Clinical Proteomics

RESEARCH

Open Access

Combined urine proteomics and metabolomics analysis for the diagnosis of pulmonary tuberculosis

Jiajia Yu[1,2], Jinfeng Yuan[2], Zhidong Liu[2], Huan Ye[2], Minggui Lin[3], Liping Ma[2], Rongmei Liu[2], Weimin Ding[2], Li Li[3], Tianyu Ma[2], Shenjie Tang[2] and Yu Pang[2*]

Abstract

Background Tuberculosis (TB) diagnostic monitoring is paramount to clinical decision-making and the host biomarkers appears to play a significant role. The currently available diagnostic technology for TB detection is inadequate. In the present study, we aimed to identify biomarkers for diagnosis of pulmonary tuberculosis (PTB) using urinary metabolomic and proteomic analysis. Methods: In the study, urine from 40 PTB, 40 lung cancer (LCA), 40 community-acquired pneumonia (CAP) patients and 40 healthy controls (HC) was collected. Biomarker panels were selected based on random forest (RF) analysis. Results: A total of 3,868 proteins and 1,272 annotated metabolic features were detected using pairwise comparisons. Using AUC ≥ 0.80 as a cutoff value, we picked up five protein biomarkers for PTB diagnosis. The five-protein panel yielded an AUC for PTB/HC, PTB/CAP and PTB/LCA of 0.9840, 0.9680 and 0.9310, respectively. Additionally, five metabolism biomarkers were selected for differential diagnosis purpose. By employment of the five-metabolism panel, we could differentiate PTB/HC at an AUC of 0.9940, PTB/CAP of 0.8920, and PTB/LCA of 0.8570. Conclusion: Our data demonstrate that metabolomic and proteomic analysis can identify a novel urine biomarker panel to diagnose PTB with high sensitivity and specificity. The receiver operating characteristic curve analysis showed that it is possible to perform non-invasive clinical diagnoses of PTB through these urine biomarkers.

Keywords Proteomics, Metabolomics, Tuberculosis, Biomarkers, Urine

Introduction

Tuberculosis (TB) is a communicable disease caused by *Mycobacterium tuberculosis* (MTB) infection and a major cause of human illness and mortality. Worldwide, an estimated 10.6 million people (95% UI: 9.9–11.4 million) developed TB in 2022, up from best estimates of 10.3 million in 2021 and 10.0 million in 2020 [1]. Clinically, TB patients most commonly present to the clinic with pulmonary tuberculosis (PTB), a form of TB with enhanced communicability that can be difficult to diagnose due to its multi-systemic features and protean presentation. To address this issue, rapid and accurate

*Correspondence:
Yu Pang
pangyupound@163.com
[1]Department of Infectious Diseases and Clinical Microbiology, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, People's Republic of China
[2]Department of Bacteriology and Immunology, Beijing Chest Hospital, Beijing Tuberculosis and Thoracic Tumor Research Institute, Capital Medical University, Beijing 101149, China
[3]School of Clinical Medicine, Beijing Tsinghua Changgung Hospital, Tsinghua University, Beijing 102218, China

diagnostics are required to achieve early point-of-care detection. Current TB diagnostic methods are hampered by challenges that either compromise their accuracy or hinder their widespread use, especially in resource-limited settings [2]. For example, conventional gold standard TB-diagnostic methods (i.e., sputum microscopy and microbiologic culture) are accessible but lack sensitivity [1, 2]. By contrast, GeneXpert MTB/RIF offers rapid detection in this regard and has shown good sensitivity (97%) and correlation time to culture positivity but suffers from poor specificity ranging from 49 to 72% [3, 4]. For these reasons, improved tests are needed that can accurately detect additional host biomarkers associated with active TB disease, while also distinguishing TB from other respiratory illnesses with similar symptomatology.

Metabolomics, an emerging systems-level technology that enables unbiased, multiplexed metabolite profiling and comparative analysis of biological samples, can generate a fingerprint of all metabolites within a cellular system [5]. Meanwhile, another advanced systems-level approach, proteomics, has enabled the characterization of proteomes, the total protein contents of a cell, tissue, or organism. Within the domain of clinical infectious disease management, metabolomics and proteomics have proven to be invaluable assets, enabling early diagnosis, precise prognosis, and effective monitoring of disease progression [5, 6]. Urine collection is both less invasive and more convenient than collection of blood and other biological fluids. Furthermore, urine differs from blood in that urine is not subject to homeostatic control and thus its composition mirrors even small systemic changes [7]. Nevertheless, although metabolomics and proteomics approaches have been used successfully for TB biomarker discovery in a few reported studies [8–11], a single omics-based approach may fall short of providing strongly predictive and reliable biomarkers for use in TB diagnosis and differential diagnosis. To address these issues, here we applied proteomics and metabolomics to test urine specimens for PTB-specific protein and metabolite changes in order to identify potential PTB-diagnostic biomarkers.

Methods

Study population

Urine specimens were collected from subjects who had provided written informed consent for participation in the study. Confirmed PTB inpatients of Beijing Chest Hospital, Capital Medical University ranging in age from 18 to 65 years were enrolled from April 2022 to April 2023. During the same period, healthy controls (HCs) and patients with lung cancer (LCA) or community-acquired pneumonia (CAP) were recruited in a 1:1:1 ratio. As shown in Fig. 1 (apply Biorender to make the figure), enrollment of four cohorts of participants resulted in the collection of 160 urine specimens that yielded metabolome profiles for PTB patients (n=40), LCA patients (n=40), CAP patients (n=40), and HC subjects (n=40). For proteomic analysis, urine specimens obtained from an equal number of participants from each group (n=15) were randomly selected for analysis.

Case definitions

The diagnostic criteria for TB (WS 288–2017) issued by the National Health Council of the People's Republic of China were our standard for the inclusion of PTB patients into the study: (1) positive result of sputum acid-fast bacilli (AFB) smear or culture assay; (2) positive result obtained from a molecular-based MTB-detection assay; (3) lung specimen-based pathological diagnosis of TB. A diagnosis of LCA was confirmed based on cytological or surgical histopathological findings. CAP was diagnosed based on detection of a lung infiltrate (appearing as a shadow on a chest radiograph) accompanied by at least one of the following symptoms: cough, sputum production, fever, dyspnea, and/or chest pain [12], as directed by the "Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults" [12]. The subjects in the HC group were healthy individuals. Patients received no drug treatments prior to urine collection. Subjects excluded from the study were those with HIV infection, viral hepatitis, neurodegenerative disorders, immunodeficiencies, those who were pregnant or breastfeeding or planning to become pregnant, those taking immunosuppressant medications, those who refused to sign the informed consent form, and those who participated in other clinical trials within one month prior to the study enrollment date.

Sample collection and processing

Random, midstream urine samples obtained from all participants (40 mL) were collected into 50-mL Falcon tubes and transported to the laboratory at 4–8℃. Coded urine samples were centrifuged at 5000 × g for 10 min at 4 °C then supernatants were aliquoted into multiple 2 mL cryovials and stored at -80 °C until needed for testing.

Untargeted metabolomic analysis of urine specimens

Metabolomics profiling analysis was conducted using a UPLC-ESI-Q-Orbitrap-MS system (UHPLC, Shimadzu Nexera X2 LC-30AD, Shimadzu, Japan) coupled with Q-Exactive Plus (Thermo Scientific, San Jose, USA). For liquid chromatography (LC) separation of samples, we used an ACQUITY UPLC® HSS T3 column and followed specific steps as described in the supplementary material.

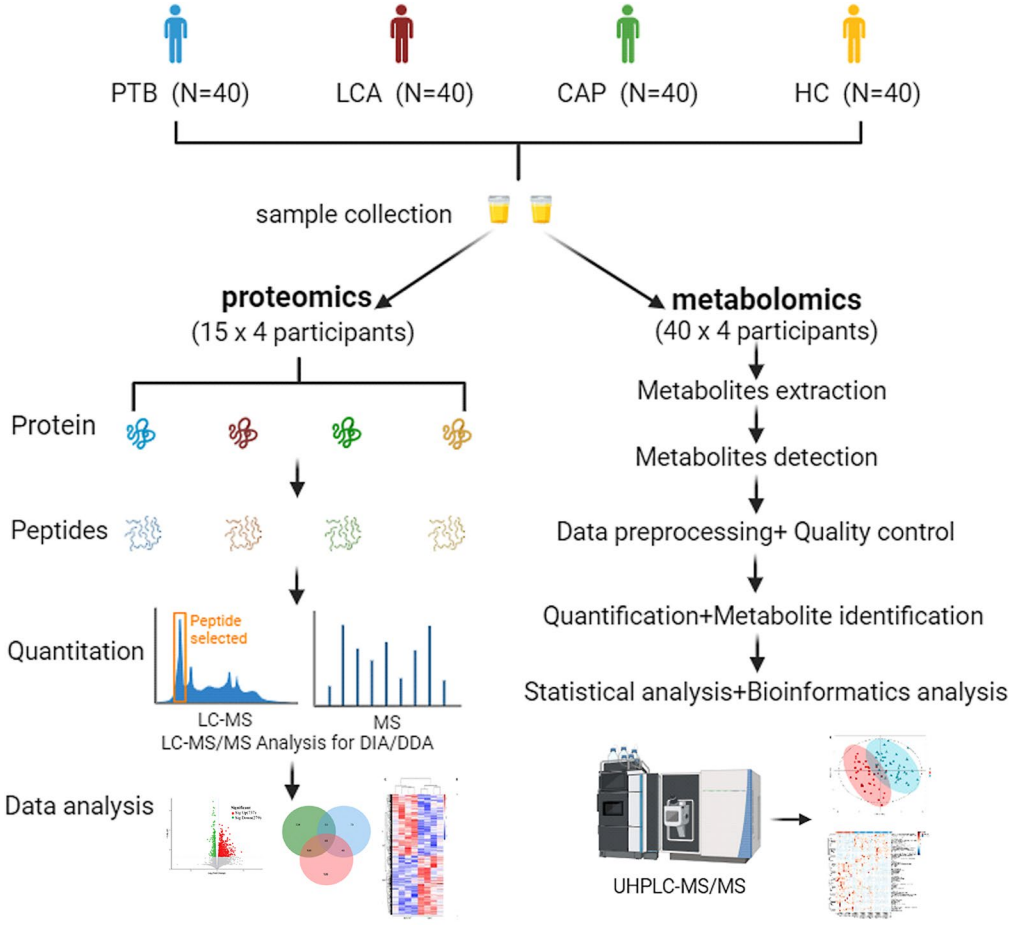

Fig. 1 Summary of the study design and cohort details. Urine samples were collected and subjected to proteomics and metabolomics. PTB: pulmonary tuberculosis; LCA: lung cancer; CAP: community-acquired pneumonia; HC: Healthy controls; UHPLC-MS/MS: Ultra-High Performance Liquid Chromatography-ESI-Q-Orbitrap-MS/MS; DDA: Data-dependent Acquisition; DIA: Data-independent Acquisition

Proteomic analysis of urine specimens

Proteins were extracted from urine samples using SDT lysis buffer (4% SDS, 100 mM DTT, 100 mM Tris-HCl pH 8.0) then digested using the filter-aided sample preparation (FASP) method [13]. The pooled peptide mixture prepared from each urine sample was fractionated using an Agilent 1260 high-performance liquid chromatography (HPLC) system equipped with a Waters XB ridge BEH C18 column; specific protein extraction and digestion steps and HPLC and tandem MS settings are described in the supplementary material.

Statistical analysis

Statistical analysis of clinical data

Clinical data were analyzed using IBM SPSS Statistics 24 and expressed as the mean ± SD. The baseline characteristics of the study population were statistically analyzed by Chi-square test, Kruskal-Wallis H test, and Mann-Whitney U test, followed by Bonferroni's multiple comparison test. P value < 0.05 indicated statistical significance.

Statistical and bioinformatic analysis

Bioinformatics analysis was carried out using Microsoft Excel and R statistical computing software. The data quality of metabolites in terms of homogeneity and reproducibility was evaluated by the principal component analysis (PCA). Then, the orthogonal partial least squares discriminant analysis (OPLS-DA) method was applied to remove irrelevant variables. Meanwhile, the variable importance in the projection (VIP) values were obtained from each variable to measure the contribution of variables to the model. The quality of the OPLS-DA was validated by the permutation test. The t-test was used to obtain P values of each individual variables, followed by the adjustment of false discovery rate (FDR) by multiple hypothesis tests. The bioinformatic analysis includes hierarchical clustering and volcano plot generation using the statistical programming language R, which is appropriate for omics data analysis.

Biomarkers were ranked using an ensemble feature selection (EFS)-based approach [14], which entailed use of applied multiple feature selection algorithms to avoid

biases associated with individual feature selection algorithms [15]. The EFS approach combines Mann-Whitney U tests, logistic regression, Pearson and Spearman correlations and two random forest algorithm implementations, *cforest* and *randomForest*, into a single, rankable score. Statistically significant, differentially abundant metabolites (DAMs) were identified based on the criteria of a FC (fold change) > 1.2 or < 0.83, VIP > 1, FDR value < 0.05, and removal of exogenous metabolites. Differentially abundant proteins (DAPs) were analyzed using MS stats with linear mixed-effects models [16]. DAPs that were deemed significant met conditions of a FC > 1.2 or < 0.83 and FDR value < 0.05.

Selection of biomarker candidates

To identify biomarkers, receiver operating characteristic (ROC) analysis was performed then the predictive power of each potential protein and metabolite biomarker was evaluated as based on its ROC area-under-the-curve (AUC) value. Next, three machine learning classifiers, including linear support vector machine (SVM), logistic regression (LR), and orange line for random forest (RF) were used to generate diagnostic models. The best diagnostic model was selected based on accuracy and error rates acquired using a ten-fold cross-validation method. Finally, ROC curves were employed to assess accuracy rates of biomarker combinations incorporating protein and metabolite candidate markers.

Results

Clinical characteristics of participants

Unbiased selection of patients was performed. There were 160 samples in total, with 40 samples in each group. All sample identities were renamed with codes instead of the patient's name and hospital number. The age and gender distribution of the four groups is roughly the same. Demographic and clinical characteristics of study participants are presented in Table 1.

Urine screening for DAPs

Ultimately, analysis of liquid chromatography-tandem mass spectrometry (LC-MS/MS) data obtained from the 60 urine samples of the four groups enabled the detection of 41,015 peptides and 3,868 proteins. After data were pre-processed and filtered to remove missing values, 2601 proteins were subjected to further analysis using various methods. Differential expression analysis was conducted to identify DAPs between PTB versus LCA, CAP, or HC groups (Supplementary Figure S1). Volcano plots were created to enable visualization of numbers of DAPs that were significantly down- and up-regulated (Fig. 2A-C). Based on pairwise two-sample *t*-test results, 88 DAPs were identified across the four groups (Fig. 2D, Supplementary Tables S1-S4). DAPs were grouped into four clusters based on protein expression patterns using Euclidean distance matrix hierarchical clustering (Fig. 2E).

To identify potential PTB-diagnostic protein markers in urine specimens obtained from PTB, LCA, CAP, and HC subjects, the EFS approach and Student's *t*-test were employed to aggregate and rank potential protein markers according to PTB-diagnostic accuracy (Table 1, Supplementary Table S5). The top 20 ranked DAPs for each of the four groups are listed in Fig. 2F-H. Use of the EFS approach minimizes the risk that top-ranked biomarkers are inter-correlated. Potentially reliable PTB-diagnostic urinary biomarkers were selected based on average EFS ranking, Student's *t*-test results, patterns of DAPs overlap between PTB and other groups, and clinical relevance. The final five protein biomarkers that were selected included carboxypeptidase B2 (CPB2), serine protease HTRA1 (HTRA1), beta-hexosaminidase (HEXA), filamin-A (FLNA), and neuropilin-1 (NRP1). Subsequently, each of the five protein biomarkers was evaluated for diagnostic potential using relative quantitative comparison and ROC curve analysis. The results demonstrated that all five proteins performed well as PTB-diagnostic biomarkers, as based on AUCs of individual proteins of ≥0.80 and *p* values of <0.01 (Supplementary Table S6, Supplementary Figure S2).

In order to assess the combined PTB-diagnostic performance of the five-biomarker set, we conducted ten-fold cross-validation on all pairwise predictions generated using three machine learning algorithms: SVM, RF, and

Table 1 Demographic characteristics of participants enrolled

					P value			
Characteristics	PTB($n=40$) [c]	LCA($n=40$) [c]	CAP($n=40$) [c]	HC($n=40$) [c]	LCA vs. PTB [a]	CAP vs. PTB [a]	HC vs. PTB [a]	All group [b]
Age, years(mean)	45.02 ± 10.62	48.20 ± 9.63	46.15 ± 11.58	44.17 ± 12.76	0.507	0.538	0.922	0.507
Sex (female)	26(65%)	24(60%)	17(45%)	22(55%)	0.646	0.074	0.364	0.316
BMI (kg^2/m)	21.61 ± 3.28	22.81 ± 3.01	22.82 ± 3.19	23.85 ± 2.71	0.695	0.564	0.011	0.021
Sputum Acid Fast Bacilli Smear (+/-)	15/25	NA	NA	NA	NA	NA	NA	NA
Sputum Culture (+/-)	22/18	NA	NA	NA	NA	NA	NA	NA
GeneXpert/MTB RIF assay (+/-)	18/12	NA	NA	NA	NA	NA	NA	NA

PTB: pulmonary tuberculosis; LCA: lung cancer; CAP: community-acquired pneumonia; HC: Healthy controls; IQR, interquartile range; SD, standard deviation; BMI: Body Mass Index. [a] Kruskal-Wallis test. [b] Mann–Whitney U test (continuous variables), Chi-square test (categorical variables). [c] Data are mean ± SD for the continuous variables, number (%) for categorical variables

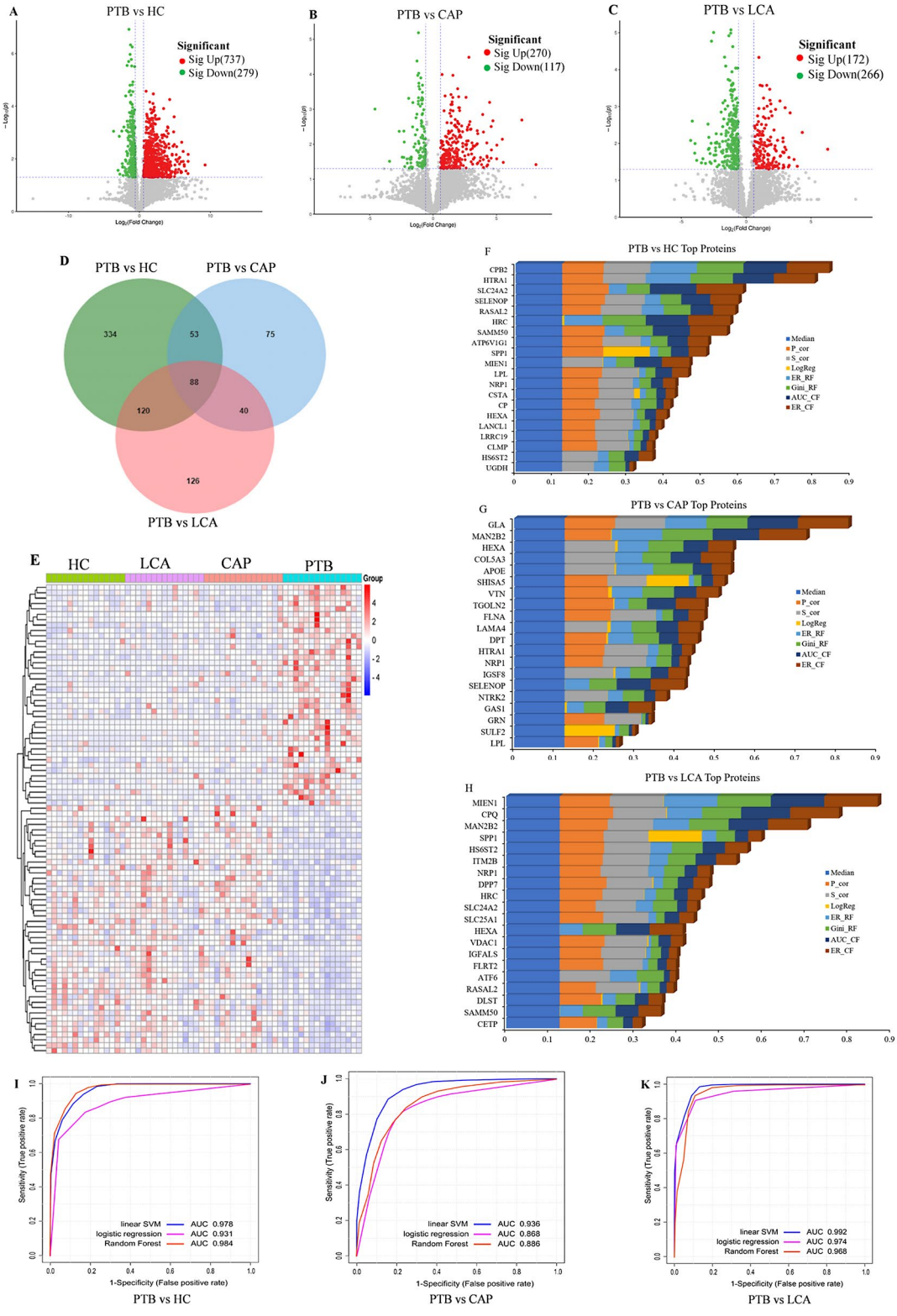

Fig. 2 Ensemble feature selection analysis of proteomics data and ROC analysis of potential biomarkers among PTB, CAP, LCA and HC groups (**A**-**C**) Volcano plots of the proteomics data for each pairwise comparison. The x-axis is log2 fold-change and the y-axis represents the minus log10p-value. Number of significantly down- (green) and up- (red) regulated proteins are shown on top. (**D**) Venn diagram of the number of DAPs. (**E**) Heatmap for biomarker detection from the proteomics data. Each row represents a protein marker, and each column represents a loading sample across the four groups, including PTB, LCA, CAP, and HC. Protein markers are grouped into four clusters via hierarchical clustering. (**F**-**H**) EFS analysis for Group PTB/HC, PTB/CAP and PTB/LCA. (**I**-**K**) Receiver operating characteristic (ROC) curves for each pairwise prediction by three different machine learning methods. Blue line for linear support vector machine (SVM), purple line for logistic regression (LR), and orange line for random forest (RF). AUC: the area under the curve

LR (Fig. 2I-K). Notably, similar results were obtained for all three algorithms, thus prompting us to present RF analysis results as representative results in this report. RF analysis yielded the following respective sensitivity, specificity, accuracy, and AUC values for the three pairwise comparisons: PTB/HC (93.33%, 86.67%, 90%, 0.9840), PTB/CAP (93.33%, 93.33%, 93.33%, 0.9680), and PTB/LCA (100%, 86.67%, 93.33%, 0.9310) (Table 2).

Pairwise comparisons of urinary DAMs

To unveil PTB-associated urine metabolite profiles, we conducted metabolomic analysis of all 160 urine samples using a UPLC-ESI-Q-Orbitrap-MS system. The results of this analysis revealed 47,528 features, subsequent analysis was based on annotated 1272 features. Prior to data analysis, data integrity was assessed and detected no missing values. PCA was performed to evaluate the results for overall variability across the four experimental groups. PTB and HC results were almost completely separate (Fig. 3A), while PTB/CAP and PTB/LCA comparisons exhibited overlapping results (Fig. 3B-C). The data of urine metabolites in the four groups (PTB, LCA, CAP, and HC) were analyzed using the OPLS-DA method. According to the characteristics of the model, the data of any two groups could be well distinguished (Fig. 3D-F). Remarkably, verification model data indicated that the OPLS-DA model was not over-fitting and was reliable to screen metabolic biomarkers (Supplementary Figure S3). After excluding exogenous metabolites that were identified by comparing our DAMs to entries within the Lumingbio untargeted LC-MS database, we identified 76, 38, and 54 significantly different endogenous DAMs (Supplementary Table S7-S9) from PTB/HC, PTB/CAP, and PTB/LCA pairwise comparisons, respectively.

Subsequently, we employed the EFS approach and Student's *t*-test to rank endogenous DAMs (Fig. 3G-I). Guided by the results obtained from the aforementioned pairwise comparisons, we identified the five endogenous DAMs for PTB/HC (oleamide, C17-sphingosine, FA 13:3+1O, 3-nitro-L-tyrosine, 5-hydroxy-L-tryptophan), PTB/CAP (dihydrouracil, leucyl-leucine, 16-hydroxypalmitate, uridine, nicotinamide), and PTB/LCA (leucyl-leucine, hexylamine, uridine, choline, 1-methylnicotinamide). This selection was based on metabolite biomarker AUCs of ≥0.71 and *p* values of <0.01 (Supplementary Table S10, Supplementary Figure S4), alongside additional criteria including: (1) high EFS value; (2) high FC value; and (3) established associations with PTB metabolic processes. Thereafter, we conducted ROC analysis to determine AUC values for the combined five-metabolite biomarker set identified from the PTB/HC, PTB/CAP, and PTB/LCA pairwise comparisons (Fig. 3J-L). RF machine learning analysis yielded respective sensitivity, specificity, accuracy, and AUC values for the combined five-metabolite set for PTB/HC of 97.44%, 97.44%, 97.44%, 0.9940, for PTB/CAP of 77.50%, 89.74%, 83.54%, 0.8920, and for PTB/LCA of 82.05%, 87.50%, 84.81%, 0.8570 (Table 2).

Discussion

The accurate and rapid diagnosis of TB, especially at the point-of-care, is crucial for curbing the spread of the disease, treatment monitoring, risk analysis and prognosis [9]. A pressing challenge in TB diagnosis lies in the development of more streamlined, accurate, and rapid point-of-care tests [17]. Although efforts to develop novel tools for infectious disease diagnosis have been focused on discovery of single disease-specific biomarkers, this focus will likely shift towards discovery of bio-profiles or biosignatures comprised of well-defined sets of reliable molecular indicators [18]. In the present study, we identified five urine proteins with strong potential applicability as single PTB-diagnostic biomarkers (CPB2, HTRA1, HEXA, FLNA, NRP1) that all yielded AUCs of >0.80. Notably, a five-marker biosignature incorporating the same proteins effectively distinguished PTB from CAP, LCA, and HC, as based on AUCs of 0.9680, 0.9310, and 0.9840, respectively. However, the use of individual proteins as diagnostic tools for infectious diseases such as PTB is not generally recommended, since single inflammatory biomarkers tend to have low disease specificities [19]. As expected, our study demonstrated superior PTB-diagnostic accuracies of biomarker combinations as compared to individual markers, as reflected by EFS ranking and Student's *t*-test results, as observed in previous studies [20–22].

Meanwhile, our results also indicate that proteomics is valuable tool for achieving effective PTB diagnosis and

Table 2 Relevant parameters of RF machine learning method

	AUC	Cut off	Sensitivity	Specificity	Accuracy	Likelihood Ratio
PTB vs. CAP ^{M}	0.8920	0.6390	0.7750	0.8974	0.8354	7.5563
PTB vs. CAP ^{P}	0.9680	0.5001	0.9333	0.9333	0.9333	14.0000
PTB vs. HC ^{M}	0.9940	0.7026	0.9744	0.9744	0.9744	38.0000
PTB vs. HC ^{P}	0.9840	0.4467	0.9333	0.8667	0.9000	7.0000
PTB vs. LCA ^{M}	0.8570	0.4381	0.8205	0.8750	0.8481	6.5641
PTB vs. LCA ^{P}	0.9310	0.1678	1.0000	0.8667	0.9333	7.5000

*M: Metabolome biomarkers; *P: Proteome biomarkers

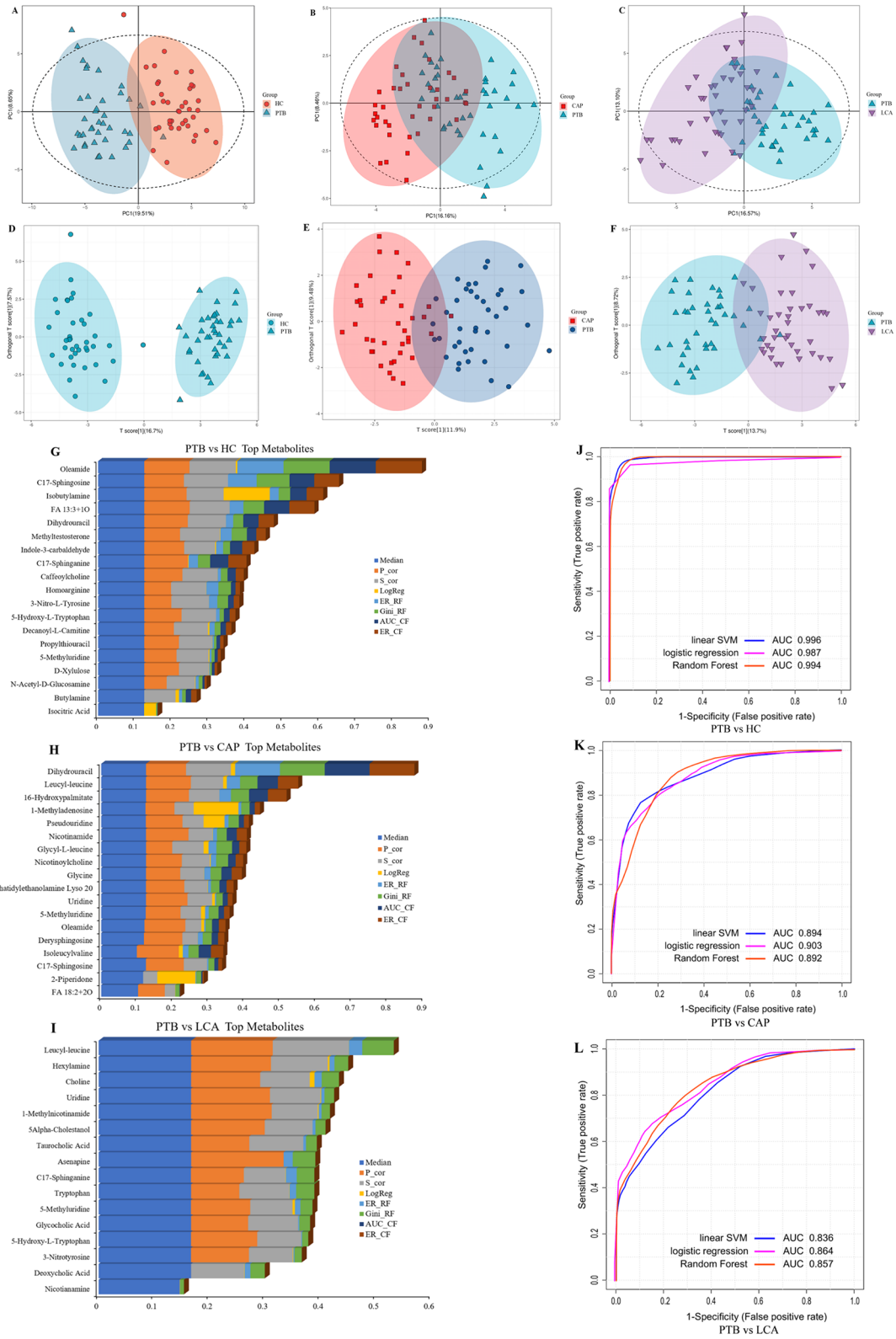

Fig. 3 Ensemble feature selection analysis of metabolomics data and ROC analysis of potential biomarkers among PTB, CAP, LCA and HC groups (**A**-**C**) PCA score plots for Group PTB/HC, PTB/CAP and PTB/LCA. (**D**-**F**) OPLS-DA score plots for Group PTB/HC, PTB/CAP and PTB/LCA. The R2(X), R2(Y)and Q2 values for the three OPLS-DA models were: R2(X) = 0.306, R2(Y) = 0.944, Q2 = 0.885 for D; R2(X) = 0.213, R2(Y) = 0.753, Q2 = 0.645 for E and R2(X) = 0.224, R2(Y) = 0.754, Q2 = 0.637 for F. PCA: principal components analysis; OPLS-DA: orthogonal partial least squares discriminant analysis. (**G**-**I**) EFS analysis for Group PTB/HC, PTB/CAP and PTB/LCA. (**J**-**L**) ROC curves for each pairwise prediction by three different machine learning methods

differential diagnosis, as consistent with results obtained by Liu et al. who analyzed and compared urinary proteomic profiles of TB patients and healthy controls. They found that the combined set of P22352, Q9P121, P15151, Q13291, and Q8NDA2 biomarkers could be useful for diagnosing TB, as based on a TB-diagnostic sensitivity rate of 82.70% and latent TB infection (LTBI)-diagnostic specificity rate of 92.30% [20]. Recently, P01011, Q8NCW5, P28072, A0A2Q2TTZ9, and Q99574 were identified via MS conducted using a Q-Exactive Orbitrap mass spectrometer then these biomarkers were combined to generate a five-protein biosignature that, after leave-one-out cross-validation, yielded an AUC of 1.00 (95% CI, 1.00–1.00) and exhibited a TB-diagnostic sensitivity rate of 100% (95% CI, 76.20–100%) and specificity rate of 90.9% (95% CI, 58.70–99.80%) [21]. One important issue is that biomarkers from different studies rarely overlap even when the groups to be compared are the same. The reasons can be the small sample size, statistical fluctuations in individual systemic biological state, race factor, experimental conditions, and analytical apparatuses differ from one study to another [9, 22]. However, one must consider the possibility that multiple independent systemic regulated states exist, each manifesting the same typical TB symptoms, yet none of the biomarkers are consistently up- or down-regulated across these different states. This is akin to a non-linear complex equation often having multiple solutions [9].

Metabolomics, like proteomics, has been increasingly used for PTB diagnosis in recent years. For example, in 2014–2015, Mrinal et al. [23] and Mahapatra et al. [24] reported that LC-MS and gas chromatography-mass spectrometry methods, respectively, could be used to identify metabolites in urine samples of TB patients. Since then, a multitude of metabolomics-based studies have been reported describing the successful use of small molecule metabolites as urine PTB-diagnostic biomarkers [10]. However, these studies only explored a subset of all potential PTB-predictive DAMs in urine of PTB patients, as LC-MS detection of all urinary compounds is not yet feasible. To broaden our repertoire of identified metabolites, we employed ultra-high-performance liquid chromatography (UHPLC)-MS/MS techniques to screen for DAMs present in urine specimens of PTB, LCA, CAP, and HC groups. Our findings pinpointed five potentially useful PTB-diagnostic and differentially diagnostic DAMs for each group, as based on AUCs of ≥ 0.71 and p values of <0.01. Interestingly, a biosignature incorporating the five metabolite biomarkers was capable of effectively differentiating PTB from CAP, LCA, and HC, as based on respective AUCs of 0.8920, 0.8570, and 0.9940. These DAMs are mainly by-products of amino acid, nucleotide, and lipid metabolism and are often generated during oxidative stress and inflammatory responses [10, 25–27].

In a similar vein, metabolomics results obtained by Jiang et al. [25] and Cho et al. [27] demonstrated that changes in blood serum levels of multiple amino acids in active TB patients may serve as useful biomarkers for achieving adjunctive, rapid, and noninvasive PTB diagnosis. Oleamide has been shown to be a useful biomarker for distinguishing between tuberculous pleural effusion and malignant pleural effusion [28]. Furthermore, a biosignature incorporating these three metabolites effectively discriminated between TB patients and HCs (AUC = 0.97). Another noteworthy study highlighted the potential of free 3-NT levels in biological samples to differentiate drug-sensitive TB from drug-resistant TB [26]. More recently, a nuclear magnetic resonance-based urinary metabolomics study identified eight metabolites; a urine metabolic fingerprint based on these metabolites could be used to effectively discriminate active TB patients from pneumococcal pneumonia patients, those with LTBI, and HCs [10]. This urinary metabolic fingerprint may also be capable of distinguishing PTB patients from HCs, LCA patients, and CAP patients and thus may be a useful noninvasive biosignature for achieving effective PTB diagnosis and differential diagnosis, warranting further research.

Importantly, different omics methodologies detect different subsets of diagnostic biomarkers present in complex clinical specimens. In fact, the power of Multi-omics approach has already been shown by in vivo and in vitro modelling studies that reconciled multilayered omics data acquisitions and metabolic and other phenotypes over a large number of experiments (performed in different conditions) and were able to accurately predict biological behavior [29]. Here, we integrated proteomics and metabolomics to discover urine PTB-diagnostic biomarkers (five proteins and five metabolites) using pairwise comparisons of biomarkers of PTB, LCA, CAP, and HC groups. As an additional contribution, we present a multi-omics-derived PTB-diagnostic biomarker signature, a rarity in the literature. In addition, the quantification of statistically significant biomarkers identified in Multi-omics would allow this technology to be adapted to a point-of-care test, especially in resource-constrained settings. This study had several limitations. First, the number of urine specimens subjected to proteomics and metabolomics analyses was relatively small and thus prevented us from ruling out potential bias due to sample heterogeneity, an issue that can be addressed through validation of biomarkers using large sample sets. Second, although we included samples based on our criteria for our comparative analyses, additional undetected, uncontrolled genetic, clinical, or environmental confounding factors may have influenced our results. Third, the

clinical efficacy of biomarkers identified in our cohort was also not assessed in a larger population; thus, further validation is urgently required. Finally, due to the lack of our ability to interpret the complex nature and biological mechanisms of metabolism between diverse analytes, we focused on the analysis of diagnostic performance to the five protein and five metabolism biomarkers, we made the raw data publicly available.

Conclusions

This study employed metabolomic and proteomic analysis to identify urine biomarkers capable of distinguishing PTB and LCA, CAP, and HC subjects with high sensitivity and specificity. Although it is too early to conclude that these markers will replace the invasive approach for diagnosing TB in the clinic, our findings indeed shed light on the development of valuable diagnostic tools for effective clinical implementation and combat TB to achieve earlier prevention, earlier diagnosis, and earlier treatment of TB. We believe that these results lay a foundation to support development of novel methods for diagnosing and identifying PTB patients, meanwhile also enhancing our understanding of underlying PTB disease mechanisms.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12014-024-09514-4.

Supplementary Material 1

Acknowledgements
The authors thank Shanghai Bioprofile Technology Co. Ltd. for technical support in proteomics and metabolomics.

Author contributions
TSJ, PY contributed to the idea and design. TSJ, PY contributed to methodology. YJJ, LZD, YH, LMG, MLP, LRM, DWM, LL, MTY participated in urine specimen collection and participant information collection. TSJ, PY, YJJ, YJF participated in data collection and data management. TSJ, PY, YJJ participated in data analysis and wrote the initial draft of the report and all authors revised and approved the final report. TSJ and PY are corresponding authors.

Funding
This work was supported by National Key Research and Development Program of China (2022YFC2304803) and Beijing Nova Program (20220484169&20230484295); Program of Tongzhou District Science & Technology Commission, Grant No. JCQN2023015 and by National Key R&D Program of China (2022YFC2304802).

Data availability
All data generated or analyzed during this study are included in this published article and its supplementary information files. The mass spectrometry proteomics raw data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD048717. The metabolomics raw data have been deposited to the metabolight with dataset identifier MTBLS9330.

Declarations

Ethics approval and consent to participate
This study was conducted in compliance with principles outlined in the Declaration of Helsinki and was approved by the Ethics Committee of Beijing Chest Hospital (YJS-2022-04). All study protocols complied with existing Ethics Committee guidelines.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 12 September 2023 / Accepted: 14 November 2024
Published online: 18 December 2024

References

1. World Health Organization. Global tuberculosis Report 2023. Geneva: World Health Organization; 2023.
2. Yong YK, Tan HY, Saeidi A, et al. Immune biomarkers for diagnosis and treatment monitoring of tuberculosis: current developments and future prospects. Front Microbiol. 2019;10:2789. https://doi.org/10.3389/fmicb.2019.02789.
3. Marlowe EM, Novak-Weekley SM, Cumpio J, et al. Evaluation of the Cepheid Xpert MTB/RIF assay for direct detection of Mycobacterium tuberculosis complex in respiratory specimens. J Clin Microbiol. 2011;49(4):1621–3. https://doi.org/10.1128/JCM.02214-10.
4. Friedrich SO, Rachow A, Saathoff E, et al. Assessment of the sensitivity and specificity of Xpert MTB/RIF assay as an early sputum biomarker of response to tuberculosis treatment. Lancet Respir Med. 2013;1(6):462–70. https://doi.org/10.1016/S2213-2600(13)70119-X.
5. Ellis DI, Dunn WB, Griffin JL, et al. Metabolic fingerprinting as a diagnostic tool. Pharmacogenomics. 2007;8(9):1243–66. https://doi.org/10.2217/14622416.8.9.1243.
6. Aslam B, Basit M, Nisar MA, et al. Proteomics: technologies and their applications. J Chromatogr Sci. 2017;55(2):182–96. https://doi.org/10.1093/chromsci/bmw167.
7. Gao Y. Urine-an untapped goldmine for biomarker discovery? Sci China Life Sci. 2013;56(12):1145–6. https://doi.org/10.1007/s11427-013-4574-1.
8. Liebenberg C, Luies L, Williams AA. Metabolomics as a Tool to investigate HIV/TB Co-infection. Front Mol Biosci. 2021;8:692823. https://doi.org/10.3389/fmolb.2021.692823.
9. Yu Y, Jiang XX, Li JC. Biomarker discovery for tuberculosis using metabolomics. Front Mol Biosci. 2023;10:1099654. https://doi.org/10.3389/fmolb.2023.1099654.
10. Izquierdo-Garcia JL, Comella-Del-Barrio P, Campos-Olivas R, et al. Discovery and validation of an NMR-based metabolomic profile in urine as TB biomarker. Sci Rep. 2020;10(1):22317. https://doi.org/10.1038/s41598-020-78999-4.
11. Guo J, Zhang X, Chen X, et al. Proteomics in Biomarker Discovery for Tuberculosis: current status and future perspectives. Front Microbiol. 2022;13:845229. https://doi.org/10.3389/fmicb.2022.845229.
12. Mandell LA, Wunderink RG, Anzueto A, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis. 2007;44(Suppl 2):S27–72. https://doi.org/10.1086/511159.
13. Wiśniewski JR, Zougman A, Nagaraj N, et al. Universal sample preparation method for proteome analysis. Nat Methods. 2009;6(5):359–62. https://doi.org/10.1038/nmeth.1322.
14. Neumann U, Genze N, Heider D. EFS: an ensemble feature selection tool implemented as R-package and web-application. BioData Min. 2017;10:21. https://doi.org/10.1186/s13040-017-0142-8.
15. Wozniak JM, Mills RH, Olson J, et al. Mortality risk profiling of Staphylococcus aureus Bacteremia by multi-omic serum analysis reveals early predictive and pathogenic signatures. Cell. 2020;182(5):1311–e132714. https://doi.org/10.1016/j.cell.2020.07.040.
16. Choi M, Chang CY, Clough T, et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics. 2014;30(17):2524–6. https://doi.org/10.1093/bioinformatics/btu305.

17. Steingart KR, Schiller I, Horne DJ, et al. Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. Cochrane Database Syst Rev. 2014;2014(1):CD009593. https://doi.org/10.1002/14651858.
18. Mamas M, Dunn WB, Neyses L, et al. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Arch Toxicol. 2011;85(1):5–17. https://doi.org/10.1007/s00204-010-0609-6.
19. Olivier M, Asmis R, Hawkins GA, et al. The need for Multi-omics Biomarker signatures in Precision Medicine. Int J Mol Sci. 2019;20(19):4781. https://doi.org/10.3390/ijms20194781.
20. Liu L, Deng J, Yang Q, et al. Urinary proteomic analysis to identify a potential protein biomarker panel for the diagnosis of tuberculosis. IUBMB Life. 2021;73(8):1073–83. https://doi.org/10.1002/iub.2509.
21. Mutavhatsindi H, Calder B, McAnda S, et al. Identification of novel salivary candidate protein biomarkers for tuberculosis diagnosis: a preliminary biomarker discovery study. Tuberculosis (Edinb). 2021;130:102118. https://doi.org/10.1016/j.tube.2021.102118.
22. Shi H, Yuan J, Zhang Y, et al. Discovering significantly different metabolites between Han and Uygur two racial groups using urinary metabolomics in Xinjiang, China. J Pharm Biomed Anal. 2019;164:481–8. https://doi.org/10.1016/j.jpba.2018.11.016.
23. Das MK, Bishwal SC, Das A, et al. Deregulated tyrosine-phenylalanine metabolism in pulmonary tuberculosis patients. J Proteome Res. 2015;14(4):1947–56. https://doi.org/10.1021/acs.jproteome.5b00016.
24. Mahapatra S, Hess AM, Johnson JL, et al. A metabolic biosignature of early response to anti-tuberculosis treatment. BMC Infect Dis. 2014;14:53. https://doi.org/10.1186/1471-2334-14-53.
25. Jiang J, Li Z, Chen C, et al. Metabolomics Strategy Assisted by Transcriptomics Analysis To Identify Potential Biomarkers Associated with Tuberculosis. Infect Drug Resist. 2021;14:4795–807. https://doi.org/10.2147/IDR.S330493.
26. Bolajoko EB, Arinola OG, Odaibo GN, et al. Plasma levels of tumor necrosis factor-alpha, interferon-gamma, inducible nitric oxide synthase, and 3-nitrotyrosine in drug-resistant and drug-sensitive pulmonary tuberculosis patients, Ibadan, Nigeria. Int J Mycobacteriol. 2020;9(2):185–9. https://doi.org/10.4103/ijmy.ijmy_63_20.
27. Cho Y, Park Y, Sim B, et al. Identification of serum biomarkers for active pulmonary tuberculosis using a targeted metabolomics approach. Sci Rep. 2020;10(1):3825. https://doi.org/10.1038/s41598-020-60669-0.
28. Liu Y, Mei B, Chen D, et al. GC-MS metabolomics identifies novel biomarkers to distinguish tuberculosis pleural effusion from malignant pleural effusion. J Clin Lab Anal. 2021;35(4):e23706. https://doi.org/10.1002/jcla.23706.
29. Rinschen MM, Ivanisevic J, Giera M, et al. Identification of bioactive metabolites using activity metabolomics. Nat Rev Mol Cell Biol. 2019;20(6):353–67. https://doi.org/10.1038/s41580-019-0108-4.

Publisher's note