Journal of Chromatography Open 7 (2025) 100204

Contents lists available at ScienceDirect

Journal of Chromatography Open

journal homepage: www.elsevier.com/locate/jcoa

Identifying skin surface chemicals as potential tuberculosis diagnostic biomarkers using ultra performance liquid chromatography-high resolution mass spectrometry

Madelien Wooding *, Kornelis van Pletzen, Yvette Naudé

Department of Chemistry, University of Pretoria, Private Bag X20, Hatfield 0028, Pretoria, South Africa

ARTICLE INFO

Keywords:
Wearable PDMS sampler
Non-invasive sampling
Human surface skin compounds
Tuberculosis biomarkers, Ion mobility
UPLC-MS

ABSTRACT

Tuberculosis (TB) remains a significant cause of morbidity and mortality globally, despite its preventability and curability. Early and accurate diagnosis of active TB is essential for enhancing patient care, improving outcomes, and interrupting the transmission cycles of *Mycobacterium tuberculosis* (*M.tb*). Metabolomics proves to be an emerging area of study for the development of a non-invasive approach to TB diagnostics.. High-resolution mass spectrometry combined with ion mobility spectrometry enhances the confidence in identifying and annotating biological markers during metabolomic research. This study outlines an analytical workflow encompassing sample preparation through to multivariate analyses for detecting potential TB diagnostic biomarkers. A custom-designed wearable polydimethylsiloxane (PDMS) sampler was employed as a passive sampling device, effectively concentrating chemical compounds from the skin surface. The sampler was directly desorbed into solvent within an LC vial, streamlining the extraction-to-analysis process. Utilising accurate mass and collision cross sections (CCS), fourteen biomarkers were tentatively identified, demonstrating the ability to differentiate TB patients from control groups. Receiver operating characteristic (ROC) analysis yielded an area under the curve (AUC) of 0.911. Among these, para-aminobenzoic acid (PABA) emerged as a promising biomarker for TB, achieving a specificity of 1, sensitivity of 0.9, and an AUC of 0.961. Method limits of detection for the 1-hour non-invasive skin sampling method ranged from 6 (PABA) to 172 ng (phenylalanine) for a calibration working range of 10 – 800 ng with a R^2 of $\geq$ 0.99. These first results demonstrate the potential of using skin surface compounds in TB diagnostics.

1. Introduction

In 2022, the World Health Organisation (WHO) classified tuberculosis (TB) as the world's second deadliest disease caused by a single infectious agent, following the coronavirus (COVID-19) [1]. TB is a contagious disease transmitted through the inhalation of *Mycobacterium tuberculosis* (*M.tb*) bacteria, which are expelled into the air by infected individuals, for instance, through coughing. [2,3]. While the disease primarily affects the lungs, known as pulmonary TB, it can also impact other body sites [1,3,4]. An estimated quarter of the global population harbours latent *M.tb*, which can progress to active TB if the host immune response is compromised. If left untreated, the mortality rate for TB infection is approximately 50 % [1,3]. In 2022, the WHO reported, 7.5 million new TB cases and 1.3 million deaths [1]. This statistic is particularly alarming because TB is a treatable disease, with appropriate treatment plans capable of curing approximately 85 % of infected individuals [1,3]. Therefore, early diagnosis of TB is crucial to reduce the disease burden.

Current diagnostic tests for TB often lack sensitivity and specificity, are expensive, invasive, time-consuming, and require medical professionals. Examples include GeneXpert units and gastric aspirates needed for individuals who cannot produce sufficient sputum, as well as smear microscopy [3,5–7]. Given the challenges and limitations of current diagnostic tools new approaches in TB diagnosis are needed to meet the WHO End TB Strategy: 2025 milestones of decreasing TB incidence rate by 50 % and TB deaths by 75 % compared to 2015 levels [1]. One such approach is the use of biomarkers. Advances in analytical techniques such as two-dimensional gas chromatography (GC × GC) and high-resolution mass spectrometry (HRMS), along with sophisticated statistical tools, have led to an increase in biomarker discovery for

* Corresponding author.
E-mail address: madelien.wooding@up.ac.za (M. Wooding).

https://doi.org/10.1016/j.jcoa.2025.100204
Received 17 December 2024; Received in revised form 28 January 2025; Accepted 1 February 2025
Available online 2 February 2025

diagnostic purposes. Metabolomics plays a significant role in this new diagnostic approach, offering non-invasive sample analysis through breath, urine, sputum, serum, and skin [5,8–12]. This method provides a comprehensive metabolomic profile of the sample and can be implemented in a targeted strategy [6] or for discovery work [7,13]. Several metabolomic studies have analysed breath, sputum, blood, and urine to identify new volatile organic compounds (VOCs) as potential TB biomarkers, primarily using gas chromatography with mass spectrometry (GCMS–) [8–11]. However, there is a notable lack of studies on skin analysis for TB diagnostics [5,7]. This approach is noteworthy as it offers a non-invasive sampling option from a non-infectious area compared to breath. Turner et al. demonstrated the feasibility of using skin as a matrix in a comparative study of volatile organic compounds (VOCs) from breath and skin following glucose ingestion. The study found that most VOCs detected in breath were also present on the skin, even though at lower concentrations, establishing a correlation between VOCs emitted from the skin and those exhaled in breath [12]. Furthermore, there is a significant gap in studies investigating non-volatile and semi-volatile compounds using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) for TB metabolomic studies. However, research by Magdalena et al. and Cho et al. has demonstrated the feasibility of using LC-MS/MS to analyse serum samples for TB biomarkers [6,13]. The benefit of adding ion mobility spectrometry (IMS) to HRMS for biomarker discovery was demonstrated by Wooding et al. in a skin metabolomic study investigating mosquito attractiveness to humans [14,15]. IMS enhances spectral quality in data-independent acquisition (DIA), improves product-precursor alignment through drift time matching, and provides an additional dimension for compound identification by utilising collision cross section (CCS) values [16]. This makes IMS highly valuable in biomarker discovery.

In this study, we utilised a passive, non-invasive wearable PDMS sampler with micro-solvent desorption, combined with ultra-performance liquid chromatography coupled with ion mobility spectrometry and high-resolution mass spectrometry (UPLC-IMS-HRMS), to profile metabolites present on the skin surface of active TB patients and healthy controls. The feasibility of this sampler for human skin surface sampling has been previously demonstrated by our group in Rood et al., Wooding et al., and Makhubela et al. [5,14,15,17]. Multivariate analysis, supported by appropriate software tools, facilitated feature alignment and TB biomarker identification through chromatographic retention times, accurate mass, and ion mobility drift times. A targeted approach was followed to validate the skin surface sampling method by using TB biomarkers as reported in literature from blood cultures and serum.

2. Materials and methods

2.1. Reagents and chemicals

Acetone, methanol (MeOH), and deionised water used for sampler conditioning were obtained from Merck, South Africa. Ultra-pure MeOH, acetonitrile (ACN), and water were sourced from Romil (Romil-UpS™, Waterbeach, Cambridge, United Kingdom), while formic acid (99+ % purity) was procured from Thermo Scientific, South Africa. *p*-Aminobenzoic acid (PABA), analytical standard (Supelco), was purchases from Merck, South Africa.

2.2. Standard solutions

A 0.5 μmole/ml amino acid standard solution prepared in 0.2 N lithium citrate (pH 2), containing 2 % thiodiglycol and 0.1 % phenol, was sourced from Sigma-Aldrich (Pty) Ltd., Kempton Park, South Africa. l-phenylalanine, l-tyrosine (Holistix), and l-lysine (Dis-Chem Gold) were purchased from Dis-Chem, Pretoria, South Africa.

A 100 ng/μl PABA stock solution was prepared by dissolving 5 mg of PABA in 50 ml of methanol. Individual stock solutions were prepared as follows: 1000 ng/μl for l-phenylalanine (250 mg in 250 ml) and l-lysine (500 mg in 500 ml) in water, and 900 ng/μl for l-tyrosine (450 mg in 500 ml) in water. A final working standard solution, containing a mixture of the target analytes at 10 ng/μl, was prepared in 100 ml of ACN:water (1:1, v/v) by combining 10 ml of the 1000 ng/μl l-phenylalanine and l-lysine stock solutions, 1.11 ml of the 900 ng/μl l-tyrosine stock solution, and 10 ml of the 100 ng/μl PABA stock solution. All standard and working solutions were stored in glass Schott bottles at 4 °C. Study Cohort

The proof of concept study involved two groups: a TB patient and a control group, each comprising both male and female participants ($n = 28$). Details of the study cohort can be found in Makhubela et al. [5]. Each participant was sampled in triplicate, two of the samples were analysed with GC × GC-TOFMS during a study by Makhubela et al. [5], the remaining sample was analysed for this study. Ten patients were clinically diagnosed with TB. The control group ($n = 18$) was confirmed to be TB-negative. TB-positive status was determined using the GeneXpert MTB/RIF assay (Cepheid, Sunnyvale, CA, USA), while TB-negative status was validated through the QuantiFERON-TB Gold assay (Cellestis/Qiagen, Carnegie, Australia) and the Fourth Generation HIV-1/2 ELISA assay (Abbott Diagnostics Medical Co. Ltd., Lake Forest, IL, USA) to minimise false-positive results. Samples from these participants were collected in 2018, prior to initiating TB treatment. The TB-positive cohort was sampled at the Steve Biko Academic Hospital and Tshwane District Hospital, both located in Pretoria, South Africa. The TB-negative cohort included students and staff from the University of Pretoria (UP), with sampling conducted on UP premises in 2018. All samplers were stored at −18 °C. Analyses were done in 2023..

2.3. Skin surface sampling

Details of sampler construction, solvent modification, sampling of participants and sampler storage can be found in Makhubela et al. [5]. In short, wearable PDMS samplers were constructed in-house, followed by modification with ethanol, sampling bands were applied to the inner arm skin surface of each participant (Fig. 1A), covered with a Mylar® patch, and secured using adhesive surgical paper tape (Fig. 1B). Before sampling, the inner arm area was cleansed using 70 % v/v isopropanol pads (Medic + dressings, Dischem, Hillcrest, South Africa).. After a sampling duration of one hour the bands were removed and stored in an ethanol saturated environment at −18 °C [5,18]. A one-hour sampling period balanced aspects such as ethical considerations, number of chromatographic peaks collected and resolution of these peaks. In terms of ethical considerations, a longer sampling duration would delay urgent commencement of treatment of the TB-patients. The application of ethanol as an extraction aid stems from its ability to induce approximately 2 % swelling in silicone, combined with its capacity to enhance the diffusion and partitioning characteristics of the target compounds during extraction [19].

Before analysis, the sampling loops were retrieved from the storage vials using clean stainless-steel tweezers and individually placed into 200 μl glass conical tip inserts (Fig. 1C) (Macherey-Nagel GmbH & Co, Separations, South Africa). These inserts were then positioned inside 1.5 ml glass screw-neck LC vials (Macherey-Nagel GmbH & Co, Separations, South Africa) containing 1 ml of deionised water to facilitate energy transfer during extraction [20]. Solvent back extraction (BE) was performed by adding 200 μl of ultra-purity MeOH:ACN (1:1, v/v) to each glass insert containing the individual samplers, following the procedure as outlined by Margoum et al. [21].. The LC vials were then sealed with PTFE pre-slit screw caps (Macherey-Nagel GmbH & Co, Separations, South Africa) and sonicated for 15 min at room temperature. After sonication, the samplers were removed using clean stainless steel tweezers, and the LC vials were resealed and transferred to the sample manager of a UPLC system, maintained at 8 °C, for UPLC-IMS-HRMS analysis.

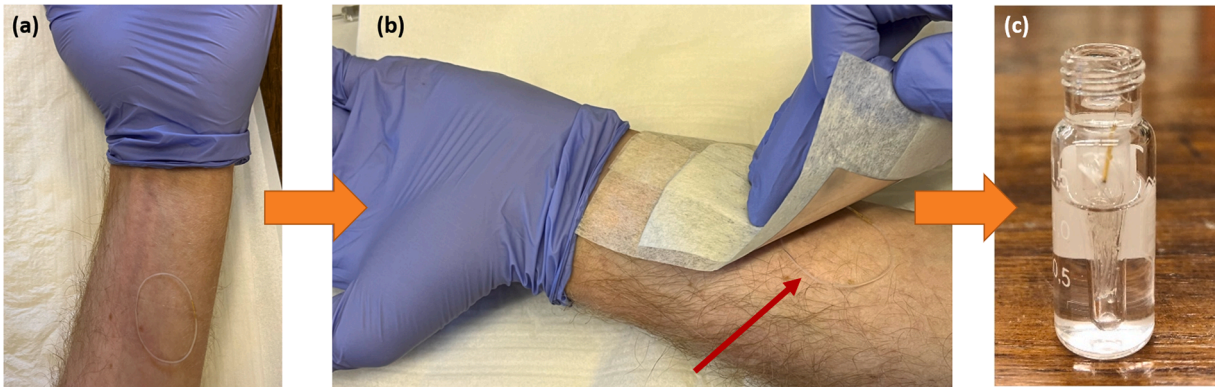

Fig. 1. Skin surface sampling procedure. (A) The ethanol modified PDMS sampler on the inner arm. (B) the Mylar® sheeting with medical adhesive tape securing the sampler. (C) Micro-solvent back extraction of the PDMS sampler in an LC vial insert.

2.4. Ethical considerations

Ethical approval for this study was granted by the ethics committees of the Faculty of Health Sciences and the Faculty of Natural and Agricultural Sciences at the University of Pretoria (reference numbers: 300/2018 and EC180306–176). Authorisation to conduct the study within its hospitals was obtained from Steve Biko Academic Hospital and Tshwane District Hospital in Pretoria, South Africa. Written informed consent was secured from all participants prior to their involvement in the study.

2.5. Quality control protocols

Extracts were stored either in the LC autosampler at 8 °C or in a refrigerator at 4 °C until analysis. To account for potential laboratory background compounds, three laboratory method blanks were analysed. Additionally, an instrument solvent blank was run prior to the UPLC analysis of the samples. Two pooled samples were prepared by combining 10 µl from each extract, one for the TB-positive group and one for the control group, and analysed as single samples. All individual samples were analysed in a randomised order, with pooled samples and method blanks interspersed throughout the sequence. Method Validation

A method-matched calibration was used for quantification of target analytes detected on the skin surface during the one-hour sampling period. A conditioned PDMS sampler, modified with solvent, was positioned on a pre-cleaned Mylar® reflective sheet (10 cm x 6 cm), which has been sanitised using a medical grade alcohol cleansing pad. The analyte mixture (PABA, tyrosine, phenylalanine, lysine) working standard solution (10 ng/µl each) solution was spiked onto the Mylar® sheet, next to but not touching the sampler, with the following increasing volumes: 1 µL (10 ng), 2 µL (20 ng), 5 µL (50 ng), 10 µL (100 ng), 20 µL (200 ng), 40 µL (400 ng), and 80 µL (800 ng). Spiking was done in triplicate at 100 ng to determine precision (% relative standard deviation (%RSD)) and accuracy (% recovery). The Mylar® sheet was folded into a parcel and sealed using 3 M dressing tape. This parcel was then placed in a 100 ml Schott bottle and suspended in a water bath maintained at 31 °C to simulate human skin temperature [22]. After 1 hour, the sampler was removed from the Mylar® package and subjected to solvent back extracted as described in Section 2.3. The limits of detection (LODs) and quantification (LOQs) were determined as those amounts producing a signal to noise ratios (S/N) of 3 and 10, respectively. Matrix effects were evaluated by spiking a PDMS sampler ($n = 3$) worn by a TB negative volunteer with 100 ng of the standard mixture, after the sampler was removed from the skin surface, and compared to conditioned PDMS samplers ($n = 3$) spiked with 100 ng of the standard mixture pre-extraction. The sampling procedures outlined above were followed.

2.6. UPLC-IMS-HRMS

Compound separation and detection were performed using a Waters Acquity Ultra Performance Liquid Chromatography (UPLC®) system coupled to a Waters® Synapt G2 High-Definition Mass Spectrometry (HDMS) system (Waters Inc., Milford, Massachusetts, USA). Data acquisition was carried out using MassLynx™ software (version 4.1, Waters Inc., Milford, Massachusetts, USA). The source parameters were configured as follows: the capillary voltage was set to 2.8 kV for positive ionisation mode and 2.4 kV for negative ionisation mode. The source temperature was maintained at 120 °C, with a sampling cone voltage of 25 V and an extraction cone voltage of 4.0 V. Nitrogen was used as the cone gas at a flow rate of 10.0 L/hr. The desolvation temperature was set at 350 °C, with nitrogen desolvation gas flowing at 600.0 L/hr. Mass spectral scans were recorded every 0.3 ss in continuous profile mode, with a mass range of 50–1200 *m/z*. Data-independent acquisition (DIA) was carried out using two alternating acquisition functions with low and high collision energy, employing an ion mobility-enabled HDMSE approach. Tandem MS (HDMSE) fragmentation utilised high-energy collision-induced dissociation (CID) with argon gas. The transfer collision energy was set to 4 V for the low-energy function, while a ramp from 15 to 45 V was applied for the high-energy function.

For ion mobility spectrometry (IMS), nitrogen was used as the drift gas at a flow rate of 90 ml/min, and helium flowed at 180 ml/min in the helium cell. The mobility t-Wave operated at a variable IMS wave velocity, set at 650 m/s with a ramp ranging from 1000 to 300 m/s. The IMS wave height was configured at 40.0 V, with a transfer wave velocity of 224 m/s (optimised to prevent pusher phasing) and a transfer wave height of 4.0 V. The trap DC bias and helium cell DC voltages were set at 45.0 V and 35.0 V, respectively. An IMS wave delay of 1000 µs was applied for mobility separation. Calibration of the IMS was conducted using the Waters Major Mix Calibration Sample and Driftscope (version 2.8) to determine experimental collision cross-section (CCS) values (Ω), yielding a CCS error of <3.7 % (0.66 ± 0.8 %).

2.7. Chromatographic conditions

Separation was achieved using a reverse-phase step gradient with mobile phase A consisting of water containing 0.1 % formic acid and mobile phase B consisting of acetonitrile with 0.1 % formic acid. The gradient began with an isocratic hold at 3 % B for 0.1 mins, followed by a linear increase to 100 % B over 14.0 mins. The column was then washed for 2 mins, and the mobile phase transitioned back to the initial conditions over 0.5 mins, followed by reconditioning to restore the starting conditions.

The column temperature was maintained at 40 °C throughout the run, with a flow rate set at 0.4 ml/min, resulting in a total run time of 20 mins. An injection volume of 5 µl was used, and the autosampler was

maintained at 8 °C. Positive and negative ESI ionisation mode mass spectra were recorded in separate chromatographic runs using identical separation conditions and columns. All separations were performed on a Waters Acquity UPLC® HSS T3 column (1.8 µm particle size, 2.1 mm ID × 150 mm length) supplied by Microsep, South Africa.Data Processing and Statistical Analysis

Data processing was conducted using Masslynx V4.1 or the UNIFI® Scientific Information System (Waters Inc., Milford, Massachusetts, USA). Method performance was assessed by evaluating method-matched calibration linearity, limits of detection (LOD), limits of quantification (LOQ), accuracy (% recovery), precision (% relative standard deviation, %RSD), and matrix effects using the QuanLynx Method Editor V4.1. The retention time window criterion was set to 0.1 min, and the mass window criterion was set to 0.01 Da. For feature alignment and data processing in UNIFI®, the following parameters were applied: retention time (RT) tolerance of ±0.1 min, target mass tolerance of ±10.0 ppm, fragmentation match tolerance of ±10.0 mDa, and drift time tolerance of 0.2 ms. Lock mass correction was performed in UNIFI® prior to data analysis.

Features were generated using the Marker Matrix functionality in UNIFI® and exported in .csv format for statistical analysis and biomarker discovery in EZinfo (version 3.0.3.0). Principal component analysis (PCA) was employed to identify chemical composition similarities among samples. Orthogonal partial least squares discriminant analysis (OPLS-DA) was used as a supervised chemometric method to differentiate between TB-positive and control groups by assigning TB status to the data. Pareto scaling was applied to create OPLS-DA models, and S-plots were utilised to highlight distinguishing features between the groups.

Compounds contributing to group differences were annotated based on accurate mass and fragmentation patterns by comparison with the Human Metabolome Database (HMDB, version 3.6). A CCS tolerance of 5.0 % was used for library matching. Further confirmation of annotated compounds was performed using the CCS prediction tool, CCSondemand, developed by Waters®. This tool predicts CCS values based on the physiochemical properties of molecular files. The .csv feature table generated in UNIFI® was imported into MetaboAnalyst V6.0 for additional biomarker analyses based on receiver operating characteristic (ROC) curves. Normalisation was done using the pooled samples and data scaling using Pareto scaling. Data filtering was done by removing 40 % of the missing variables. The differential metabolites were defined as follows (1): P-value $\leq$ 0.05 (2); fold change $\geq$ 1.5 and $\leq$ 0.67. The RandomForest classification method was used to evaluate the performance of the model as it is ranked as one of the top algorithms based on accuracy, sensitivity, and specificity for biomarker discovery [23].

3. Results and discussion

The study cohort included 28 participants sampled in 2018. A PDMS sampler from each participant was solvent extracted and analysed in triplicate with a resulting dataset of $N = 84$. The cohort was divided into two categories: TB-positive test patients ($n = 10$, $m = 30$) and TB-negative (healthy) ($n = 18$, $m = 54$) controls. The use of Mylar® and Micropore tape prevented the inclusion of background compounds during sampling. Collecting skin compounds presents a less infectious matrix compared to exhaled breath, which is particularly advantageous for airborne infections. The PDMS sampler offers several advantages, including its portability, wearability, and ease of use for both researchers and patients, without requiring skilled healthcare personnel. Additionally, it does not disrupt medical procedures or conflict with hospital protocols. To address the nonpolar characteristics of PDMS, the silicone rubber band was treated with ethanol, a polar solvent, which improved its ability to accumulate polar compounds [5]. The complexity of the UPLC separation is illustrated in Fig. 2 showing a comparison between a control and TB positive sample, on average 1 200 features (ranging from 352 to 3 583) were detected in an LC analysis.

3.1. Method validation

The performance of the method employing a non-invasive sampler with solvent BE and UPLC-IMS-HRMS was evaluated by plotting multilevel calibration curves (method-matched) using seven concentration levels 10 – 800 ng of *p*-aminobenzoic acid (PABA), l-phenylalanine, l-tyrosine, and l-lysine. The analytes were chosen as they had been previously identified as a potential TB biomarker by Magdalena et al. when analysing serum using LC-MS [6]. The linear regression fit (R^2) for all analytes was above 0.99 (Table 1). A R^2 value of $\geq$ 0.99 was likewise reported by Magdalena et al. for TB-biomarkers in serum for a calibration range of 1–25 ng/ml for PABA [6]. The LODs and LOQs ranged from 6 to 172 ng and 20 to 575 ng, respectively (Table 1). Serum LODs were

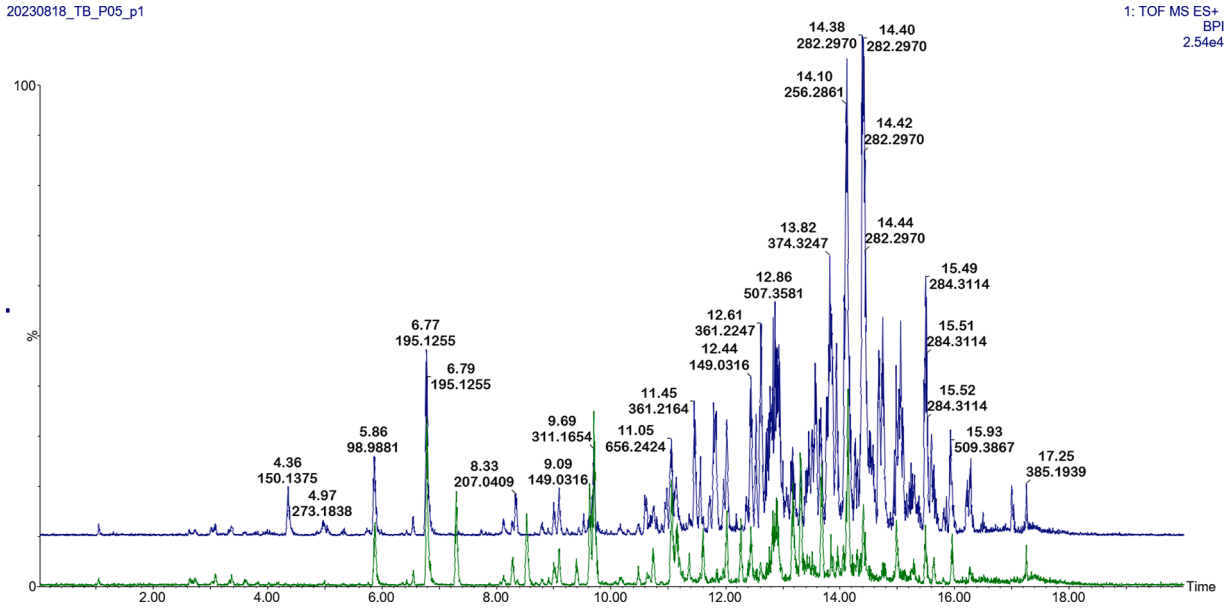

Fig. 2. Representative overlaid base peak ion (BPI) chromatograms of a TB positive individual (top) and TB negative individual (bottom). The chromatograms illustrate the complexity of the LC separation.

Table 1
Method matched calibration: Linearity, LODs and LOQs.

Compound	Quantification Ion (m/z)	Adduct	$^{TW}CCS_{N2}$ (Å^2)[1]	Linearity				LOD[3] (ng)	LOQ[4] (ng)
				Range (ng)	Number of Points	Regression Equation[2] (y=mx+c)	R^2		
PABA	138.0549	$[M+H]^+$	132.4	10 - 800	7	$y = 0.0265x - 0.4208$	0.995	6	20
L-phenylalanine	120.0822	$[M\text{-}COOH]^+$	126.6	10 - 800	6	$y = 0.0027x - 0.0099$	0.988	172	575
L-tyrosine	136.0776	$[M\text{-}COOH]^+$	129.8	10 - 400	5	$y = 0.0022x + 0.0218$	0.989	28	94
L-lysine	130.0874	$[M\text{-}COOH]^+$	125.9	10 - 800	5	$y = 0.0004x + 0.0692$	0.996	51	171

1 travelling wave collision cross section (CCS) experimental values.
2 y = compound peak area; x = compound amount (ng).
3 LOD calculated as the concentration which gives a signal-to-noise ratio (S/N) of 3.
4 LOQ calculated as the concentration which gives a S/N of 10.

stated as $S/N \geq 3$ without reporting the LOD concentrations corresponding to this particular S/N in Magdalena et al. [6]. Percentage recoveries (accuracy) for the target analytes were between 90 % (phenylalanine) and 137 % (tyrosine), while the accuracy (%RSD) range from 3 % (lysine) to 15 % (phenylalanine). Matrix effects were determined by spiking a conditioned PDMS sampler and compared to a PDMS sampler that has been exposed to the human skin surface for 1 hour and then similarly spiked pre-extraction with 100 ng of the working standard solution. Matrix effects were determined to be 77 % for PABA, 99 % for phenylalanine, 46 % for tyrosine, and 43 % for lysine, when comparing peak areas. The significant matrix suppression, 46 % and 43 %, observed for tyrosine and lysine, respectively, suggests potential ionisation suppression by matrix components on the human skin, however the skin matrix will differ between individuals sampled necessitating further investigation to mitigate the effect.

3.2. Biomarkers discovery for TB associated compounds

Multivariate chemometric techniques were employed to investigate chemical differences on the human skin surface to find biomarkers for TB diagnostics using a non-invasive approach. Features were aligned using m/z (Da), chromatographic retention time (min) and ion mobility drift time (ms). Over 14,000 and 594 unique features were identified using ESI+ and ESI- ionisation modes, respectively, employing UNIFI®'s Marker Matrix software function. An unsupervised PCA score plot, Fig. S1, using positive ionisation mode data, yielded two principal components with R2X = 0.381 and Q2 = 0.335 while 13 components explained 71 % of the variance (R2X(Cum)). The ESI- data set yielded two principal components with R2X = 0.763 and Q2 = 0.582 with five components accounting for 86 % of the variance (R2X(Cum)). A supervised OPLS-DA model (Fig. 3A) was used to investigating differences in semi-volatile and non-volatile skin chemical compounds between TB positive (test) patients and control groups. The OPLS-DA model for the ESI+ dataset explained 95 % of the variance (R2Y (cum)) and 67 % of the predicted variance (Q2 (cum)). The OPLS-DA model for the ESI- dataset only explained 22 % of the variance and had no predictive power. Due to the poor predictive nature for the ESI- dataset, it was decided to continue with the ESI+ ionisation mode results only. The Student's t-test revealed that 463 of the biomarkers were significant in contributing to the difference between the test (TB-positive) and control groups ($p < 0.05$). An S-plot (Fig. 3B) was constructed to filter potential biomarkers for TB from the skin surface chemicals using a non-invasive approach (test vs. control). The extreme ends of the S-plot show the compounds responsible for the separation into the two observed groups (Fig. 3B). From the S-plot 40 features were selected (VIP > 1) for further investigation as potential biomarkers.

These markers were further reduced to 13 potential biomarkers by comparison to the top features determined by fold-change analysis, threshold >1.5, (Table S1) and a t-test, threshold > 0.05, (Table S2) using MetaboAnalyst V6.0. A ROC curve analysis was performed on the entire feature table to determine the predictive ability of potential biomarkers (Fig. 4A). When using between ten and 100 features the area under the curve (AUC) is larger than 0.9 indicating the potential diagnostic value of the model. The most significant features using the best ROC model is shown in Fig. 4B. The final list of potential biomarkers was increased to 14 potential biomarkers when considering the ROC prediction features.

A ROC curve analysis was done on the 14 features that showed promise as potential biomarkers for TB diagnostic purposes (Fig. 5A). The area under the curve (AUC) for the model was 0.911 highlighting the predictive power of the model. To ensure the reliability and stability of the supervisory model, we performed 100 permutation tests

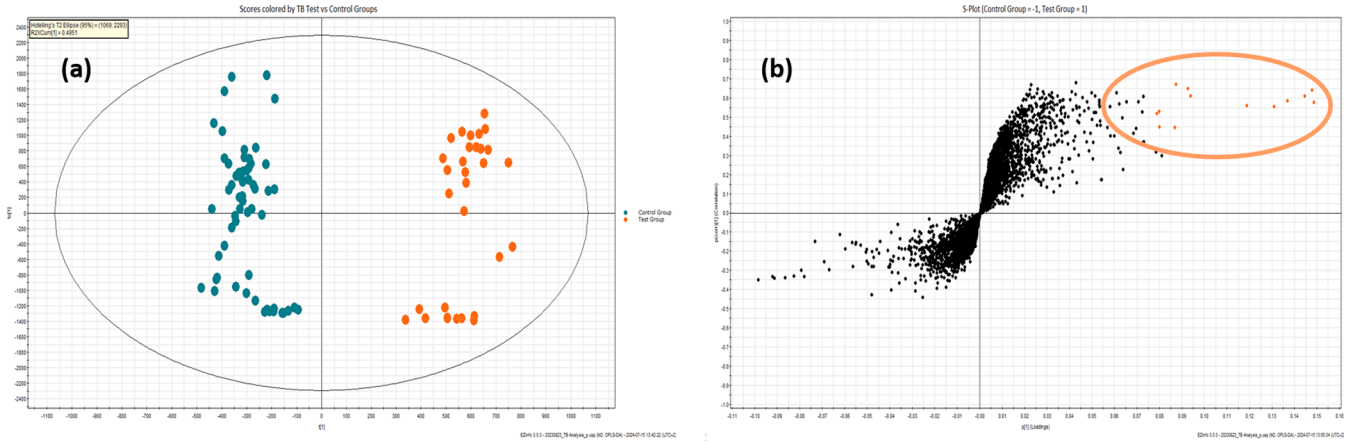

Fig. 3. (A) OPLS-DA score plot using ESI positive mode with UPLC-IMS-HRMS showing inner arm skin surface data, over 14 000 markers, revealing separation between test (TB-positive), $n = 10$, $m = 30$, (orange) and control, $n = 18$, $m = 54$, (blue) groups. (B) S-plot showing features closely related to the test group (orange), and thus potential TB diagnostic markers.

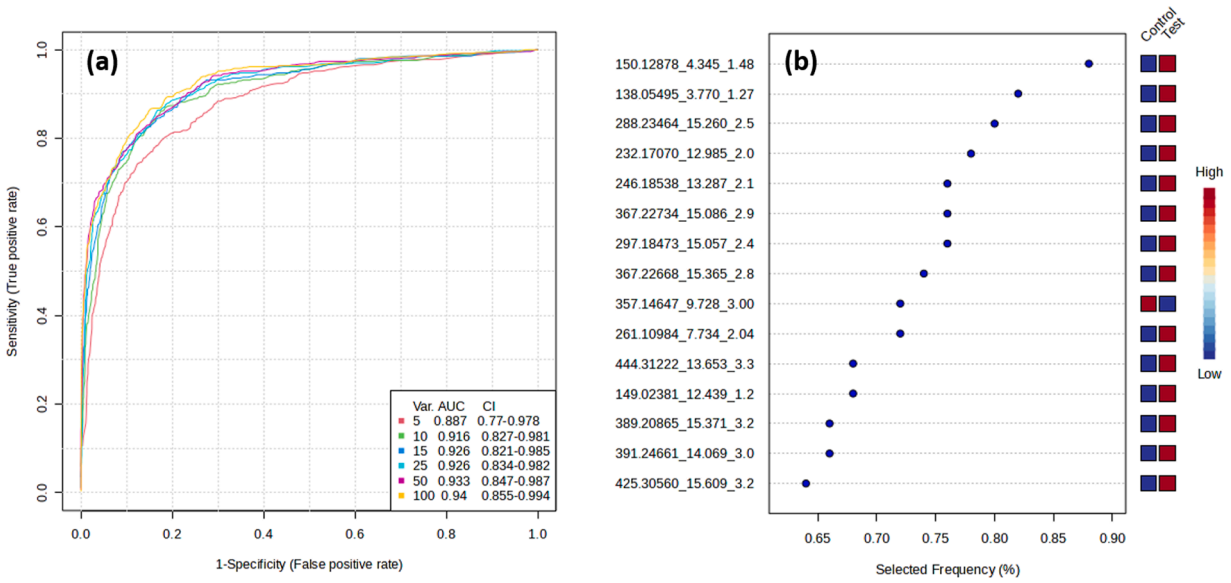

Fig. 4. (A) Plot of ROC curves for all biomarker models based on its average performance across all Monte-Carlo cross validation (MCCV) runs. (B) Plot of the most important features of the best model ranked from most to least important. Features are labelled in the following order: *m/z* (Da)_retention time (min)_drift time (ms). Colour scale on the right indicates the association of features with specific groups. High-intensity red blocks represent strong associations with a group (e.g., the TB-positive test group), while blue blocks indicate weak, or no association, with a group.

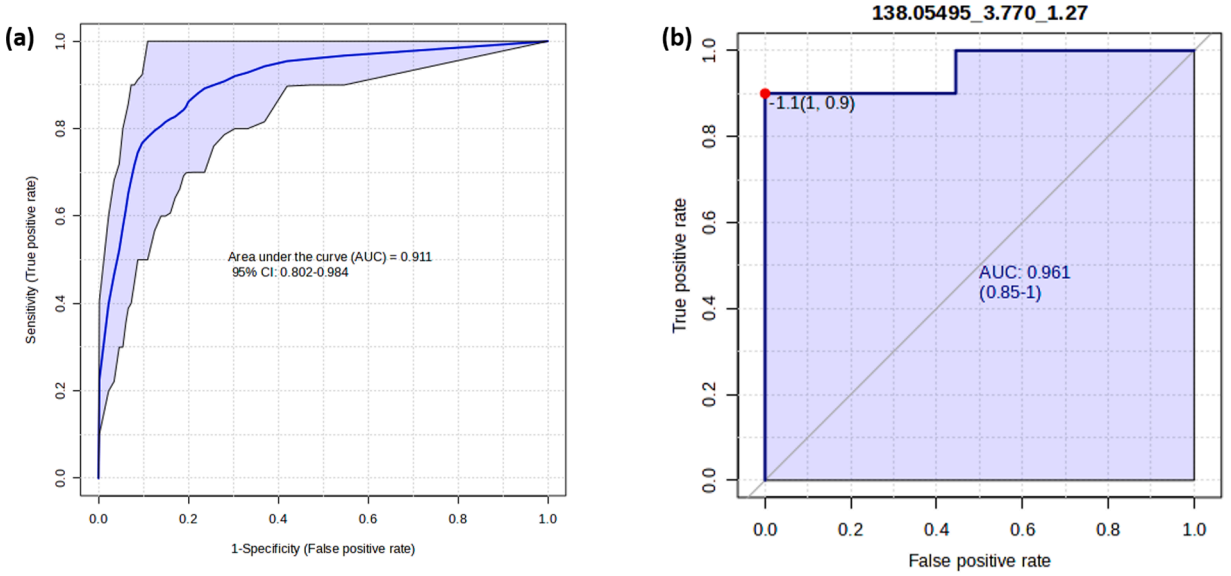

Fig. 5. (A) Plot of ROC curve for the created biomarker model based on its average performance across all Monte-Carlo cross validation (MCCV) runs using ESI+ UPLC-IMS-HRMS. Biomarkers used: 150.12878_4.345_1.48, 337.19933_12.016_2.62, 232.17070_12.985_2.07, 444.31222_13.653_3.30, 274.21794_14.777_2.40, 367.22734_15.086_2.98, 389.20875_15.087_3.37, 288.23464_15.260_2.50, 479.34862_14.371_3.41, 261.10984_7.734_2.04, 315.23182_14.726_2.59, 138.05495_3.770_1.27, 246.18538_13.287_2.18, and 425.30560_15.609_3.27. (B) ROC curve for the individual biomarker PABA, 138.0549 *m/z*, AUC = 0.956 with a specificity of one and a sensitivity of 0.9.

(Figure S2). The 14 compounds were annotated based on comparison with the HMDB and online ChemSpider databases using accurate mass, i-Fit values and, when available, fragmentation patterns. The CCS prediction tool, CCSondemand, further confirmed tentative identification (Table 2). Table 2 gives a count of the number of individuals a compound was detected on. This is after subtracting the detector response from the laboratory blank from that of the participants.

Eight of the 14 annotated markers were identified using the HMDB, showing promise as potential biomarkers for TB diagnostics (Table 2). The predictive ability of each potential biomarkers was assessed using a ROC curve analysis to evaluate for TB diagnostic purposes. The results identified three compounds with potential diagnostic value, namely PABA (AUC = 0.956) and 2,6-diethylaniline (AUC = 0.989, Figure S3A) markers for the test group, and diethyl [4-(2,6-dimethylphenoxy)butyl] malonate a marker for the control group (AUC = 0.9333, Figure S3B). These results compare excellently to Magdalena et al. where AUC-values of 0.519–0.593 were reported for PABA in serum [6]. As the control group marker is present in all samples, the marker assignment is based on increase in detector response in the control group and not absent/present criteria. PABA has a specificity of one and a sensitivity of 0.9 (Fig. 5B), while 2,6-diethylaniline has a specificity of 0.9 and a sensitivity of one. The AUC values for three other differential features were above 0.8, indicating good diagnostic value (Figure S3C-E), namely p-octyloxybenzonitril, 4-dodecyloxybenzonitrile, and *n*-isobutyl-2E,4E,

Table 2
Potential biomarkers on the human skin surface for TB diagnosis using UPLC-IMS-HRMS annotated using accurate mass (observed adduct $[M+H]^+$), fragmentation patterns, and CCS values.

Compound	m/z[1] (mass error [ppm])	RT[2] (min)	$^{TW}CCS^{3}_{N2}$ (Å^2)	Fragment Count	Count[4]	
					Test Group[5] ($n = 10$)	Control Group ($n = 21$)
p-Aminobenzoic acid (PABA)	138.0549 (−0.4)	3.74	135.2	1	10	0
2,6-Diethylaniline	150.1280 (1.7)	4.34	144.4	6	10	0
4-Amino-N'-[(E)-(2-methoxyphenyl)methylene]−1,2,5-oxadiazole-3-carbohydrazonamide[6]	261.1098 (3.6)	7.73	167.8	0	10	2
Diethyl [4-(2,6-dimethylphenoxy)butyl]malonate[6]	337.1993 (5.5)	12.02	193.5	2	10	21
p-Octyloxybenzonitril[6]	232.1695 (−0.4)	12.98	170.4	6	10	0
N-Isobutyl-2E,4E,10Edodecatriene-8-ynamide[6]	246.1859 (2.6)	13.29	174.5	9	10	2
N-Linoleoyl Tyrosine	444.3122 (3.0)	13.65	222.9	1	8	0
2-[{[2-(Dicyclohexylcarbamoyl)cyclohexyl]carbonyl}(2-hydroxyethyl)amino] ethyl propionate[6]	479.3486 (1.4)	14.37	227.8	1	9	1
Progesterone	315.2320 (0.5)	14.72	189.1	4	9	1
N-Isobutyl-2,4,8,10,12-tetradecapentaenamide	274.2172 (2.3)	14.78	182.1	12	8	0
Umbelliprenin	367.2280 (3.4)	15.07	208.8	5	10	0
Ramiprilat	389.2012 (3.7)	15.09	193.5	0	10	0
4-Dodecyloxybenzonitrile[6]	288.2324 (0.7)	15.26	189.8	0	10	1
17-Estradiol cyclooctyl acetate	425.3051 (0.1)	15.61	223.7	0	10	2

1 m/z: mass-to-charge ratio.
2 RT: retention time.
3 travelling wave collision cross section.
4 Number of individuals compound was detected on based on detector counts > lab blank.
5 Annotation based on match to online ChemSpider database.

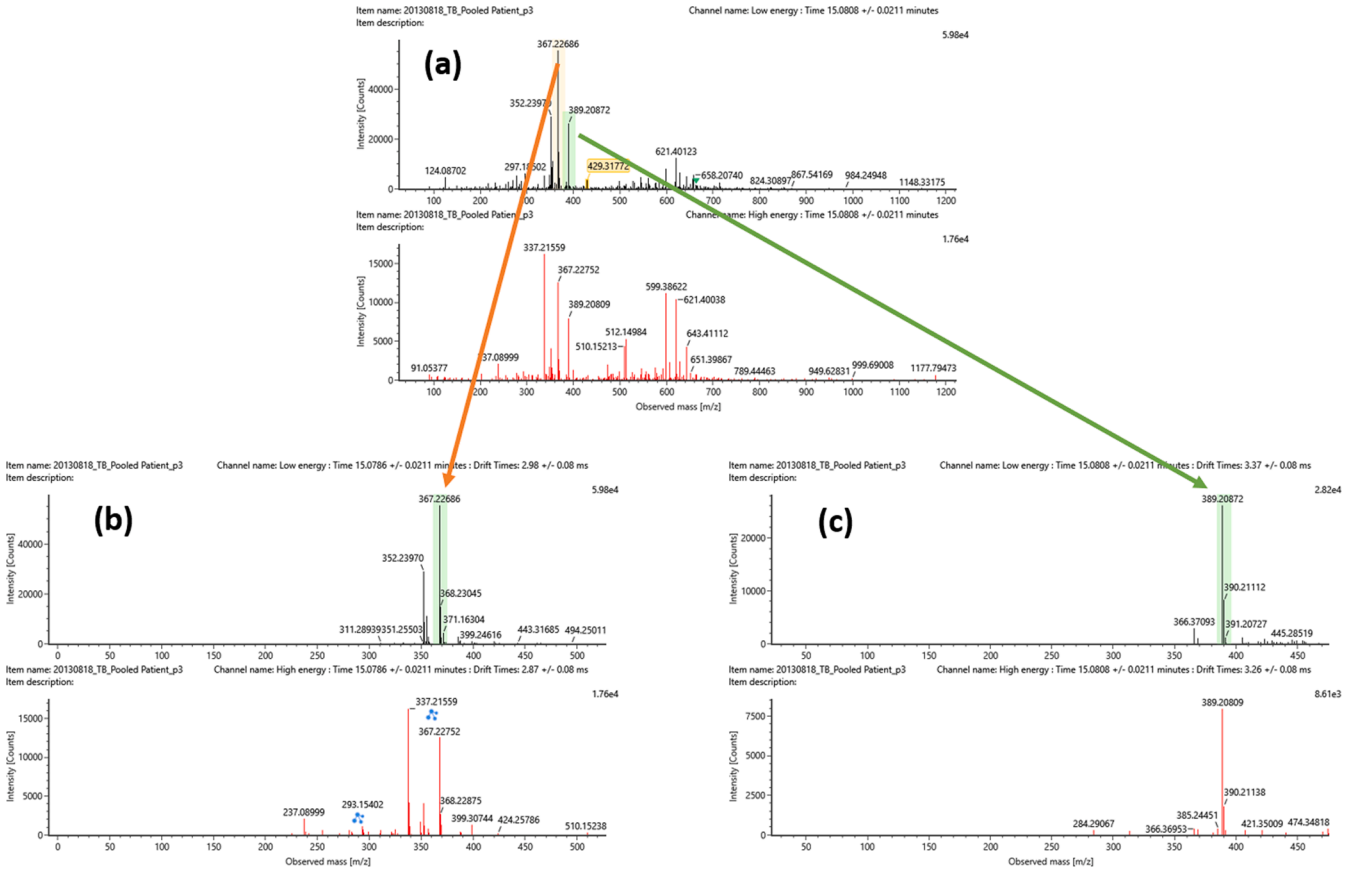

Fig. 6. (A) ESI+, low energy (top) and high energy (bottom), mass spectra showing pre-drift spectral clean-up for biomarkers 367.2273 m/z and 389.2088 m/z at retention time 15.08 min. Spectral clean-up using ion mobility drift time for (B) 367.2273 m/z, umbelliprenin, and (C) 389.2088 m/z, ramiprilat, simplifying the assignment of fragment to precursor ions.

10Edodecatriene-8-ynamide. Notably, the metabolite showing the most promise as TB diagnostic marker, PABA, was not identified when using the S-Plot, *t*-test or fold analysis but only came to light during the ROC analysis (Fig. 4B). PABA is involved in various metabolic pathways, namely aminobenzoate degradation, folate biosynthesis, biosynthesis of phenylpropanoids, biosynthesis of cofactors, and other metabolic pathways (Kyoto Encyclopedia of Genes and Genomes (KEGG)). The metabolites putatively assigned as umbelliprenin, a plant metabolite, as well as 2,6-diethylaniline, which has as role as nutrient (HMDB), could be indicative as potential self-medication of patients with herbal or other remedies prior to hospital treatment. The putatively assigned biomarker, ramiprilat, a metabolite of ramipril which is used to treat high blood pressure, was detected in all hospitalised patients but not in the control group, indicating that this compound is not a marker for TB but may serve as an indicator of hospitalisation, thereby demonstrating the applicability of the analytical method employed. However, this needs to be further investigated.

The incorporation of drift time for compound separation demonstrates significant benefits in the annotation of biomarkers, as demonstrated by biomarkers 367.22734_15.086_2.98 and 389.20875_15.087_3.37 (*m/z*_retention time (min)_drift time (ms)). These compounds exhibit identical retention times, complicating the mass spectrum. However, the application of ion mobility and post-drift fragmentation (Fig. 6) effectively achieves spectral clean-up, allowing for the clear differentiation between umbelliprenin and ramiprilat. This method enhances spectral clarity and enables more accurate assignment of fragment ions to their precursor ions, thereby simplifying compound annotation through ion mobility drift.

3.2.1. Targeted analysis to known TB-associated compounds

To evaluate skin as a diagnostic media we screened the dataset for compounds associated with TB, as reported in the literature from serum [6]. The following compounds were therefore included in the initial screening for TB-associated compounds: PABA, pyridoxal, isopyridoxal, inosine, phenylalanine, kynurenine, neopterin, isoleucine/leucine, tryptophan, methionine, proline, valine, tyrosine, pyroglutamic acid, cisteine, alanine, hydroxyproline, threonine, GABA, glycylo-l-valine, glycine, glutamine, asparagine, serine, citrulline, GSH, arginine, lysine, and cystine [6]. The PDMS sampler was previously validated for its ability to detect amino acids on the human skin surface in a mosquito host biting context [14] confirming the suitability of the method for TB diagnostic purposes. Magdalena et al. (2022) identified PABA and kynurenine as TB biomarkers of interest in serum and blood cultures as these had the best AUC when comparing uninfected healthy controls and non-mycobacterial pneumonia patients to latently *M.tb* infected individuals and TB active patients when using ROC analysis [6]. This is in line with our findings for PABA as a highly promising biomarker for TB diagnosis via human surface skin chemicals. However, kynurenine was detected on only four TB positive individuals and was also present for six control participants, thereby showing less potential as a biomarker. The following literature TB biomarkers as reported in other biofluids were also detected in our study on the human skin surface: cystine, glycylo-l-valine, citrulline, valine, and inosine. However, using multivariate analyses these showed limited potential as TB biomarkers. These results demonstrate the ability of the PDMS sampler, the sorptive non-invasive sampling technique, and simplified extraction method to detect TB biomarkers previously described in literature when sampling skin. Furthermore, these findings confirm the stability of the skin sampler matrix for metabolomics even with extended storage periods, comparing remarkably well to blood which can be stored up to 30 months at −80 °C [24].

3.2.2. Quantification of a potential biomarker

Four compounds were unequivocally identified on the human skin from volunteers using reference standards, a passive sampling approach and UPLC-IMS-HRMS. PABA was detected on average at ≥ 800 ng (~1221 ng semi-quantified) on the TB patients for a one-hour sampling period, and not detected on any of the control group individuals sampled (Figure S4). The amount PABA detected ranged from 52 to ≥ 800 ng (~4198 ng semi-quantified), demonstrating the quantification range of the sampler. In comparison Magdalena et al. reported an upper quantification limit of 25 ng/ml for serum [6]. Phenylalanine was detected on four control volunteers and seven TB positive individuals (<LOQ). Lysine was detected on nine control volunteers and six TB positive individuals (<LOQ); two samples were above the LOQ for the lysine (135 ng detected on a TB patient and 120 ng on a control volunteer). Tyrosine was detected on average at 592 ng (n = 5) for the TB positive individuals, and at 409 ng (n = 10) on the control volunteers. These results demonstrated the methods' capabilities in quantifying TB biomarkers on the human skin surface using non-invasive sampling.

4. Conclusion

We report a non-invasive analytical metabolomic approach for the identification of biological markers for TB diagnostic purposes form human skin. The simplified and cost-effective sampling strategy enabled the detection of 14 lead compounds of diagnostic importance. Two, namely PABA and 2,6-diethylaniline, showed excellent potential as TB diagnostic markers. 2,6-Diethylaniline has not previously, to the authors' knowledge, been reported as a potential TB marker. Coupling of high-resolution mass spectrometry and CCS data provided additional confidence in marker identification, as does post-drift fragmentation alignment of product and precursor ions. Marker alignment for multivariate statistics was strengthened using three unique features, namely *m/z*, chromatographic retention time and ion mobility drift time. We have demonstrated that skin semi-volatiles and non-volatile compounds can discriminate between TB-positive and TB-negative individuals. The skin test has great merit as a promising tool for diagnostic protocols, and as such warrants further investigation involving a larger study cohort.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Madelien Wooding: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Methodology, Investigation, Conceptualization. **Kornelis van Pletzen:** Investigation, Formal analysis. **Yvette Naudé:** Writing – review & editing, Supervision, Resources, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no conflicts of interest related to this manuscript. There are no personal or financial interests that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors gratefully acknowledge the late Prof Anton Stoltz (†2020), Department of Internal Medicine, Division of Infectious Diseases, University of Pretoria, for access to patients and funding, we thank Dr Katherine Anderson for phlebotomy services, Ms Tebogo Magongoa for assistance with the hospital administration, Ms Marizane du Pisanie for the NHLS labtrek services, Ms Portia Makhubela for sampling of the volunteers, Ross McAdam for assisting with the method validation, and Prof Egmont Rohwer for his original concept of sampling skin surface chemicals with PDMS.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jcoa.2025.100204.

Data availability

Data will be made available on request.

References

[1] World Health Organization, Global Tuberculosis Report 2023, 2023.
[2] S.D. Lawn, A.I. Zumla, Tuberculosis, Lancet 378 (2011) 57–72, https://doi.org/10.1016/S0140-6736(10)62173-3.
[3] B.M. Nogueira, S. Krishnan, B. Barreto-Duarte, M. Araújo-Pereira, A.T. Queiroz, J. J. Ellner, P. Salgame, T.J. Scriba, T.R. Sterling, A. Gupta, Diagnostic biomarkers for active tuberculosis: progress and challenges, EMBO Mol. Med 14 (2022) e14088, https://doi.org/10.15252/emmm.202114088.
[4] M. Gengenbacher, S.H.E. Kaufmann, Mycobacterium tuberculosis: success through dormancy, FEMS Microbiol. Rev. 36 (2012) 514–532, https://doi.org/10.1111/j.1574-6976.2012.00331.x.
[5] P.C. Makhubela, E.R. Rohwer, Y. Naudé, Detection of tuberculosis-associated compounds from human skin by GCxGC-TOFMS, J. Chromatogr. B. 1231 (2023) 123937, https://doi.org/10.1016/j.jchromb.2023.123937.
[6] D. Magdalena, S. Michal, S. Marta, K.-P. Magdalena, P. Anna, G. Magdalena, S. Rafał, Targeted metabolomics analysis of serum and Mycobacterium tuberculosis antigen-stimulated blood cultures of pediatric patients with active and latent tuberculosis, Sci. Rep. 12 (2022) 4131, https://doi.org/10.1038/s41598-022-08135-2.
[7] R. Vishinkin, R. Busool, E. Mansour, F. Fish, A. Esmail, P. Kumar, A. Gharaa, J. C. Cancilla, J.S. Torrecilla, G. Skenders, M. Leja, K. Dheda, S. Singh, H. Haick, Profiles of volatile biomarkers detect tuberculosis from skin, Adv. Sci 8 (2021) 2100235, https://doi.org/10.1002/advs.202100235.
[8] M. Beccaria, C. Bobak, B. Maitshotlo, T.R. Mellors, G. Purcaro, F.A. Franchina, C. A. Rees, M. Nasir, A. Black, J.E. Hill, Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques, J. Breath Res. 13 (2019) 016005, https://doi.org/10.1088/1752-7163/aae80e.
[9] D. Beukes, M. van Reenen, D.T. Loots, I. du Preez, Tuberculosis is associated with sputum metabolome variations, irrespective of patient sex or HIV status: an untargeted GCxGC-TOFMS study, Metabolomics 19 (2023) 55, https://doi.org/10.1007/s11306-023-02017-7.
[10] N.A. Dang, H.-G. Janssen, A.H.J. Kolk, Rapid diagnosis of Tb using GC–MS and chemometrics, Bioanalysis. 5 (2013) 3079–3097, https://doi.org/10.4155/bio.13.288.
[11] F. Buljubasic, G. Buchbauer, The scent of human diseases: a review on specific volatile organic compounds as diagnostic biomarkers, Flav. Fragr. J 30 (2015), https://doi.org/10.1002/ffj.3219.
[12] C. Turner, B. Parekh, C. Walton, P. Španěl, D. Smith, M. Evans, An exploratory comparative study of volatile compounds in exhaled breath and emitted by skin using selected ion flow tube mass spectrometry, Rapid Commun. Mass Spectrom 22 (2008) 526–532, https://doi.org/10.1002/rcm.3432.
[13] Y. Cho, Y. Park, B. Sim, J. Kim, H. Lee, S.-N. Cho, Y.A. Kang, S.-G. Lee, Identification of serum biomarkers for active pulmonary tuberculosis using a targeted metabolomics approach, Sci. Rep. 10 (2020) 3825, https://doi.org/10.1038/s41598-020-60669-0.
[14] M. Wooding, T. Dodgen, E.R. Rohwer, Y. Naudé, Advancing the analytical toolkit in the investigation of vector mosquito host biting site selection, J. Mass Spectrom. 59 (2024) e4992, https://doi.org/10.1002/jms.4992.
[15] M. Wooding, E.R. Rohwer, Y. Naudé, Chemical profiling of the human skin surface for malaria vector control via a non-invasive sorptive sampler with GC× GC-TOFMS, Anal. Bioanal. Chem. 412 (2020) 5759–5777, https://doi.org/10.1007/s00216-020-02590-2.
[16] A. De Villiers, P. Venter, H. Pasch, Recent advances and trends in the liquid-chromatography–mass spectrometry analysis of flavonoids, J. Chromatogr. A. 1430 (2016) 16–78, https://doi.org/10.1016/j.chroma.2015.12.013.
[17] A.P. Roodt, Y. Naudé, A. Stoltz, E. Rohwer, Human skin volatiles: passive sampling and GC× GC-ToFMS analysis as a tool to investigate the skin microbiome and interactions with anthropophilic mosquito disease vectors, J. Chromatogr. B. 1097 (2018) 83–93, https://doi.org/10.1016/j.jchromb.2018.09.002.
[18] M. Wooding, T. Dodgen, E.R. Rohwer, Y. Naudé, Mass spectral studies on the human skin surface for mosquito vector control applications, J. Mass Spectrom. 56 (2021) e4686, https://doi.org/10.1002/jms.4686.
[19] Y. Maitani, H. Sato, T. Nagai, Effect of ethanol on the true diffusion coefficient of diclofenac and its sodium salt in silicone membrane, Int. J. Pharm. 113 (1995) 165–174, https://doi.org/10.1016/0378-5173(94)00192-8.
[20] P. Serodio, J. Nogueira, Development of a stir-bar-sorptive extraction–liquid desorption–large-volume injection capillary gas chromatographic–mass spectrometric method for pyrethroid pesticides in water samples, Anal. Bioanal. Chem. 382 (2005) 1141–1151, https://doi.org/10.1007/s00216-005-3243-1.
[21] C. Margoum, C. Guillemain, X. Yang, M. Coquery, Stir bar sorptive extraction coupled to liquid chromatography-tandem mass spectrometry for the determination of pesticides in water samples: method validation and measurement uncertainty, Talanta 116 (2013) 1–7, https://doi.org/10.1016/j.talanta.2013.04.066.
[22] Z. Szelényi, S. Komoly, Thermoregulation: From basic Neuroscience to Clinical neurology, Part 2, Taylor & Francis, 2019.
[23] V. Balakrishnan, Y. Kherabi, G. Ramanathan, S.A. Paul, C.K. Tiong, Machine learning approaches in diagnosing tuberculosis through biomarkers-a systematic review, Prog. Biophys. Mol. Biol. 179 (2023) 16–25, https://doi.org/10.1016/j.pbiomolbio.2022.10.002.
[24] V.L. Stevens, E. Hoover, Y. Wang, K.A. Zanetti, Pre-analytical factors that affect metabolite stability in human urine, plasma, and serum: a review, Metabolites 9 (2019) 156, https://doi.org/10.3390/metabo9080156.