scientific **data**

OPEN

DATA DESCRIPTOR

In-depth profile of biosignatures for T2DM cohort utilizing an integrated targeted LC-MS platform

Shurong Ma [1,2,5], Lu Yang [3,5], Jinwen Lai [1,2,5], Shan Cheng [1,2], Yunshu Zhang [1,2], Zeming Wu [4], Anliang Huang [1,2], Tianfu Wei [1,2], Qiuying Luo [4], Mimi Wang [1,2], Jianling Du [3] ✉ & Peiyuan Yin [1,2] ✉

The profiling of metabolites provides an immediate snapshot that depicts crucial physiological information, holding immense potential for the early diagnosis and prognosis of diseases, including diabetes. Herein, we proposed an optimized and in-depth target-based metabolome platform through an integration of six distinct conditions, including a normal phase, a pre-column chemical derivatization and four reversed phase separation methods for the quantification of a total of 1609 small molecules (32 sub-classes) in serum after normalization using isotope-labeled internal standards. After undergoing rigorous methodological validation and comprehensive comparison with untargeted strategies, we present a new dataset of metabolomic profile encompassing a cohort of 200 healthy individuals and 100 newly diagnosed Type 2 diabetes mellitus (T2DM) patients from the northern region of China. The overall differential analysis results indicated obvious metabolic disturbance of amino acid, fatty acids, lysophosphatidyl-choline and triacylglycerol in T2DM. We hereby make these technical validation results and the profiling dataset publicly available to the scientific community, showcasing its exceptional sensitivity and robustness as an invaluable tool for the comprehensive targeted metabolome analysis.

Background & Summary

Metabolic biosignatures in biofluids are intricately linked to the human's phenotypic informative dimension, serving not only as substrates/intermediates or end-products of various biological processes, but also as critical signals that could can impact metabolic pathways or modulate other regulatory proteins[1,2]. They offer a comprehensive readout and snapshot depicting the essential physiological information[3] and hold immense potential as biomarkers for early diagnosis and prognosis, as well as providing insights into the effective interventions for future precise medicine[4,5]. Over the past decade, there has been significant interest in the unique variations in levels under diseases and their corresponding bioactive functions[3,6,7]. Fortunately, advancements in metabolomics technologies, such as the utilization of liquid chromatography coupled with mass spectrometry (LC-MS), have facilitated the measurement of thousands of biosignatures to portray a specific metabolic profile under disease[8–12].

Unlike untargeted metabolomics, which aims to detect as many peaks as possible using high resolution mass spectrometers in full-scan MS mode, targeted metabolomics are considered the gold standard for quantitating those metabolites[13,14]. By utilizing liquid chromatography triple quadrupole mass spectrometer (LC−QQQ-MS)-based strategies, it serves as an indispensable tool with prominent advantages for the absolute quantification of those small molecule substances[15] and enables direct comparisons and analysis across batches and even labs, driving a large improvement of data legibility in practical applications[16,17]. Yet the main challenges lie in the relatively low coverage and throughput capacities, prompting efforts to expand the quantification coverage[18,19].

The combination of three chromatography conditions by Yan *et al.*[20], along with the implementation of a derivatization strategy for approximately 300 metabolites in a single run by Xie *et al.*[4], encompassing amino acids,

[1]Laboratory of Integrative Medicine, The First Affiliated Hospital of Dalian Medical University, Dalian, Liaoning, 116000, P. R. China. [2]Institute (College) of Integrative Medicine, Dalian Medical University, Dalian, Liaoning, 116000, P. R. China. [3]Department of Endocrinology, The First Affiliated Hospital of Dalian Medical University, Dalian, Liaoning, 116000, P. R. China. [4]IPhenome Biotechnology Inc., Dalian, Liaoning, 116000, P. R. China. [5]These authors contributed equally: Shurong Ma, Lu Yang, Jinwen Lai. ✉e-mail: dujianling63@163.com; yinpeiyuan@dmu.edu.cn
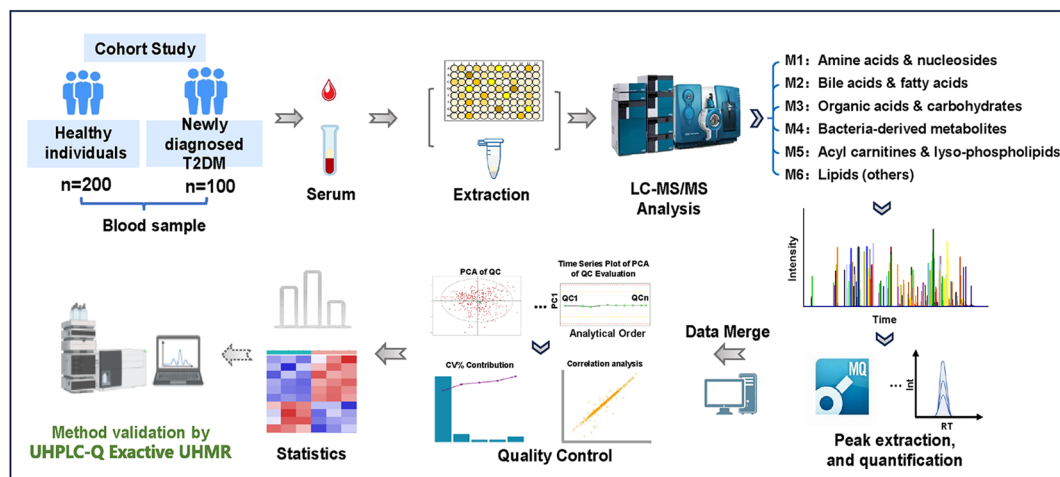
Fig. 1 The workflow of the cohort study and metabolomics analysis.

organic acids, and carbohydrate pathways. Alternatively, the Globally Optimized Targeted Mass Spectrometry (GOT-MS) approach by Gu *et al.* and the Sequential Window Acquisition of All Theoretical Mass Spectrometry (SWATH-MS), a data-dependent acquisition (DDA)-based technique by Wang *et al.*, aims to cover a significantly increased detecting flux[21–23]. However, those methods necessitate the integration of two types of instruments and yields semi-quantitative results as output. Moreover, the selection of precursor ions heavily relies on peak abundance, which entails labor-intensive steps to transfer the ion-pairs between instruments[23,24]. The vast array of chemical diversity poses a significant challenge in devising precise quantification strategies for the profiling of metabolites.

Herein, we proposed an optimized and integrated target-based metabolome platform for biosignature profiling of over 1600 metabolites, spanning across 10 classes and 32 sub-classes. The quantification was performed using UPLC-MS/MS in positive or negative MRM modes with ESI source, and it comprises six separation conditions: a normal phase separation for amino acids, nucleotides, and derivatives; a pre-column chemical derivatization and reversed separation for organic acids and carbohydrates; as well as four reversed phase separation methods for bile acids, fatty acids, intestinal bacterial metabolites, acylcarnitine, and lipids. Absolute quantitative results were obtained for each compound through the mixed isotope internal standard correction. After undergoing methodological validation and a comprehensive comparison between the untargeted strategies, we present a new dataset of metabolomic profile involving a cohort of 200 healthy individuals and 100 newly diagnosed Type 2 diabetes mellitus (T2DM) patients in the northern region of China (in Fig. 1). We have made these methodological validation results and the profiling dataset publicly available to the scientific community. This demonstrates its exceptional sensitivity and robustness as an invaluable tool for metabolome analysis across various laboratories.

Methods

Inclusion and exclusion criteria for the T2DM cohort. In this dataset, we totally recruited 200 individuals as healthy controls, and 100 patients of newly diagnosed Type 2 diabetes mellitus (T2DM). All the subjects were enrolled at the physical examination center and Department of Endocrinology and Metabolism in The First Affiliated Hospital of Dalian Medical University from July 2022 to November 2022. The primary T2DM patients were pathologically diagnosed by standards of medical care in diabetes revised by The American Diabetes Association (2021 edition)[25].

Inclusion criteria for T2DM patients were as follows: (1) they were initially diagnosed with diabetes without known history of this disease and the diagnosis conforms to the T2DM criteria established by the American Diabetes Association[25]; (2) they perform with no prior history of diabetes medication for more than three months; (3) they are willing and able to comply with the research protocol and have provided their informed consent. The criteria for the healthy control group were: (1) they have no prior medical history of diabetes mellitus; (2) the fasting plasma glucose (FPG) is below 6.1 mmol/L, and their 2-hour post-load plasma glucose following the oral glucose tolerance test (OGTT) does not exceed 7.8 mmol/L; (3) they have provided their informed consent.

Exclusion criteria included: (1) they aged below 20 or above 70 years old; (2) they have Type 1 diabetes mellitus or other specialized types of diabetes; (3) they have been previously diagnosed with metabolic syndrome and hyperlipidemia or exhibit severe impairment of cardiovascular, hepatic, or renal functions; (4) they have serious complications, such as diabetic ketoacidosis (DKA) or hyperglycemic hyperosmolar state (HHS); (5) they are pregnant or lactating women; (6) they are currently utilizing corticosteroids or other medications that may influence the blood glucose level; (7) they had received a prior diagnosis or been concomitant with a tumor; (8) they could not tolerate blood collection for severe hemorrhagic disease; (9) they were onset of other diseases such as cardiovascular or cerebrovascular diseases; (10) they refused to participate in the study or request the withdrawal of informed consent.

Group*	Healthy individuals	New diagnosed T2DM	P value
Size	200	100	—
Sex (M/F)	101/99	71/29	**(0.0047)
Age, y	47.24 (1.15)	49.84 (1.19)	NS (0.0719)
BMI, kg/m2	23.88 (0.37)	27.62 (0.41)	***(<0.001)
SBP, mmHg	121.44 (1.55)	138.07 (1.84)	***(<0.001)
DBP, mmHg	74.75 (1.11)	86.63 (1.3)	***(<0.001)
HBA, %	5.47 (0.02)	8.41 (0.2)	***(<0.001)
FBG, mM	5.13 (0.18)	9.81 (0.32)	***(<0.001)
ALT, U/L	20.9 (1.48)	47.84 (7.56)	***(<0.001)
AST, U/L	20.65 (0.65)	33.64 (4.12)	**(0.0024)
Urea, mM	7.02 (2.77)	5.3 (0.13)	NS (0.3851)
Cre, μM	69.07 (5.77)	65.75 (1.5)	NS (0.4467)
UA, μM	338.31 (9.93)	352.87 (11.39)	NS (0.2798)
ALB, g/L	90.7 (44.85)	46.55 (0.32)	NS (0.1666)
TC, mM	5.17 (0.19)	5.33 (0.11)	NS (0.3487)
TG, mM	1.35 (0.08)	2.44 (0.21)	***(<0.001)
HDL-C, mM	1.47 (0.03)	1.34 (0.03)	**(0.0013)
LDL-C, mM	2.67 (0.06)	2.94 (0.08)	**(0.0043)
ALP, U/L	70.46 (2.02)	86.09 (2.45)	***(<0.001)
γ-GT, U/L	27.71 (2.59)	49.54 (3.59)	***(<0.001)
T-BIL, μM	14.67 (0.68)	14.62 (0.5)	NS (0.9436)
D-BIL, μM	3.94 (0.18)	3.88 (0.15)	NS (0.7648)
TBA, μM	2.75 (0.2)	3.19 (0.18)	NS (0.0547)
CHE, U/L	362.14 (9.19)	451.57 (7.49)	***(<0.001)
WBC, 10^9/L	8.67 (4.39)	6.67 (0.19)	NS (0.5220)
NEUT, %	56.3 (0.71)	57.25 (0.71)	NS (0.2801)
LYMPH, %	51.89 (23.75)	32.52 (0.74)	NS (0.2517)
EOSO, %	2.19 (0.18)	1.93 (0.13)	NS (0.1621)
BASO, %	0.49 (0.03)	0.51 (0.03)	NS (0.4645)

Table 1. Baseline information for the T2DM cohort. *Data are presented as mean (SD).

The protocol has received approval from the institutional research ethics committee of The First Affiliated Hospital of Dalian Medical University (approval no. PJ-KS-KY-2022-325). Written informed consents were obtained from all the participants, encompassing the comprehensive information regarding the study's objectives, procedures, potential risks and benefits. This ensured that participants fully understood their rights, and all the participants have given their consent to share their information for scientific research purposes. The study population was planned to consist of an approximately equal distribution of male and female subjects, with individuals failing to meet the predefined eligibility criteria being systematically screened and excluded from the trial. For each enrolled subject, the general medical and biochemical examination parameters were recorded, and the demographic parameters for this cohort are summarized in Table 1 as baseline information.

Fasting blood samples were collected from each individual in the morning (before 9:00 am) via venipuncture at the antecubital fossa into a sterilized 5 mL glass tube without any additives. After allowing the blood to clot for an hour at room temperature, the samples were centrifuged at 5000 g and 4 °C for 15 minutes. The supernatant (serum) was then carefully transferred into clean and labeled EP tubes, with each containing 200 μL of serum within a time frame of two hours. Subsequently, the packaged serum was immediately frozen at −80 °C in a refrigerator until further analysis.

Chemicals and reagents. The majority of metabolite standards, derivatization reagents and mobile phase additives (acids and salts) were procured from suppliers such as Sigma Chemical Co. (St. Louis, MO), Macklin Inc. (Shanghai, China), TRC Inc. (Toronto, Canada) and Cayman Inc. (Michigan, USA) etc. Additionally, the stable-isotope labeled internal standards (IS), were obtained from isosciences LLC. (Ambler, USA) and Avanti Polar Lipids, Inc. (Alabaster, Alabama), *et al*. Detail list was uploaded as a dataset on Science Data Bank (Dataset 13920, Table 1)[26]. Mass spectrometry-grade mobile phases and the extraction reagent were purchased from Thermo-Fisher Scientific (Waltham, USA). All the single standard, including the stable-isotope labeled internal standards (IS), was firstly prepared at 1.00 mg/mL in methanol or methyl-sulfoxide and stored at −20 °C until use. Purified water for sample preparation and mobile phase was obtained from Millipore-Q Academic system (Waltham, USA).

Preparation of samples and standards. In order to achieve accurate quantification for each metabolite in the samples, precise proportions were employed when preparing the mixed stock solutions, taking into

consideration their practical amounts in serum and detailed information of the concentrations for standards was summarized and uploaded on Science Data Bank (Dataset 13920, Table 2)[26]. Quality control (QC) samples were prepared by pooling all the residual test samples and subsequently evenly divided into predetermined aliquots before sample pre-processing. Ten levels of working solutions were generated through serial dilution with methanol from the mixed stock solutions and prepared for calibration and validation purposes, except that acyl carnitines and lipid levels were calculated by pseudo-targeted strategies, where the concentration values were determined by calculating the area ratio to the internal standard of known level at one point.

M1: for amine acids and nucleosides. 50 μL of serum samples, QCs or different concentration of standards were accurately transferred to the 750 μL 96-well sample collection plates (Waters, Milford, USA). Subsequently, 10 μL of mixed internal standard (M1-IS) and 190 μL of methanol were added to the serum samples and vortex shaking at 1500 rpm for 3 min. Next, the mixture was centrifuged at 3500 g for 20 min under 4 °C and 180 μL of supernatant was transferred, with a total of 2 μL injected for LC-MS/MS analysis.

M2: for bile acids and fatty acids. Briefly, 50 μL of serum samples, QCs or standards were firstly combined with 180 μL of the extraction reagent (including methanol and acetonitrile, V:V = 1:1) into the sample collection plates. Then, 20 μL of mixed internal standard (M2-IS) was added and the mixture performed vortex shaking at 1500 rpm for 3 min, followed by centrifugation at 3500 g for 20 min. Afterward, 180 μL of the supernatant was transferred to the microplates for analysis (3 μL).

M3: for organic acids and carbohydrates. The analysis of organic acids employed derivatization strategies[4]. In brief, 20 μL of samples were pretreated with 80 μL mixture of acetonitrile-methanol solution (V:V = 1:1), followed by a short vortex and centrifugation at 16000 g for 20 min to precipitate the protein. Then, 40 μL of the supernatant was transferred into another tube, and 5 μL of M3-IS, 20 μL of the derivatization reagent 1 and 20 μL of reagent 2 were added together into the bottom. Reagent 1 was prepared by adding 200 mM of 3-nitrophenylhydrazine hydrochloride (3-NPH) to a solution of acetonitrile and water in a volumetric ratio of 1:1. The preparation of Reagent 2 involved the addition of 240 mM 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide (EDC) and 6% (m/V) pyridine to the acetonitrile-water solution (V:V = 1:1). The obtained system was reacted at 40 °C for 30 min after vortex shaking and centrifugation for 30 s using miniSpin. After that, the mixture was placed immediately under −20 °C to terminate the reaction and dried using Vacuum Concentrator for 30 min. Then, the residue was redissolved in 800 μL of methanol-water solution (V:V = 1:1) and 2 μL of the supernatant was injected.

M4: for bacteria-derived metabolites. To determine the levels of bacteria-derived metabolites, 50 μL of samples or standards were firstly transferred into the 750 μL 96-well sample collection plates. Then, 190 μL of the extraction reagent and 10 μL of M4-IS was added, followed by vortex at 1500 rpm for 3 min. Then, the mixture in the 96-well plate was centrifuged at 3500 g for 20 min and both 120 μL of the supernatant and ultra-pure water were transferred to the 450 μL Nunc™ microWell™ 96-Well microplates. Finally, 200 μL of the supernatant was transferred to the microplates for analysis (5 μL).

M5: for acyl carnitines and lysophospholipids. 50 μL of samples or standards, 50 μL of M5-IS and 150 μL of the extraction reagent methanol was transferred and added into the 750 μL 96-well sample collection plates. After vortex at 1500 rpm for 3 min and centrifugated at 3500 g for 20 min under 4 °C, 180 μL of the supernatant was transferred to the 96-Well microplates (injection volume: 5 μL).

M6: lipids. 20 μL of serum samples or QC samples were accurately transferred to the corresponding centrifuge tube, followed by adding 120 μL of methanol containing mixed M6-IS and vortex at 1500 rpm for 3 min. Subsequently, 100 μL of water and 360 μL of MTBE were added and vortex at 1500 rpm for 5 min. After placed at 4 °C for 3 min, the mixture was centrifuged at 2000 g for 5 min, in which 300 μL of upper layer was transferred. Then, 360 μL MTBE was additionally added to the residue, and repeat the extracted operation, followed by the upper layer transferred to the same tube or plate. After dried under vacuum for 1 h, 80 μL of isopropanol-acetonitrile-hexane (5/5/2, V/V/V) was used as the redissolved solution. Finally, 2 μL of supernatant was injected for analysis.

Instrument parameters for LC-MS/MS analysis. All the detection for metabolomics was performed on the LC-20ADXR HPLC series system (Shimadzu, Japan) coupled with QTRAP hybrid triple quadrupole mass spectrometer (5500+, AB Sciex, Singapore) except special notes. The metabolites were ionized by a TurboV™ heated electrospray ionization source and scheduled multiple reaction monitoring (MRM) was conducted in both positive and negative modes with detail information for instrument setting parameters recorded on Science Data Bank (Dataset 13920, Table 2)[26]. The optimal parameters of mass spectrometer were as follows: curtain gas, 35 psi; ion source gas 1, 50 psi; ion source gas 2, 50 psi; ionSpray voltage in negative mode, −4.5 kV; ionSpray voltage in positive mode, 5.5 kV; ion source temperature 550 °C.

LC parameters for targeted metabolomic analysis. *M1: for amine acids and nucleosides.* Amine acids and nucleosides were separated on an Acquity UPLC BEH Amide column (1.7 μm, 2.1 mm × 100 mm, Waters). The mobile phase was consisting of acetonitrile-water solution (V:V = 9:1) as phase A, and acetonitrile-water solution (V:V = 5:5) as phase B with 10 mM ammonium formate and 0.1% (m/V) formic acid as additives in both phases. The binary gradient condition was set as follows: 0–0.5 min at 0% B, 0.5–12.0 min from 0 to 40%,

12.0–15.0 min from 40% to 70%, 15.0–16.0 min at 70%, 16.0–17.0 min from 70% to 0%B with additional 3 min for re-equilibration. The column was kept at 50 °C and the total flow rate was set at 0.3 mL/min.

M2: for bile acids and fatty acids.　The separation was performed using a Hypersil GOLD column (1.9 μm, 2.1 mm × 100 mm, Thermo) at a column temperature of 50 °C. The mobile phase consisted of 2 mM ammonium acetate in water (A), and acetonitrile (B). The binary gradient condition was optimized as follows: 0–0.5 min at 17% B, 0.5–12.0 min from 17% to 30%, 12.0–15.5 min from 30% to 55%, 15.5–16.5 min at 55%, and 24.5–27.0 min at 95% with another 3 min for equilibration with total flow rate set at 0.4 mL/min.

M3: for organic acids and carbohydrates.　A typical C_{18} column namely ACQUITY UPLC® HSS C_{18} column (1.8 μm, 2.1 mm × 100 mm, Waters) was conducted for the analysis of the derivatives of carbohydrate. The mobile phase consisted of 0.01% (m/V) formic acid in water (A), and acetonitrile (B). The gradient condition for phase B was as follows: 0–1.0 min at 15%, CTO.RVR value at 0; 1.0–10.0 min at 15%, CTO.RVR value at 1; 16.0–23.0 min from 40% to 95%; 23.0–26.0 min, at 95%, 26.0–27.0 min from 95% to 15% with the column temperature kept at 40 °C. The total flow rate was set at 0.3 mL/min.

M4: for bacteria-derived metabolites.　Bacteria-derived metabolites were separated on a PFP C_{18} column (2 μm, 2.1 mm × 100 mm, ACE). The mobile phase was consisting of 0.01% (m/V) formic acid-water solution (phase A) and acetonitrile (phase B). The binary gradient condition was set as follows: 0–1.0 min at 5% B, 1.0–9.0 min from 5 to 75%, 9.0–9.5 min from 75% to 95%, 9.5–12.0 min at 95%, 12.5 min at 5%B with 2.5 min for re-equilibration and the column was maintained at 40 °C. The total flow rate was set at 0.3 mL/min.

M5: for acyl carnitines and lyso-phospholipids.　The separation for acyl carnitines and lyso-phospholipids was performed using a PFP C_{18} column (2 μm, 2.1 mm × 100 mm, ACE) on the LC-20ADXR HPLC series system coupled with QTRAP hybrid triple quadrupole mass spectrometer. The flow rate was set at 0.4 mL/min with the column temperature kept at 50 °C. The mobile phase consisted of water (A) and acetonitrile (B) with 0.1% formic acid as additives in both phases. The binary gradient condition was optimized as follows: 0–0.5 min at 2% B, 0.5–12.0 min from 2% to 98%, 12.0–13.5 min at 98%, 13.60 min at 2% with another 2 min for equilibration.

M6: for lipids.　Lipids in different structures were separated on a ACQUITY UPLC® HSS C_{18} column (1.8 μm, 2.1 mm × 100 mm, Waters) and detected by QTRAP hybrid triple quadrupole mass spectrometer (4500 MD). Mobile phase A was acetonitrile/water (60/40, V/V) with 10 mM ammonium acetate and mobile phase B was isopropanol/acetonitrile (90/10, V/V) containing 10 mM ammonium acetate. The binary gradient condition was set as follows: 0–2.0 min at 20% B, 2.0–5.0 min from 20 to 70%, 5.0–17.0 min from 70% to 95%, 17.0–17.5 min at 95%, 17.6 min at 20%B with 2.4 min for re-equilibration. The column was kept at 50 °C and the flow rate was set at 0.3 mL/min.

Sampling sequence.　Nunc™ microWell™ 96-Well microplates (450 μL or 200 μL) were utilized for automatic sampling. Following the injection of standard curve, each sample within the batch was given a randomized number as the injection sequence and a QC sample was inserted after every 20–30 samples on each 96-Well plate. The entire process of sample analysis commenced with quality control (QC) injections and concluded with QC samples to assess the measurement variation and detect the flawed integration resulting from inferior peak forms, which may exhibit systematic upward or downward trends in area[27].

Data extraction and processing.　The mass chromatogram peaks of metabolites were extracted and integrated using the in-house software OS or MultiQuant (AB Sciex, Singapore). Firstly, each peak was automatically identified based on the retention time and MRM parameters of standards. The software parameters were configured to their default values, including a minimum peak width of 3 points and peak splitting at 2 points, etc. Any peak with a signal-to-noise ratio below 3 was excluded from the analysis. In addition, an overall afterward manual check was conducted to confirm that all peak identifications are consistent, and the automatic integrations are based on uniform peak shapes. A linear regression model incorporating a weighting factor of 1/x was applied to each metabolite in the standard curve and the quantitative results were subsequently calculated using the standard curve after correction with the isotope-labelled internal standards.

The metabolites with missing values exceeding 50% of the total sample size were excluded from the analysis during the data pre-processing stage to ensure the analytical accuracy. Subsequently, we employed the K-Nearest-Neighbor (KNN) Algorithm to estimate and fill the remaining sporadic missing values for further quality control. Quality control assessment was performed on our established omics platform namely IPOS (http://82.157.20.231:3838/ipos/), and data was processed by SERRF normalization to avoid systematic errors for samples in large scale[28].

Methodological validation for M1-M6.　The validation protocol for M1-M5 encompassed six key aspects, namely the linear range, accuracy and precision, carryover, spiked recovery and stability assessment. Linear range involved utilizing matrix-free standards with consideration given to instrument sensitivity as well as each level of molecule in serum. Least square regression analysis with an internal standard was employed to assess the linearity. And accuracy and precision were evaluated using working solutions at middle level. Carryover was calculated by ratio of the peak area in blank solvent after consecutive injections of upper limit of quantitation (ULOQ) versus the area in lower limit of quantitation (LLOQ) samples. Spiked recovery was conducted by measuring the observed increased concentration of metabolite after spiking known levels in serum samples and calculated

as the ratio counts (in two different levels). Stability data mainly included the short-term stability after sample preparation.

For M6, considering the limitation of obtaining all the lipid standards, pooled human serum samples from healthy subjects were utilized as quality control (QC) to ensure the reproducibility. Carryover effects were assessed by injecting blank solvent replicates sequentially after serum extracts. And short-term autosampler stability was evaluated by subjecting QC samples to extraction and subsequent analysis after placed in the autosampler at 4 °C for 24 h.

Untargeted metabolomic analysis. The untargeted metabolomics analysis was performed using an Ultimate 3000 ultra-high performance liquid chromatography (UHPLC) coupled with a Q Exactive Quadrupole-Orbitrap High-Resolution Mass Spectrometer (Thermo Fisher Scientific). A C_{18}-PFP column (1.8 μm, 2.1 mm × 100 mm, ACE) was employed for the separation in positive mode by eluting with 0.1% (m/V) formic acid-water solution and acetonitrile. For the negative detection mode, the metabolites were separated on an Acquity HSS C_{18} column (Waters, 1.8 μm, 2.1 mm × 100 mm), and mobile phases containing additives of ammonium bicarbonate at a concentration of 400 mg/L were used to adjust the pH value during elution[29].

The gradient was programmed as follows: initially, phase B was ramped from 2% to 100% over a period of 10 minutes, followed by a 5-minute washing and equilibration step. The flow rate, injection volume, and column temperature were maintained at 0.4 mL/min, 5 μL, and 50 °C respectively in both conditions[30]. The biosignatures were analyzed using full scan mode with a resolution of 70,000 FWHM and a scan range of m/z 70–1,000. Other ionization parameters included sheath gas (45 arb), aux gas (10 arb), heater temperature (355 °C), capillary temperature (320 °C), and S-Lens RF level (55%).

In brief, 150 μl of serum sample was placed in a 1 mL 96-Deep well plate (Thermo, Delaware, USA), and then 600 μL of the extract reagent (a mixture of methanol and acetonitrile with a volume ratio of 1:1) was added to each sample. After gentle vortexing and centrifugation, two aliquots of the supernatant (200 μL each) were transferred to a 450 μL 96-well plate (Thermo). The samples were subsequently concentrated and dried for 2 h. To redissolve the dried residue, 80 μL of methanol-water solution at a concentration ratio of 50% was used. Detection was performed in both positive and negative ion modes. Additionally, the remaining upper layers were collected to prepare quality control (QC) samples.

The peak area under curve for the untargeted analysis as quantitative information was obtained with TraceFinder software (Thermo Fisher Scientific), followed by normalization. And the annotations for the metabolites referred to the self-established spectrum library, with the high-resolution mass error within ± 5 ppm and retention time limited. The untargeted metabolomic data were merged and the metabolites with variables exceeding 50% of the total missing values and relative standard deviation (RSD) of quality controls above 30% were excluded.

Data Records

The integrated metabolite counts, sample preparation steps, detailed parameters, annotations and corresponding clinical data for all the samples are available for download on Metabolomics Workbench at https://www.metabolomicsworkbench.org (Project ID PR002021) and Science Data Bank at https://www.scidb.cn (CSTR 31253.11.sciencedb.13920). The data can be accessed directly via it's Project DOI: 10.21228/M81V7G (for data) and DOI: 10.57760/sciencedb.13920 (for method)[26,31]. The raw data for all the samples are provided in a zip format (ST003390_YLab_T2DM_Raw_Dataset.zip), which could be downloaded as a whole. The dataset contains both the raw metabolomics data (a folder of six pairs of wiff. and wiff.scan. files for M1-M6 and the converted raw data files in mzML format for each sample) and the absolute quantitative data (YLab_T2DM_Metabolomics Data.xlsx), the post-processed data with missing values filled (YLab_T2DM_Metabolomics Post-Processed Data.xlsx), the differential analysis results (YLab_T2DM_Differential Analysis of the Targeted Metabolomics for T2DM), the clinical information data (YLab_T2DM_Clinical Baseline Information.xlsx), the mass parameters for the metabolites (YLab_Mass Parameters in M1-M6.xlsx), as well as a document (YLab_Sample Preparation Protocols.docx) outlining the sample processing procedures[31].

By utilizing the comprehensive targeted LC-MS platform, a total of 1609 small molecules in biofluids, spanning across 10 major molecular classes and 32 sub-classes (as depicted in Fig. 2), were subjected to analysis through the integration of six distinct analytical conditions (M1-M6). M1 employed a normal phase separation mode for the analysis of 94 amino acids, 23 nucleosides, and 15 monoamines as well as the derivatives. M2 utilized a reversed phase separation method for the examination of 53 fatty acids and 57 bile acids and conjugates. The method M3 involved a pre-column chemical derivatization technique for the analysis of 85 organic acids and carbohydrates. M4 employed a fluorophenyl modified reversed phase separation mode to analyze the bacterial-derived metabolites, including 91 compounds in other types that have been reported bioactive and partially or completely produced by gut microbiota based on literature reviews[32–40]. Apart from this, another two reversed phase separation methods (M5-M6) in strong elution mode were conducted to analyze acylcarnitines (64 compounds), lysophospholipids (140 compounds), sphingomyelin (55 compounds), phospholipids (290 compounds), ceramides (66 compounds), glycerides (526 compounds), and other lipids (116 compounds).

The pre-processed data were generated following the aforementioned procedures, and undergoing additional quality control with the accompanying QCs in the batch. Then, the data by different methods (M1-M6) was merged and uploaded as a whole file (YLab_T2DM_Metabolomics Data.xlsx)[31]. The overall differential analysis results were summarized in the uploaded file (YLab_T2DM_Differential Analysis of the Targeted Metabolomics for T2DM)[31], indicating an obvious metabolic disturbance in amino acid, fatty acids, lysophosphatidyl-choline (LysoPC), and triacylglycerol (TAG) under T2DM. To protect the privacy of the participants, each individual name was represented by a numerical identifier and detailed descriptions can be found in the "Legend" tab on the website.
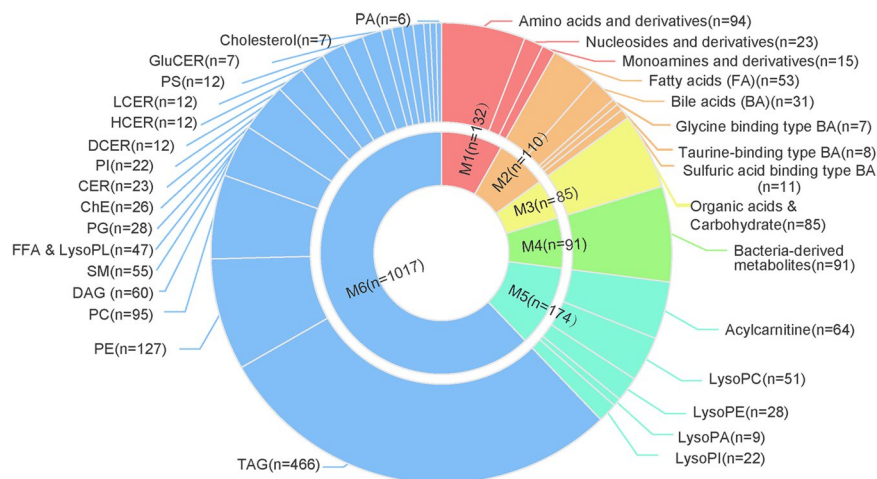
Fig. 2 The detecting lists for the targeted metabolomic platform (M1–M6), containing a total of 1609 compounds, spanning across 10 major molecular classes and 32 sub-classes. PC, phosphatidylcholine; PE, phosphatidylethanolamine; PA, phosphatidic acid; PI, phosphatidylinositol; TAG, triacylglycerol; DAG, diacylglycerol; SM, sphingomyelin; FFA, free fatty acids; PG, phosphatidylglycerol; ChE, cholesterol ester; CER, ceramide; DCER, dihydroceramide; HCER, hydroxylceramide; LCER, lactosylceramide; GluCER, glucosylceramide; PS, phosphatidylserine; PL, phospholipid.
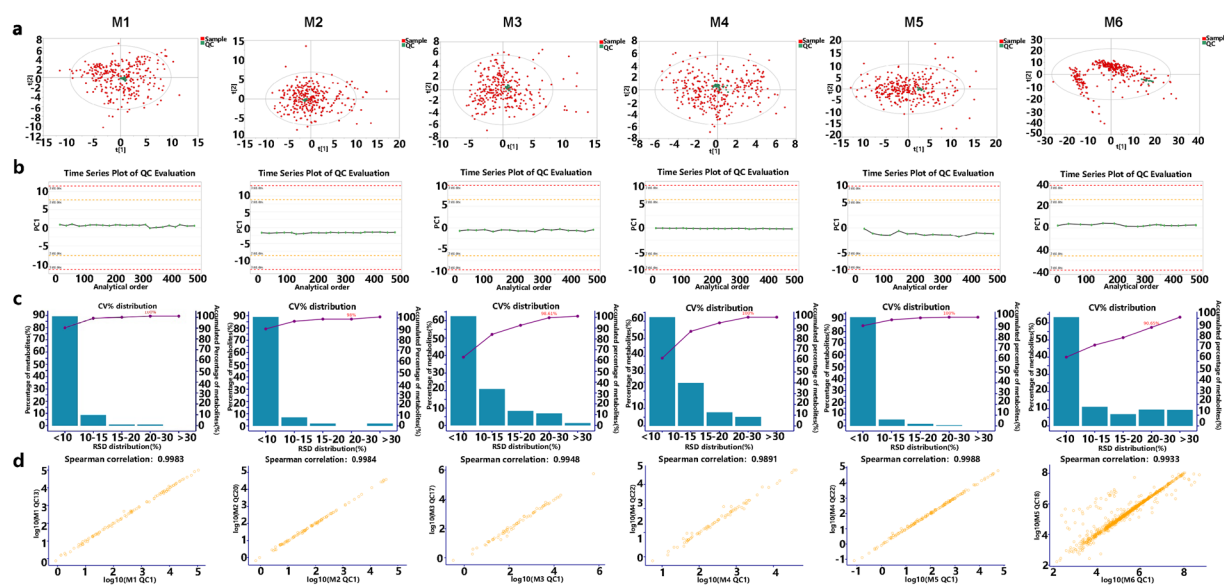
Fig. 3 Technical validation data for QC samples in methods M1-M6. (**a**) Principal components analysis (PCA) for both quality controls (QCs) and cohort samples. (**b**) The stability of the first principal component of QCs over time. (**c**) Distribution graph of relative variation coefficient for the analytes in QCs. (**d**) Spearman correlation analysis between the first and last QC data.

Technical Validation

After completing the preprocessing steps for the entire batch, Principal components analysis (PCA) was conducted including both quality controls (QCs) and samples. The score plot for each method is illustrated in Fig. 3a, where it can be observed that QCs tend to cluster together. The stability of the first principal component of QCs over time is demonstrated in Fig. 3b. Furthermore, an analysis of distribution trends confirms that more than 98% of all analytes exhibit a relative variation coefficient RSD (%) below 25% in response within QC samples (Fig. 3c). Additionally, Spearman correlation analysis between the first and last QC data reveals an exceptionally high degree of correlation as anticipated (Fig. 3d). These findings collectively indicate the feasibility and stability of the methods.

Apart from the accompanying QC samples for quality control, the technical validation for targeted LC-MS/MS methods encompasses six key aspects, namely linear range, accuracy and precision, carryover, spiked recovery and stability assessment. Detail validation results are summarized and uploaded on Science Data Bank as
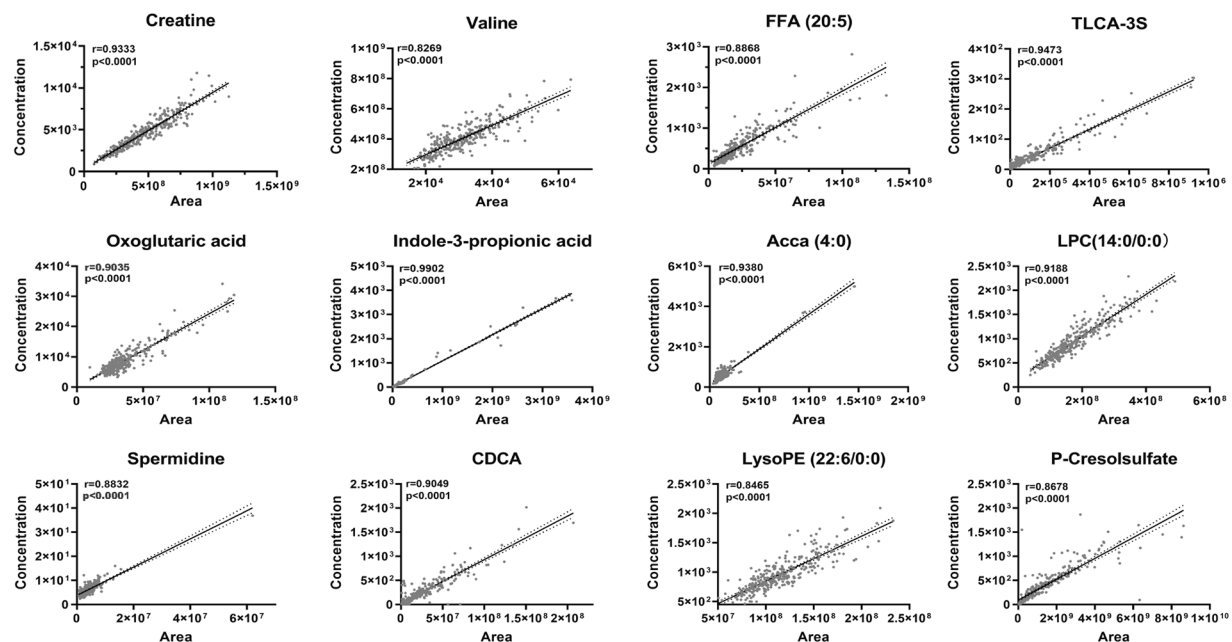
Fig. 4 The Passing Bablok regression scatter plots of represented metabolites by targeted platform (Y axis) and non-targeted metabolomics platform (X axis), respectively.
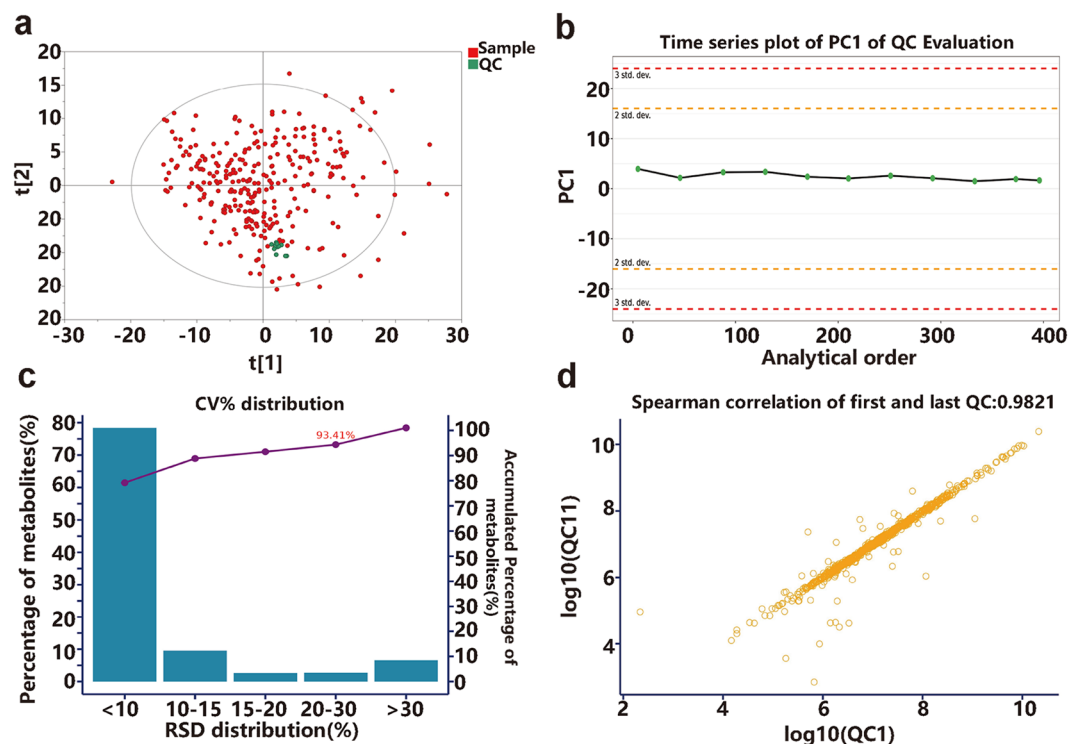

Fig. 5 Technical validation data for QC samples by the un-targeted metabolomics platform. (**a**) Principal components analysis (PCA) for both quality controls (QCs) and cohort samples. (**b**) The stability of the first principal component of QCs. (**c**) Distribution graph of relative variation coefficient for the analytes in QCs. (**d**) Spearman correlation analysis between the first and last QC data.

Dataset 13920-Table 3 (Technical validation results for M1-M6.xlsx)[26]. Finally, the targeted quantification results were cross-validated with additional un-targeted metabolomics test results using Passing Bablok regression scatter plots (in Fig. 4), with the quality control results for the un-targeted method were shown in Fig. 5. The analysis of representative metabolites consistently demonstrated the excellent quality of data. And all these findings

above provide compelling evidence for the analytical reliability and exceptional performance of our methods following meticulous correction of isotope labelled internal standards.

Code availability

Metabolomics data analysis was conducted based on R Studio (4.3.1), where tidyr (1.3.1) and dplyr (1.1.4) were employed for untargeted metabolomics data cleaning, the impute (1.74.1) was utilized for missing value imputation, and FactoMineR (2.11) were used for PCA, PLS (-DA), and OPLS (-DA) analysis. Statistical analysis was performed using the stats (4.3.1). The codes are available at https://cran.r-project.org/web/packages/available_packages_ by_name. html.

References

1. Baker, S. A. & Rutter, J. Metabolites as signalling molecules. *Nat Rev Mol Cell Biol.* **24**, 355–374 (2023).
2. Haas, R. *et al*. Intermediates of Metabolism: From Bystanders to Signalling Molecules. *Trends Biochem. Sci.* **41**, 460–471 (2016).
3. Qiu, S. *et al*. Small molecule metabolites: discovery of biomarkers and therapeutic targets. *Signal Transduct. Target. Ther.* **8**, 132 (2023).
4. Xie, G. *et al*. A Metabolite Array Technology for Precision Medicine. *Anal. Chem.* **93**, 5709–5717 (2021).
5. Aderemi, A. V., Ayeleso, A. O., Oyedapo, O. O. & Mukwevho, E. Metabolomics: A Scoping Review of Its Role as a Tool for Disease Biomarker Discovery in Selected Non-Communicable Diseases. *Metabolites* **11**, 418 (2021).
6. Rinschen, M. M., Ivanisevic, J., Giera, M. & Siuzdak, G. Identification of bioactive metabolites using activity metabolomics. *Nat. Rev. Mol. Cell Biol.* **20**, 353–367 (2019).
7. Newgard, C. B. Metabolomics and Metabolic Diseases: Where Do We Stand? *Cell Metab.* **25**, 43–56 (2017).
8. Zhang, A., Yan, G., Han, Y. & Wang, X. Metabolomics approaches and applications in prostate cancer research. *Appl. Biochem. Biotechnol.* **174**, 6–12 (2014).
9. Zhang, X. *et al*. Mass spectrometry-based "omics" technologies in cancer diagnostics. *Mass Spectrom. Rev.* **26**, 403–431 (2007).
10. Dettmer, K., Aronov, P. A. & Hammock, B. D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **26**, 51–78 (2007).
11. Zhang, A., Sun, H., Wang, P., Han, Y. & Wang, X. Modern analytical techniques in metabolomics analysis. *Analyst* **137**, 293–300 (2012).
12. Serafimov, K. & Lammerhofer, M. Metabolic profiling workflow for cell extracts by targeted hydrophilic interaction liquid chromatography-tandem mass spectrometry. *J. Chromatogr. A* **1684**, 463556 (2022).
13. Shikshaky, H. *et al*. A Novel Approach of SWATH-Based Metabolomics Analysis Using the Human Metabolome Database Spectral Library. *Int. J. Mol. Sci.* **23**, 10908 (2022).
14. van Bentum, M. & Selbach, M. An Introduction to Advanced Targeted Acquisition Methods. *Mol. Cell. Proteomics* **20**, 100165 (2021).
15. Totsch, K., Fjeldsted, J. C., Stow, S. M., Schmitz, O. J. & Meckelmann, S. W. Effect of Sampling Rate and Data Pretreatment for Targeted and Nontargeted Analysis by Means of Liquid Chromatography Coupled to Drift Time Ion Mobility Quadruple Time-of-Flight Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **32**, 2592–2603 (2021).
16. Griffiths, W. J. *et al*. Targeted metabolomics for biomarker discovery. *Angew. Chem. Int. Ed. Engl.* **49**, 5426–5445 (2010).
17. Lu, W., Bennett, B. D. & Rabinowitz, J. D. Analytical strategies for LC-MS-based targeted metabolomics. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **871**, 236–242 (2008).
18. Yuan, M., Breitkopf, S. B., Yang, X. & Asara, J. M. A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. *Nat. Protoc.* **7**, 872–881 (2012).
19. Wei, R., Li, G. & Seymour, A. B. High-throughput and multiplexed LC/MS/MRM method for targeted metabolomics. *Anal. Chem.* **82**, 5527–5533 (2010).
20. Yan, Z. & Yan, R. Increase the accessibility and scale of targeted metabolomics: Construction of a human urinary metabolome-wide multiple reaction monitoring library using directly-coupled reversed-phase and hydrophilic interaction chromatography. *Anal. Chim. Acta.* **894**, 65–75 (2015).
21. Gu, H., Zhang, P., Zhu, J. & Raftery, D. Globally Optimized Targeted Mass Spectrometry: Reliable Metabolomics Analysis with Broad Coverage. *Anal. Chem.* **87**, 12355–12362 (2015).
22. Chen, Y. *et al*. Development of a Data-Independent Targeted Metabolomics Method for Relative Quantification Using Liquid Chromatography Coupled with Tandem Mass Spectrometry. *Anal. Chem.* **89**, 6954–6962 (2017).
23. Wang, L. *et al*. Ion-Pair Selection Method for Pseudotargeted Metabolomics Based on SWATH MS Acquisition and Its Application in Differential Metabolite Discovery of Type 2 Diabetes. *Anal. Chem.* **90**, 11401–11408 (2018).
24. Zha, H. *et al*. SWATHtoMRM: Development of High-Coverage Targeted Metabolomics Method Using SWATH Technology for Biomarker Discovery. *Anal. Chem.* **90**, 4062–4070 (2018).
25. American Diabetes Association. Introduction: Standards of Medical Care in Diabetes-2022. *Diabetes Care* **45**, S1–S2 (2022).
26. Ma, S. R. & Yin, P. Y. Dataset of the comprehensive description and methodological validation for an in-depth targeted metabolomics platform developed by Yin Lab. *Science Data Bank* https://doi.org/10.57760/sciencedb.13920 (2024).
27. Benedetti, E. *et al*. Plasma metabolomics profiling of 580 patients from an Early Detection Research Network prostate cancer cohort. *Sci. Data* **10**, 830 (2023).
28. Fan, S. *et al*. Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data. *Anal. Chem.* **91**, 3590–3596 (2019).
29. Luo, P. *et al*. A Large-scale, multicenter serum metabolite biomarker identification study for the early detection of hepatocellular carcinoma. *Hepatology* **67**, 662–675 (2018).
30. Liu, J. *et al*. Integrative metagenomic and metabolomic analyses reveal the potential of gut microbiota to exacerbate acute pancreatitis. *NPJ Biofilms. Microbiomes* **10**, 29 (2024).
31. Ma, S. R. In-depth profiling of biosignatures for Type 2 diabetes mellitus cohort utilizing an integrated targeted LC-MS platform. *Metabolomics Workbench* https://doi.org/10.21228/M81V7G (2024).
32. Platten, M., Nollen, E. A., Röhrig, U. F., Fallarino, F. & Opitz, C. A. Tryptophan metabolism as a common therapeutic target in cancer, neurodegeneration and beyond. *Nat. Rev. Drug Discov.* **18**, 379–401 (2019).
33. Yano, J. M. *et al*. Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell.* **161**, 264–76 (2015).
34. Chen, H. *et al*. A forward chemical genetic screen reveals gut microbiota metabolites that modulate host physiology. *Cell.* **177**, 1217–1231 (2019).
35. Krautkramer, K. A., Fan, J. & Bäckhed, F. Gut microbial metabolites as multi-kingdom intermediates. *Nat. Rev. Microbiol.* **19**, 77–94 (2021).

36. Yang, W. & Cong, Y. Gut microbiota-derived metabolites in the regulation of host immune responses and immune-related inflammatory diseases. *Cell Mol. Immunol.* **18**, 866–877 (2021).
37. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55–71 (2021).
38. Zhao, Q. *et al*. Polyamine metabolism links gut microbiota and testicular dysfunction. *Microbiome* **9**, 224 (2021).
39. Bermudez-Martin, P. *et al*. The microbial metabolite p-Cresol induces autistic-like behaviors in mice by remodeling the gut microbiota. *Microbiome* **9**, 157 (2021).
40. Zhao, M. *et al*. Gut bacteria-driven homovanillic acid alleviates depression by modulating synaptic integrity. *Cell Metab.* **36**, 1000–1012 (2024).

Acknowledgements

Author contributions

P.Y.Y. and J.L.D. performed the study design and patient recruitment. P.Y.Y. and S.R.M., and J.W.L. wrote the primary manuscript. L.Y., J.W.L. and Y.S.Z. recruited the patients and extracted the clinical annotations for this paper. P.Y.Y., Z.M.W., S.R.M. and Q.Y.L. established the metabolomics analysis methods. J.W.L, S.C., M.M.W., T.F.W. and A.L.H. performed the data extraction and J.W.L. and S.R.M. performed the data processing and analysis. All authors approved the final version of the manuscript.

Competing interests

Z.M.W. and P.Y.Y. are cofounders of IPhenome Biotechnology (Yunpukang) Inc, Dalian. Other authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.D. or P.Y.

Reprints and permissions information is available at www.nature.com/reprints.

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