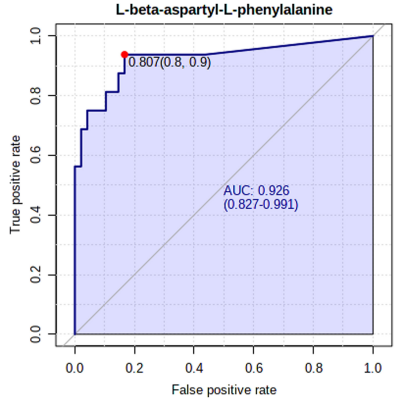
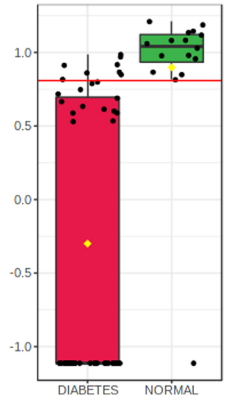
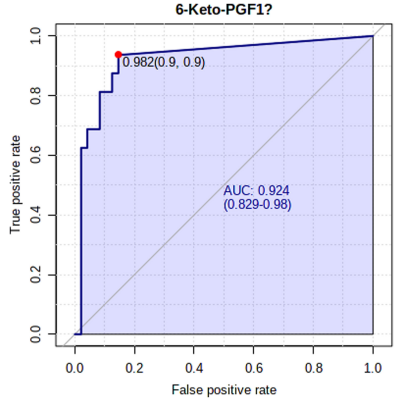
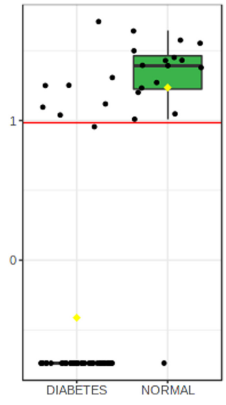
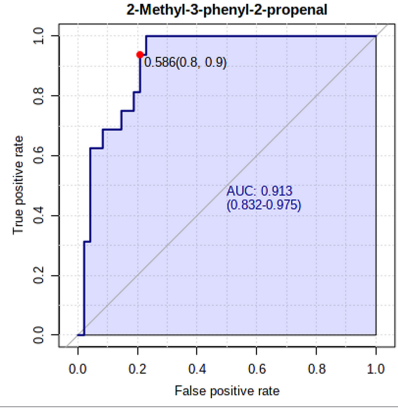
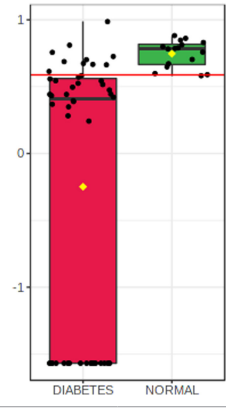
	Groups	Raw data (Aligned features)	Filter Flags (50%)	Matched metabolites from Metlin database	Recursion analysis, P-value ≤ 0.05, FC ≥ 2.0
1.	Normal no risk vs Normal with risk)	72960	1586	60	35
2.	Normal versus T2DM with and without complication	94256	1667	335	137
3.	Normal versus T2DM only	56819	1666	339	144
4.	Normal versus T2DM with IHD	55454	1935	363	151
5.	Normal versus T2DM with CKD	59349	1736	186	71

Supplementary Table S3. The top 5 metabolites and the AUC and levels in different groups of subjects

Metabolite	AUC	T-test	Log2 FC	AUC graphs	Levels of metabolites
Normal without risk vs Normal with risk					
PE(17:1(9Z)/0:0)	0.90771	2.196E-11	-1.4376		
LysoPE(0:0/20:0)	0.81396	1.6404E-4	-0.75239		
(6S)-dehydrovomifoliol	0.80859	6.8818E-7	-0.903		
9Z,12Z,15E-octa decatrienoic acid	0.80664	2.3387E-7	-1.2827		

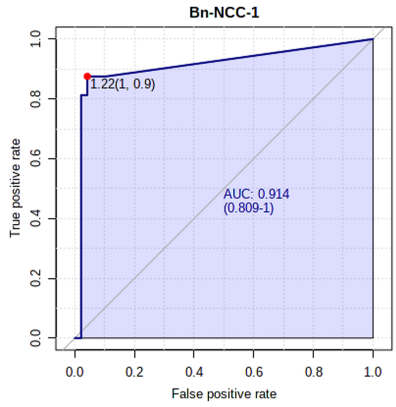
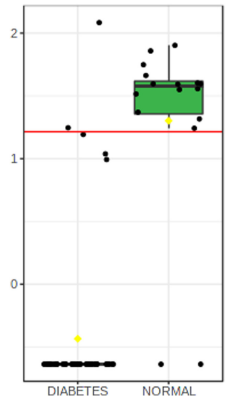
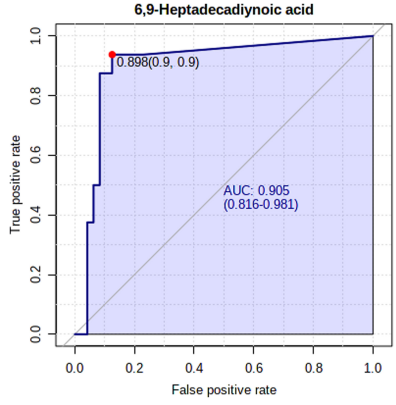
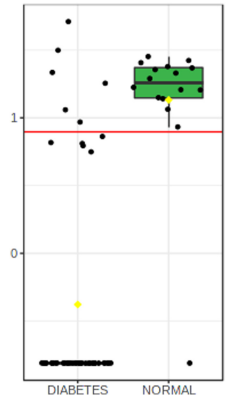
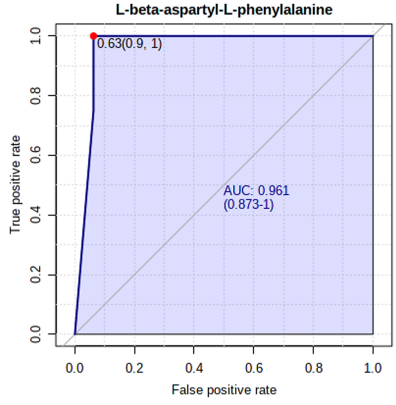
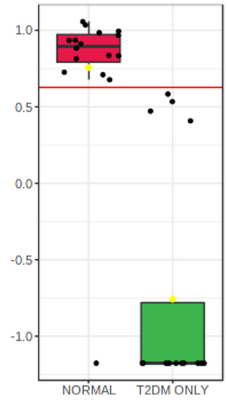
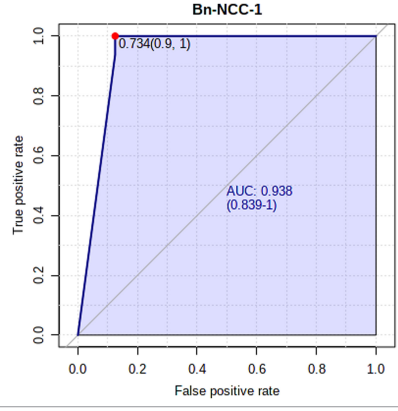
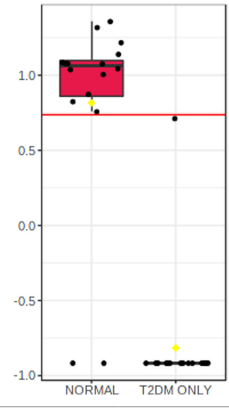
(*Continued*)

Supplementary Table S3. The top 5 metabolites and the AUC and levels in different groups of subjects (*Continued*)

Metabolite	AUC	T-test	Log2 FC	AUC graphs	Levels of metabolites
DMPO	0.79395	1.0411E-6	1.4271		
Normal vs patients with T2DM + complication					
L-beta-aspartyl -L-phenylalanine	0.92383	9.3164E-6	-0.98459		
6-Keto-PGF1	0.92057	2.21E-11	-1.6784		
2-Methyl-3-phenyl-2-propenal	0.91406	3.3723E-4	-0.64269		

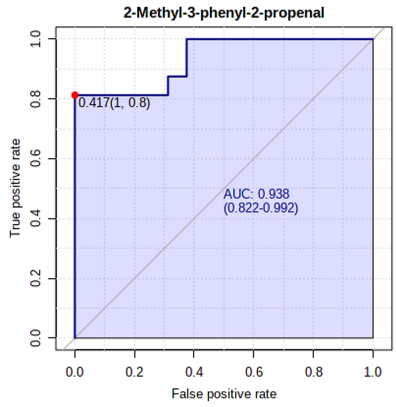
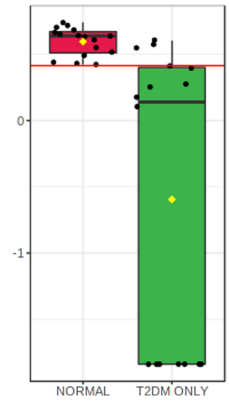
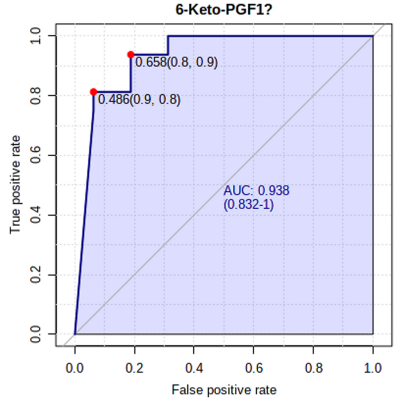
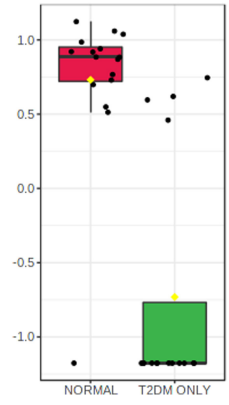
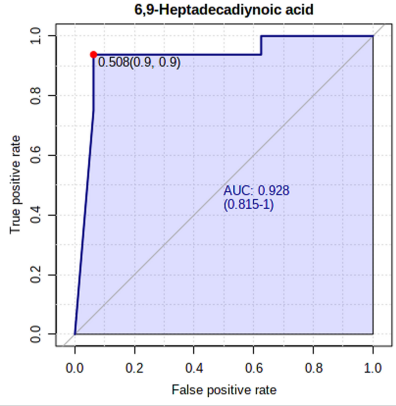
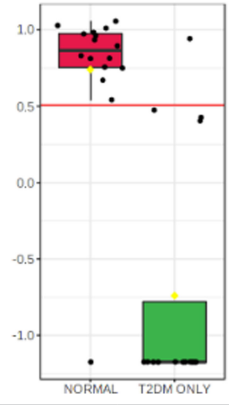
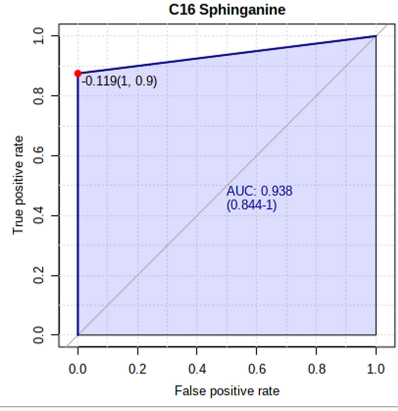
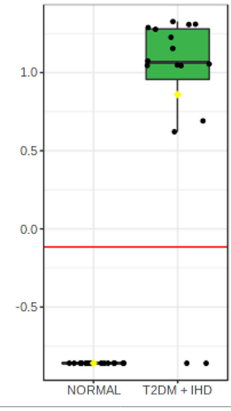
(*Continued*)

Supplementary Table S3. The top 5 metabolites and the AUC and levels in different groups of subjects (*Continued*)

Metabolite	AUC	T-test	Log2 FC	AUC graphs	Levels of metabolites
Bn-NCC-1	0.91146	4.2186E-13	-1.9428		
6,9-Heptadecadiynoic acid	0.89909	3.2131E-9	-1.5034		
Normal vs patients with T2DM only					
L-beta-aspartyl -L-phenylalanine	0.96094	2.5874E-7	1.516		
Bn-NCC-1	0.93359	4.8004E-9	2.072		

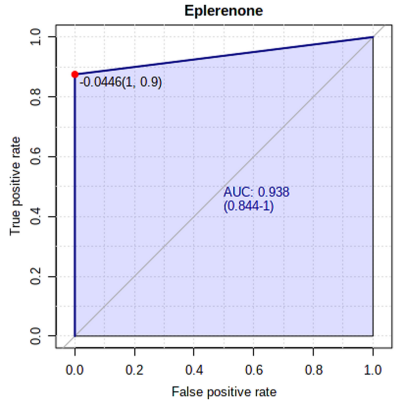
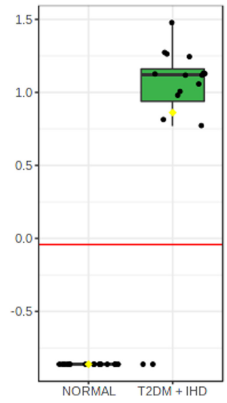
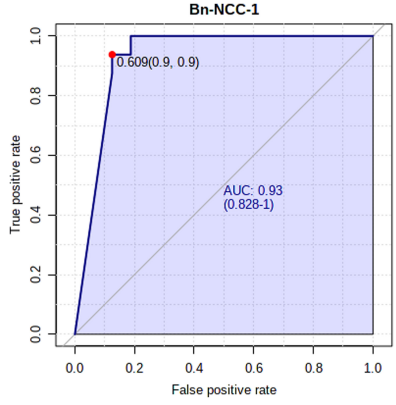
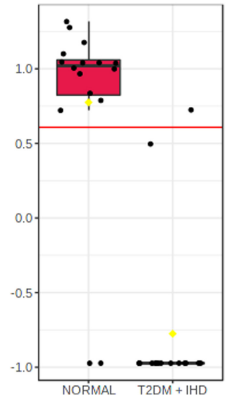
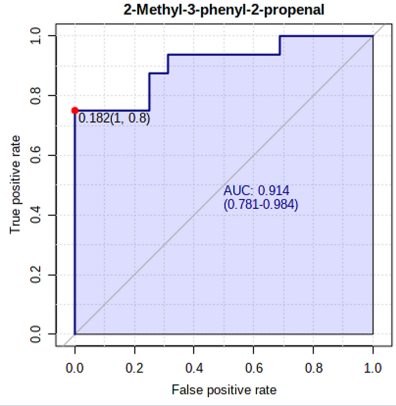
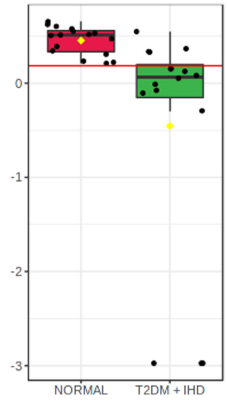
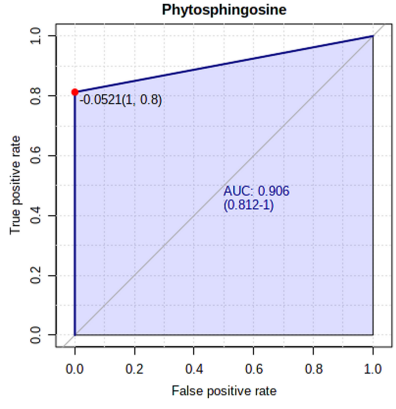
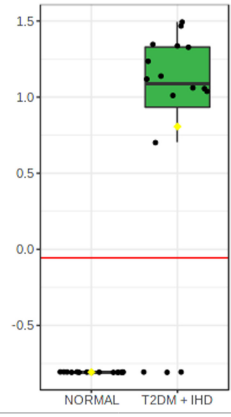
(*Continued*)

Supplementary Table S3. The top 5 metabolites and the AUC and levels in different groups of subjects (*Continued*)

Metabolite	AUC	T-test	Log2 FC	AUC graphs	Levels of metabolites
2-Methyl-3-phenyl-2-propenal	0.93359	2.4586E-4	0.75583		
6-Keto-PGF1	0.92969	1.1155E-6	1.4391		
6,9-Heptadecadiynoic acid	0.92578	6.8653E-7	1.4782		
Normal vs patients with T2DM + IHD					
C16 Sphinganine	0.9375	7.3188E-11	-2.4962		

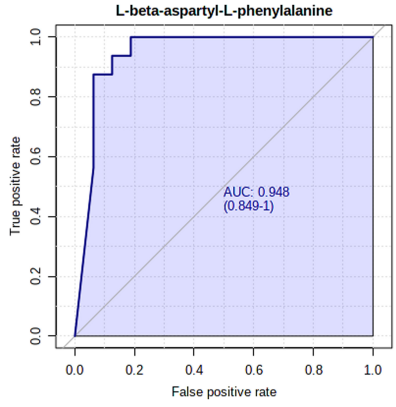
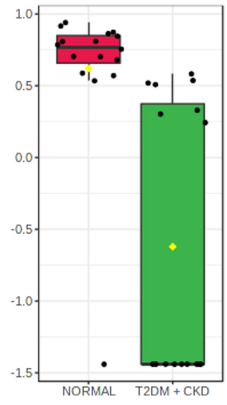
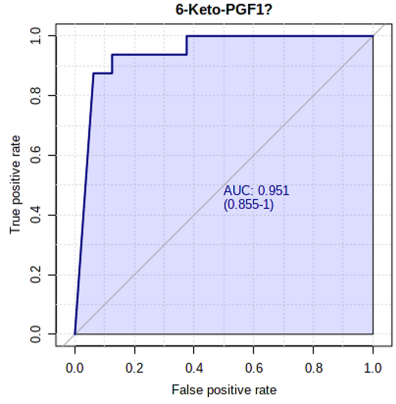
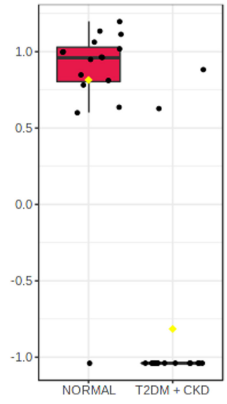
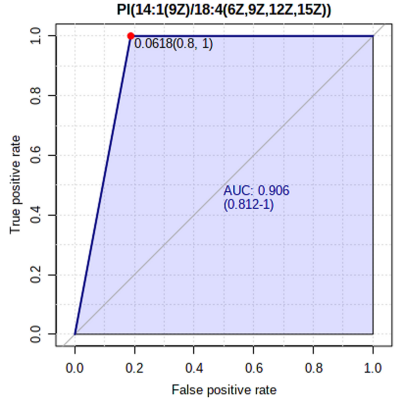
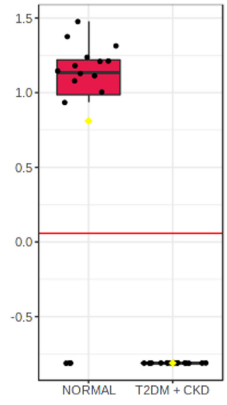
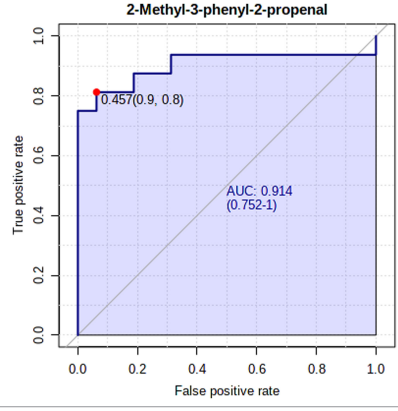
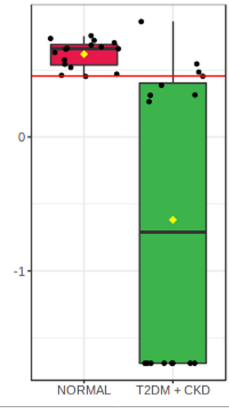
(*Continued*)

Supplementary Table S3. The top 5 metabolites and the AUC and levels in different groups of subjects (*Continued*)

Metabolite	AUC	T-test	Log2 FC	AUC graphs	Levels of metabolites
Eplerenone	0.9375	5.3541E-11	-2.3836		
Bn-NCC-1	0.92578	8.8457E-8	1.9038		
2-Methyl-3-phenyl-2-propenal	0.90625	0.0079152	0.36111		
Phytosphingosine	0.90625	9.6428E-9	-2.3996		

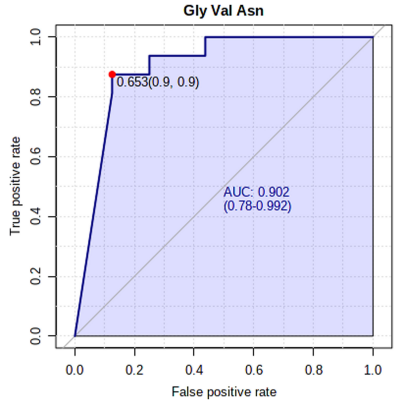
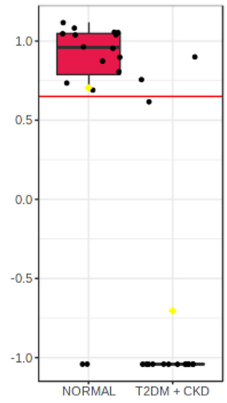
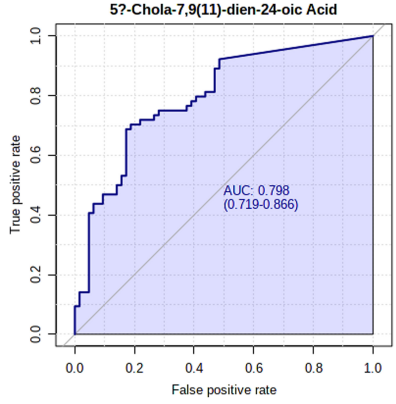
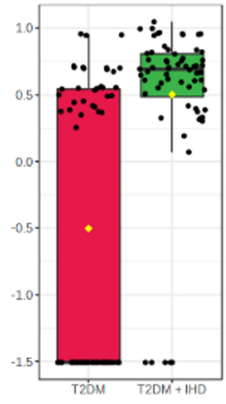
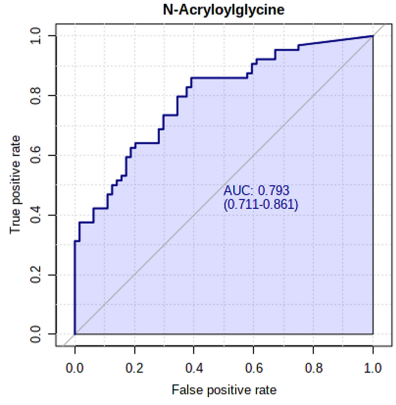
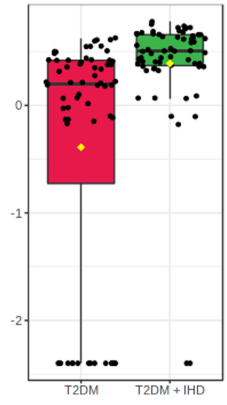
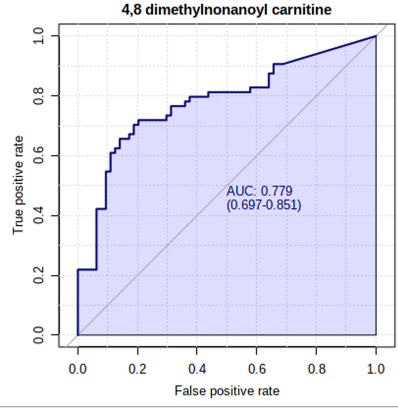
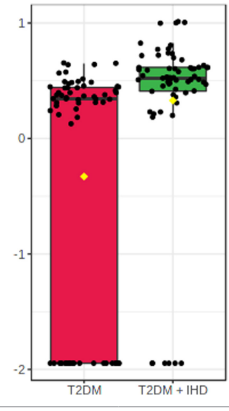
(*Continued*)

Supplementary Table S3. The top 5 metabolites and the AUC and levels in different groups of subjects (*Continued*)

Metabolite	AUC	T-test	Log2 FC	AUC graphs	Levels of metabolites
Normal vs patients with T2DM + CKD					
L-beta-aspartyl -L-phenylalanine	0.94336	1.0794E-4	1.0015		
6-Keto-PGF1	0.94141	5.0385E-9	1.8359		
PI(14:1(9Z)/18: 4(6Z,9Z,12Z,15Z))	0.90625	7.002E-9	2.2378		
2-Methyl-3-phen yl-2-propenal	0.90234	1.1724E-4	0.84287		

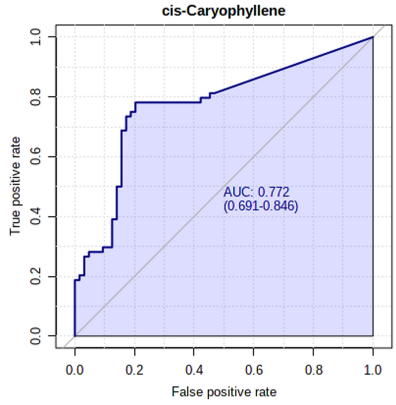
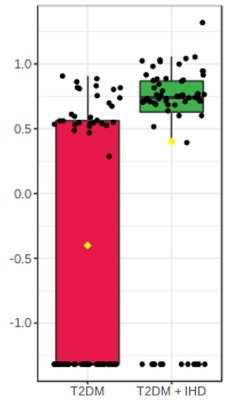
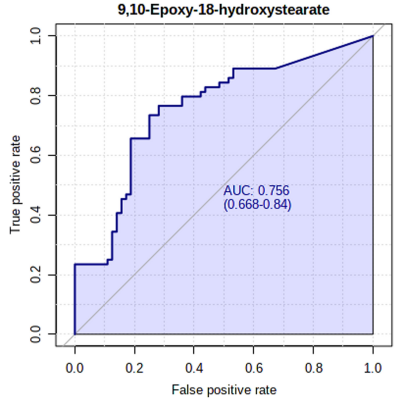
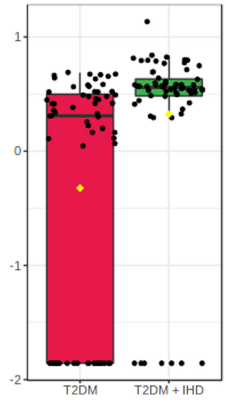
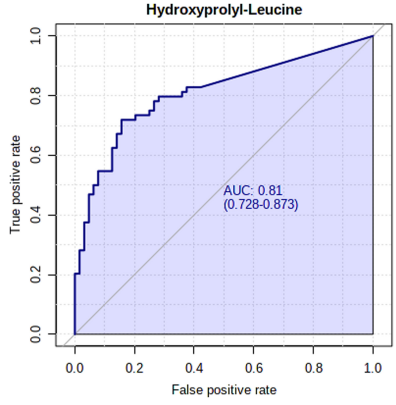
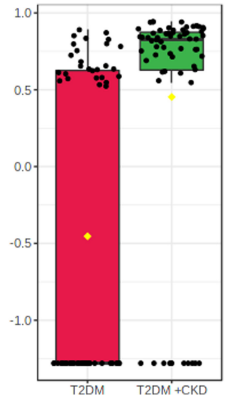
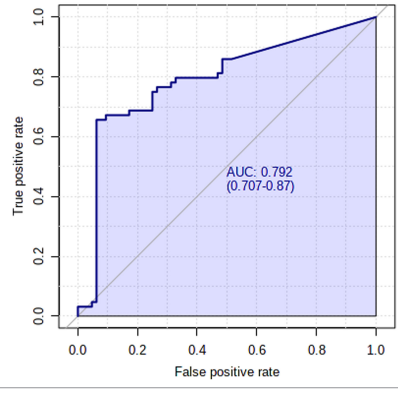
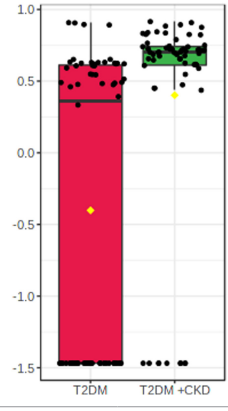
(*Continued*)

Supplementary Table S3. The top 5 metabolites and the AUC and levels in different groups of subjects (*Continued*)

Metabolite	AUC	T-test	Log2 FC	AUC graphs	Levels of metabolites
Gly Val Asn	0.89844	4.17E-6	1.523		
T2DM vs T2DM + IHD					
5?-Chola-7,9(11)-dien-24-oic Acid	0.79602	1.2596E-9	-0.7803		
N-Acryloylglycine	0.79297	5.061E-6	-0.38269		
4,8 dimethylnonanoyl carnitine	0.7793	1.3863E-4	-0.38768		

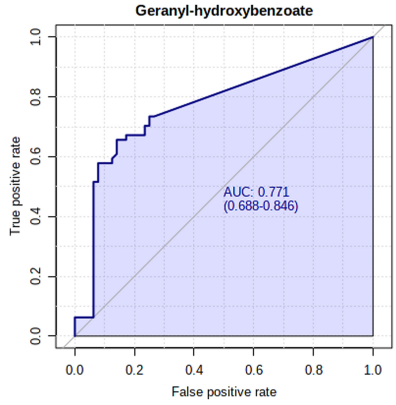
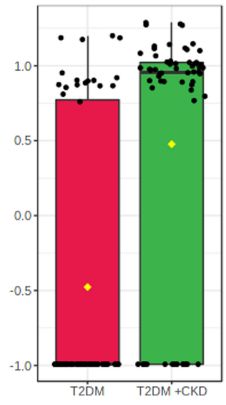
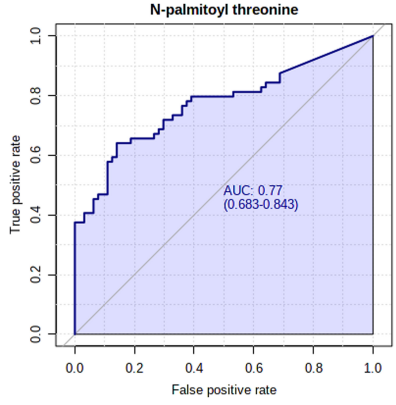
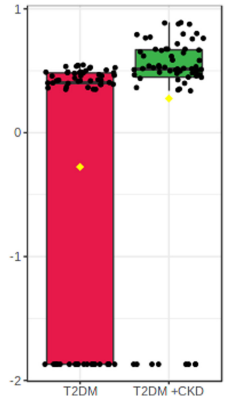
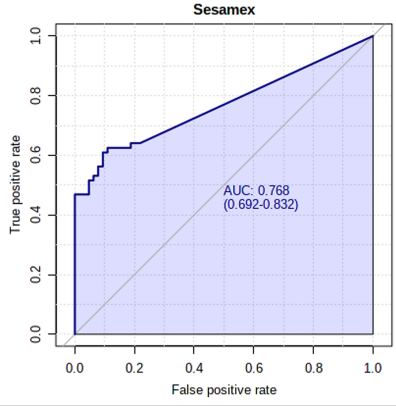
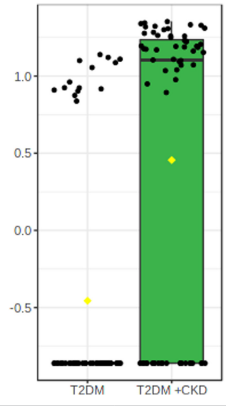
(*Continued*)

Supplementary Table S3. The top 5 metabolites and the AUC and levels in different groups of subjects (*Continued*)

Metabolite	AUC	T-test	Log2 FC	AUC graphs	Levels of metabolites
cis-Caryophyllene	0.77051	2.4993E-6	-0.68556		
9,10-Epoxy-18-hydroxystearate	0.75598	1.8658E-4	-0.40239		
T2DM Vs T2DM+CKD					
Hydroxyprolyl-Leucine	0.80872	6.8133E-8	-0.77378		
4,4'-Thiobis-2-butanone	0.79065	2.3626E-6	-0.61298		

(*Continued*)

Supplementary Table S3. The top 5 metabolites and the AUC and levels in different groups of subjects (*Continued*)

Metabolite	AUC	T-test	Log2 FC	AUC graphs	Levels of metabolites
Geranyl-hydroxy benzoate	0.77026	1.1051E-8	-1.0001		
N-palmitoyl threonine	0.77026	0.0015492	-0.33026		
Sesamex	0.76636	5.583E-8	-1.0698		

Appendix A. Important features for prediction model

	Rank Freq.	Importance	T2DM	T2DM +CKD
4,4'-Thiobis-2-butanone	1	7.894735	Low	High
Hydroxyprolyl-Leucine	1	6.932916	Low	High
Sesamex	1	6.427967	Low	High
Trp Thr Tyr	1	6.357728	Low	High
2-([4-(2-Chlorophenyl)-5-methoxycarbonyl-3-ethoxycarbonyl-6-methyl-2-pyridyl]methoxyacetic acid	1	6.210388	Low	High
Apigenin 7-glucuronosyl-(1->2)-glucuronide	1	6.088583	Low	High
N-palmitoyl threonine	1	5.84147	Low	High
methyl 4-[2-(2-formyl-vinyl)-3-hydroxy-5-oxo-cyclopentyl]-butanoate	1	5.601029	Low	High
3E,4Z,7,11-Tetramethyl-6,10-tridecadienal	1	5.239326	Low	High
Fusicoccin H	0.966667	7.223361	Low	High
Geranyl-hydroxybenzoate	0.966667	6.373256	Low	High
2,5-Dimethoxycinnamic acid	0.966667	5.47431	Low	High
methyl 15,16-epoxy-9,12-octadecadienoate	0.966667	5.08058	Low	High
alpha-Carboxy-delta-decalactone	0.933333	5.845996	Low	High
Selagine	0.933333	5.480653	Low	High
Phenacylamine	0.933333	5.08747	Low	High
14-Fluoro-11Z-tetradecenyl acetate	0.933333	4.732145	Low	High
Prosafrinine	0.9	5.792177	Low	High
N-Acryloylglycine	0.9	4.920724	Low	High
Ritodrine glucuronide	0.866667	6.388905	Low	High
Mupirocin	0.866667	5.199133	Low	High
3-(4-Hydroxyphenyl)propionic acid	0.833333	4.347908	Low	High
Tetrahydropentoxyline	0.8	4.630027	Low	High
Guaifenesin	0.8	4.318781	Low	High
Dihydropicromycin	0.8	4.257284	Low	High
Pentadecanoyl-EA	0.766667	5.39465	Low	High
Zedoarol	0.766667	4.043434	Low	High
Isoxeniaphyllenol	0.733333	4.207294	Low	High
Ethyl Oxalacetate	0.733333	3.970136	Low	High
cis-Caryophyllene	0.733333	3.848878	Low	High
N-palmitoyl glutamic acid	0.7	3.894902	Low	High
Xestoaminol C	0.666667	3.957251	Low	High
10-hydroxy-2E,8Z-Decadiene-4,6-diynoic acid	0.666667	3.926729	Low	High
Kamahine C	0.633333	5.066685	Low	High
Tracheloside	0.6	3.727567	Low	High
N,N-Didesmethyltamoxifen	0.566667	3.568566	Low	High
3-Hydroxybenzaldehyde	0.566667	3.557737	Low	High
Fludiazepam	0.5	3.346745	Low	High
7Z,11Z,14E-eicosatrienoic acid	0.433333	3.475684	Low	High
UDP-L-Ara4O	0.4	3.609028	Low	High
Urdamycin G	0.333333	3.195331	Low	High
10-Nitrooleate	0.333333	2.935037	Low	High
Pentagastrin	0.3	3.056213	Low	High
17beta-Methylestra-1,3,5(10)-trien-3-ol	0.3	3.025458	Low	High
11-methoxy-12,13-epoxy-9-octadecenoic acid	0.3	2.969978	Low	High
AX 048	0.266667	3.057409	Low	High
1?,25-dihydroxy-24-norvitamin D3 / 1?,25-dihydroxy-24-norcholecalciferol	0.266667	2.892184	Low	High
PI(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/21:0)	0.266667	2.679541	Low	High
Medroxyprogesterone glucuronide	0.233333	2.894509	Low	High
Fasoracetam	0.233333	2.754144	Low	High

(*Continued*)

Appendix A. Important features for prediction model (*Continued*)

	Rank Freq.	Importance	T2DM	T2DM +CKD
Caffeine	0.2	2.86665	Low	High
Linoleamide	0.2	2.700457	Low	High
Nocardicin C	0.2	2.661127	Low	High
N-docosahexaenoyl GABA	0.2	2.554849	Low	High
hexadecanedioic acid mono-L-carnitine ester	0.166667	2.746778	Low	High
5?-Chola-7,9(11)-dien-24-oic Acid + 12.08878	0.166667	2.408471	Low	High
Laserpitin	0.166667	2.400423	Low	High
L,L-Cyclo(leucylprolyl)	0.133333	2.607865	Low	High
2-Hydroxydesmethylimipramine glucuronide	0.133333	2.422086	Low	High
Tyr Ala Phe	0.133333	2.376717	Low	High
2(?-D-Mannosyl)-D-glycerate	0.133333	2.146971	Low	High
5?-Chola-7,9(11)-dien-24-oic Acid	0.1	2.622958	Low	High
(3a,5b,7b)-24-[(carboxymethyl)amino]-7-hydroxy-24-oxocholan-3-yl-b-D-glucopyranosiduronic acid,	0.1	2.448864	Low	High
3-Methylglutarylcarnitine	0.1	2.211743	Low	High
Cyclopassifloic acid E	0.066667	1.73861	Low	High
Isocarbostyril	0.033333	2.025416	Low	High
15-methyl-1,2-heneicosanediol	0.033333	1.88634	Low	High
Gln Tyr Lys	0.033333	1.861729	Low	High
Nitrendipine	0.033333	1.667372	Low	High