Non-Volatile Organic Compounds in Exhaled Breath Particles Correspond to Active Tuberculosis

Dapeng Chen (dapeng.chen@zeteotech.com)
Zeteo Tech, Inc
Noella A. Bryden
Zeteo Tech, Inc
Wayne A. Bryden
Zeteo Tech, Inc
Michael McLoughlin
Zeteo Tech, Inc
Dexter Smith
Zeteo Tech, Inc
Alese P. Devin
Zeteo Tech, Inc
Emily R. Caton
Zeteo Tech, Inc
Caroline R. Haddaway
Zeteo Tech, Inc
Michele Tameris
University of Cape Town
Thomas J. Scriba
University of Cape Town
Mark Hatherill
University of Cape Town
Sophia Gessner
University of Cape Town
Digby F. Warner
University of Cape Town
Robin Wood
University of Cape Town

Research Article

Keywords: non-volatile organic compounds, exhaled breath aerosol, NOC signatures

Posted Date: January 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1142980/v1

Abstract

Human breath contains trace amounts of non-volatile organic compounds (NOCs) which might inform non-invasive methods for evaluation of individual health. In previous work, we demonstrated that lipids detected in exhaled breath aerosol (EBA) could be used as markers of active tuberculosis (TB). Here, we advanced our analytical platform in characterizing small metabolites and lipids in EBA samples collected from participants enrolled in clinical trials designed to identify molecular signatures of active TB. EBA samples from 26 participants with active TB and 73 healthy participants were processed using a dual-phase extraction method, and metabolites and lipids were identified via mass spectrometry (MS) database matching. In total, 13 metabolite and 9 lipid markers were identified with optimized relative standard deviation values that were statistically different between individuals diagnosed with active TB and the healthy controls. A feature ranking algorithm reduced this number to 10 molecules, with the membrane glycerophospholipid, phosphatidylinositol 24:4, emerging as top driver of segregation between the two groups. These results support the utility of this approach to identify consistent NOC signatures from EBA samples in active TB cases and suggest the potential to apply this method to other human diseases which alter respiratory NOC release.

Introduction

Human exhaled air contains water vapor and trace amounts of organic materials including volatile organic compounds (VOCs), non-volatile organic compounds (NOCs), and particulate matter including microbes [1–5]. Molecular characterization of these organic compounds provides a noninvasive method for the investigation of human diseases [5]. Typically, human breath analysis uses a pre-concentration step since the amounts of organic and biological molecules are very low, usually in the parts per billion to parts of trillion range [1].

Since the identification of small gas molecules can be achieved by relatively simple mass spectrometry (MS)-based platforms, human breath analysis has focused on VOCs [6]. For example, an MS with moderate resolution (R=1,000) has the capacity to resolve the molecular patterns of gas molecules. However, VOC analysis has a fundamental drawback as the endogenous metabolic origins of the molecules detected are poorly understood [6]. Consequently, changes in VOC patterns might be caused by factors including internal gut and airway bacteria, host physiological state, or exogenous influences such as drinking, smoking, and eating [7]. Changes in VOCs can also be caused by the inhalation of gases generated from environmental sources such as breath samplers and bioaerosols [6, 7]. As a result, changes in VOC fingerprints acquired from non-targeted sensors, such as MS and ion mobility spectrometry, do not necessarily represent individual phenotypes and could be falsely interpreted as disease biomarkers by pattern recognition tools, limiting their application in medical diagnosis [6, 7]. Therefore, there is a need for a comprehensive assessment of biomolecules in human breath for diagnostics, individual phenotyping, and respiratory disease monitoring.

NOCs in human exhaled air include metabolites, lipids, and proteins that are contained in exhaled water vapor (exhaled breath condensate, EBC) and exhaled breath aerosols (EBA) [5]. To date, the major barriers to achieving a better understanding of nonvolatile molecules has been instrumentation limitations and technical difficulties in the analysis of these large molecules. However, recent advances in analytical techniques, such as the invention of MS ionization methods and ultrahigh-resolution MS, have overcome these hurdles, opening exciting possibilities for the characterization of NOCs in human breath [5, 8]. Given increasing evidence that particles contained in human exhaled air are associated with disease transmission, contributing to public health risks such as seasonal influenza, tuberculosis (TB), and the COVID-19 pandemic [4, 9], collection and characterization of NOCs are critical to establish candidate biomarkers for the detection of respiratory infection, medical diagnosis, disease screening and performance metrics [5, 10–12]. In previous work, we demonstrated that lipid markers detected using high-resolution MS could be used to identify active TB cases [5]. Here, we applied advanced analytical approaches to 200 EBA samples acquired from clinical trials and evaluated the correlation between respiratory NOCs and TB status [13].

Results

NOC analysis of EBA samples from CORTIS participants

In total, 99 CORTIS trial participants classified as either GeneXpert (GXP)-negative or GXP-positive were included in this study, from whom 200 EBA samples (include some revisits for some of the participants) were collected at different clinical visits (Figure 1 and Supplementary Table 1). For the 1st visit, there were 26 GXP-negative participants and 20 GXP-positive participants; for the 2nd visit, there were 63 GXP-negative participants and 11 GXP-positive participants; and for the 3rd visit, there were 53 GXP-negative and 11 GXP-positive participants. All EBA samples were processed with 0.2 μm membrane filtration (see methods) in which the bioaerosol pellet was separated from the supernatant. The concept of using dual-mobile phase extraction was based on the hydrophobicity of molecules. Since lipids are non-polar molecules 70% IPA is used for elution [18]. The dual-mobile phase method enabled us to extract both small metabolites (50% acetonitrile solvent) and lipids from the same EBA sample, deepening the molecular identification and analysis in LC-MS/MS (Figure 1).

Validation of the dual-mobile phase extraction method

In our previous work, lipids were extracted using a classic Folch solvent separation method [5]. Since the EBA samples were collected in ~10 mL of buffer solution, they required overnight lyophilization and centrifugation for extraction (Figure 2A). These steps are resource intensive and difficult to manage at scale; therefore, in the current study, we applied our newly developed solid phase extraction (SPE) approach which uses a C18 resin column as the capture matrix. For the sample preparation, EBA samples were loaded directly onto the column. After washing, small metabolite molecules and proteins, which are less nonpolar, were eluted using acetonitrile (1st elution). Lipids, which are nonpolar, were eluted with 2-propyl alcohol (2nd elution, Figure 2A). Three representative molecules – methadone, dilauroyl-sn-glycero-

3-phosphorylcholine (DLPC), and insulin – were used for method validation (Figure 2B). Methadone and insulin were only detected in the 1st elution sample, not in the washed sample (Figure 2C, red and green dots). DLPC signals were detected in the 2nd elution sample, which suggests the efficient capture of lipid molecules by the C18 resin. The SPE-based sample preparation was rapid, with the separation process taking ~3 min *per* sample, and the use of a C18 resin cartridge makes it amenable to automation. Notably, dual mobile phases for polar and nonpolar molecule separation have been used by others, reinforcing the reliability and reproducibility of this approach [19].

Molecular profiles of EBA samples with liquid chromatogram and MS

The total ion chromatogram of extracted small metabolites and lipids in EBA samples showed that molecular signals were acquired in EBA samples when compared to blank samples (Figure 3A, blue, green, and orange lines). Representative total ion chromatograms of small metabolites in EBA samples from GXP-positive and GXP-negative participants revealed no obvious differences between the two groups, suggesting a deeper analysis was required (Figure 3B). In general, ~500 features were extracted from small metabolite analysis and ~330 features were extracted from lipid analysis (Figure 3C,D). No statistical difference was observed between the two groups based on the feature numbers. Both small metabolite and lipid analyses with MS achieved outstanding dynamic range: ~5 magnitudes in small metabolite profiles and ~4 magnitudes in lipid profiles (Figure 3E,F).

Selection of definitive molecules in each study group

A global correlation heat map demonstrated that molecular profiles could be used to segregate the two groups of study participants with GXP-negative and GXP-positive participants exhibiting clear correlations by Pearson correlation coefficient analysis (Figure 4A). The relative standard deviation of each feature of either GXP-negative or GXP-positive samples was calculated (Figure 4B-E). A threshold of 30% was used to select features in each group for statistics (Figure 4B-E, red line), resulting in 347 small metabolites and 217 lipids in the GXP-negative group, and 325 small metabolites and 198 lipids in the GXP-positive group (Figure 5A). The ion intensity distributions of 10 representative molecules in each participant group were identified as proline, all-trans-retinoic acid (RA), chalcone (CC), 5-hydroxyindole-3-acetic acid (HIAA), D-2-Aminobutyric acid (D2AA), uridine, cholesteryl ester (CE) 16:4, ceramide (Cer) 8:0, diacylglycerol (DG) 21:1, phosphatidylethanolamine (PE) 24:2, phosphatidylinositol (PI) 20:3, and phosphatidylserine (PS) 5:3 (Figure 5B).

Identification of 22 metabolite and lipid molecules associated with active TB status

Correlations between metabolites and lipid profiles detected in GXP-negative and GXP-positive groups were visualized using pairwise combinations of three clinical visits (Figure 6A) and those metabolites and lipids that were statistically significant by t-test identified in the volcano plots (Figure 6B, C). The molecules that were statistically significant at all three visits were marked as either red dots, meaning increased prevalence in GXP- positive group, or green dots, meaning decreased prevalence in GXP-

positive group. In summary, 22 molecules – including 13 metabolites and 9 lipids – were identified that were statistically significant between the two groups at all three visits (Figure 6 B, C and Figure 7A, B).

We then evaluated the utility of the identified metabolites and/or lipids to distinguish GXP-negative from GXP-positive participants by generating ROC curves and calculating AUC values. The AUC for metabolites was ~87% (95% confidence interval: 0.832-0.919) and 93% (95% confidence interval: 0.889-0.971) for lipids to segregate between GXP-negative and GXP-positive study participants. When combined with metabolites and lipids, the AUC was slightly higher than using lipids only, ~97% (95% confidence interval: 0.926-0.986) in the segregation between GXP-negative and GXP-positive participants (Figure 7C). The correlation between the identified molecules was investigated by Pearson correlation coefficient analysis (Figure 7D), with the heatmap identifying seven molecules (Cer 8:0, uridine, PS 24:4, DG O-8:0, NAM, PI 20:4, PI 18:4) with the tightest correlation. In our previous work, we used significance analysis of microarrays (SAM) to identify those features with the most segregation power [5]. Applying SAM analysis to the current dataset revealed 10 molecules (Figure 7E) that were significantly different between GXP-positive and GXP-negative samples, with PS 24:4 emerging as top driver of segregation between the two groups.

Discussion

Investigations of the molecular composition of human breath have predominantly focused on VOCs [6]. This bias is driven largely by the availability of analytical tools to extract and evaluate VOCs in a comprehensive manner, such as gas absorption tubes, electronic chemical sensors, and low-resolution MS [6,7]. As a result, thousands of gas molecule signatures can be acquired and used for pattern recognition and fingerprint analysis [1]. However, there is a critical need for improved tools for NOC analysis given their biological relevance, include both host and pathogen proteins, lipids, metabolites, and nucleic acids [2-5, 14-15]. In this study, we advanced the analytical approach to extract NOCs from EBA samples that were collected from well-defined study groups [13]. Coupling our dual mobile phase extraction methodology with high-resolution MS, rich molecule signatures were acquired. Moreover, we demonstrated the potential to use the profiles of metabolites and lipids to separate active TB cases from other study subjects, and to identify individual molecules providing the greatest segregation power.

The presence in NOCs of polar metabolites and non-polar lipids poses a challenge in applying a universal extraction method for the comprehensive characterization of molecules in EBA. Phase extraction methods based on organic solvents, such as Folch extraction, can be used but are resource intensive and time-consuming as several organic solvents and drying procedures are required [5]. Therefore, while improvements have been made, solvent-based molecule extraction methods are not useful for sample processing at scale. Inspired by absorption materials used in conventional liquid chromatography and solid-phase extraction, we developed a method, dual-mobile phase extraction, using a self-packed C18 column to capture both polar and non-polar molecule in the same sample [18]. The method is based on the different affinity of metabolites and lipids to the C18 functional group as conventional acetonitrile can be used to elute metabolites and a harsher solvent, IPA, is required to elute non-polar lipids [18]. In

fact, switching the solvent system to achieve a more comprehensive analysis has precedent in on-line liquid chromatography [19]. We validated our approach using three different representative molecules: methadone for metabolites; DLPC, a lung surfactant, for lipids; and insulin for small peptides and proteins. Our results indicated that the method is rapid and complete. Indeed, using this approach, we extracted hundreds of molecule features from individual EBA sample. Although proteins are of great interest for biological analysis given that their origin (host *versus* pathogen) can be easily ascribed, our initial analysis yielded no to minimal protein content in EBA samples. Consequently, our extraction and analysis focused on metabolites and lipids – though it is likely that an additional filter cutoff may be applied to the separation from small metabolites and large protein molecules.

MS generates thousands of signals suggesting feature pre-selection is required to achieve confidence in the analysis [5]. In our study, we aimed to select the molecules most definitively associated with disease status by revealing the precision and repeatability of individual molecule in GXP-negative and GXP-positive study groups. To this end, background ion signals and signals contributing to non-biological information were first excluded by the control blank samples. In addition, to improve data repeatability, a desired relative standard deviation cutoff was applied to identified molecules in each group, resulting in a list of features that showed the least variation. The advantage of this approach is that molecules of statistical significance will be more reproducible as more study subjects are available. However, the limitations of this approach should be recognized, the main drawback being the potential for this feature reducing approach to eliminate important features, a real risk in studies with small numbers of subjects [5]. An alternative solution is to assign the false-discovery rate (FDR) to each feature and only significant features are included in data analysis. In addition, due to the nature of ion injection in the orbitrap mass spectrometer, features with low ion intensity may suffer a higher internal variability and is more vulnerable to be excluded [5, 8].

TB molecular diagnosis relies heavily on nucleic acid amplification technologies (NAATs) such as GeneXpert MTB/RIF assay [20]. An acknowledged limitation of this assay is its dependence on the production of a high-quality sputum specimen [20]. Therefore, there is an interest in the development of non-sputum specimen collection method and molecular diagnosis assay that does not require expensive reagents [5]. Considering MS has been used in the clinical labs for microorganism identification, we evaluated if molecule signatures revealed by MS acquired in EBA samples can be used for potential biomarkers in active TB cases.

Our approach identified 22 molecules that were statistically different between GXP-negative and GXP-positive cases. For biomarker discovery using metabolomic and lipidomics, it is expected that a prediction model that combining a group of molecules is established to predict the outcome. However, to avoid over-fitting, the prediction model needs to include the simplest combination of features that provide the most precise prediction. Several artificial intelligences/machine learning (AI/ML) programs, such as neural network and decision tree, can be used for this purpose [5, 21]. Nevertheless, in this study we investigated the correlation of metabolite and lipids and the importance of each molecule via Pearson's and SAM. The results indicated lipid molecules offered a better segregation power than metabolites.

More specifically, lipid molecules showed a stronger correlation as well as ranked in the feature importance evaluation. This notion is further supported by the ROC curves and AUC calculation as lipids alone contributed to ~93% segregation and adding metabolite markers only improved the score very slightly.

We previously demonstrated that molecules identified from human exhaled air samples overlapped with molecules present in blood. In the current work, we identified molecules which have been reported in other studies in which blood specimens were used in biomarker discovery for active TB [22-24]. For example, Feng and coauthors investigated serum metabolic profiles of 271 participants using LC-MS and reported that 12 metabolites contributed to the segregation between TB active and control groups with multiple logistic regression analysis [23]. Among those metabolites, palmitic acid was found to be decreased in the active TB group, which is consistent with our observation that palmitic acid was decreased in EBA samples collected from active TB cases. In another study, Weiner and coauthors reported a serum metabolic panel that could be used to distinguish between TB patients and healthy individuals [24]. That panel comprised 20 metabolites, including phenylalanine which was slightly elevated in active TB patients, consistent with our findings. Using non-targeted metabolic analysis of plasma samples, Frediani and coauthors reported over 23,000 metabolites, of which 61 were found to be associated with pulmonary TB [23]; these included phosphatidylinositol, which was identified from one of 3 patients with multidrug-resistant tuberculosis (MDR-TB) and only one of 17 smear-negative adults, suggesting that this molecule was upregulated in TB subjects. In our study, two phosphatidylinositol species, PS 18:4 and PS 20:4, were increased in active TB cases, with PS 24:4 showing the most segregation power. Since this molecule is present commonly in both humans and *M. tuberculosis*, the source in active TB cases is unclear. On the other hand, it was noted that NAM , a common precursor of bacterial cell wall peptidoglycan, was much higher in GXP-positive cases.

Conclusions

In this study, we demonstrated that rich NOCs can be extracted from human exhaled breath by using advanced analytical approach and the analysis could provide an attractive solution for human disease diagnosis. We advanced NOC analysis in human breath in several aspects. Using a comprehensive molecule extraction method, we have shown that hundreds of metabolites and lipids can be collected from human breath. Most importantly, the extraction method is rapid and has the potential to be used for high-volume sample processing in a clinical setting. By integrating MS-based non-targeted omics and sophisticated data analysis, we demonstrated that identified molecule signatures can segregate active TB cases from healthy individuals with the importance of each molecule was assigned.

Materials And Methods

Ethics statement and patient information

Written informed consent was acquired from all study participants and the study was approved by the University of Cape Town Human Research Ethics Committee (Reference number HREC680/2013). All experiments conducted in this study were performed in accordance with the relevant guidelines and regulations. 200 EBA samples were collected from 99 study participants with different visits using the methods described previously (Table 1 and Supplementary Table 1). Generally, exhaled breath aerosols and particles were collected in 10 mL of phosphate-buffered saline solution (PBS) using a wetted-wall cyclone air sampler (Coriolis μ Biological Air Sampler, Bertin Instruments, Montigny-le-Bretonneux, France) located in the respiratory aerosol sample chamber (RASC) [5, 14-15]. The collection for each study participant was conducted for 60 minutes with an air flow rate of 240 liters per minute. After collection, the liquid samples were processed with 0.2 μm polycarbonate filters (Sterlitech Corporation, WA, USA) to collect the bioaerosol pellets. The TB state of the study participants was based on analysis of the bioaerosol pellet using GeneXpert PCR. The participants were identified as GeneXpert MTB/RIF (GXP-positive) or GXP-negative (control group). The supernatant from the filtration step was processed by dual phase extraction and analyzed by mass spectrometry (LC-MS/MS) as described in the following.

Dual-phase extraction of metabolites and lipids in EBA samples

Chemicals and reagents were used in this study are either HPLC or MS-grade and commercially acquired from Fisher Scientific Chemical (Thermo Fisher Scientific, MA, USA). (±)-Methadone solution and insulin from bovine pancreas were purchased from MilliporeSigma (Burlington, MA). 1,2-Dilauroyl-sn-glycero-3-phosphorylcholine (DLPC) was purchased from Matreya, LLC (State College, PA). Filters and columns were purchased from Boca Scientific (Dedham, MA). 20 μm C18 beads were purchased from Hamilton (Reno, NV). C18 column packing methods were described previously [2]. For molecule extraction, each EBA sample (~10 mL) was loaded into a column and the liquid was pushed through using a syringe pump. The column was washed with 400 μL of 0.1% formic acid (FA) in water three times for cleaning and desalting. After desalting, two elution steps were performed. First, metabolites were eluted using 400 μL of 50% acetonitrile (ACN) in water. Next, lipids were eluted from the same column using 400 μL of 70% isopropyl alcohol (IPA). Both metabolite and lipid samples were lyophilized overnight. For metabolite analysis with MS, dried samples were resuspended in 0.05% trifluoroacetic acid (TFA) in water. For lipid analysis with MS, dried samples were suspended in 50% ACN. Samples containing MS-grade water were used as negative control samples. For quality control, a mixture of short peptides was used to monitor retention time, mass resolution, and accuracy (Thermo Fisher Scientific).

Nano-liquid chromatography and MS

Metabolite and lipid samples were centrifuged at 10,000 *g* for 10 minutes before being processed using an autosampler in an EASY-nLC 1000 system and characterized using a LTQ orbitrap system (Thermo Fisher Scientific). Chromatograms were generated using an Acclaim PepMap 100 C18 trap column (0.2 mm × 20 mm, 5 μL/min) and an analytical column (75 μm × 150 mm, 300 nL/min). The mobile phase used for metabolites was 80% ACN prepared in 0.1% FA in water from 5% to 70% in 60 minutes. The mobile phase used for lipids was 90% IPA and 10% ACN from 5% to 90% in 60 minutes. For accurate

mass measurement, the resolution for MS was set to 30,000. Data collection was conducted using positive ion mode for both metabolites and lipids. High energy collision-induced disassociation was used for ion fragmentation with 35% total energy.

Identification of metabolites and lipids using database matching

Mass spectra raw data files were converted to .abf format using an open-source software (reifycs.com/AbfConverter). Raw spectra were processed with MS-DIAL software for molecule identification following the standard operation procedures provided by the developers [16]. The database for metabolites was extracted from MassBank, which includes 8,068 records (as of August 9, 2021). The database for lipids was extracted from LipidBlast library, which contains 81 classes, 377,313 molecules, and 554,041 spectra (as of August 9, 2021) [17].

Data analysis and statistics

Metabolites and lipids identified using the software were extracted into Microsoft Excel. Feature alignment was conducted in MATLAB. An ion exclusion list was constructed using negative control samples. Total ion intensity was used for data normalization. For statistical analysis, ion intensity was log-transformed. For each patient visit, relative standard deviation (%) of each identified molecule was calculated and subsequently ranked in the GXP-negative and GXP-positive groups. Molecules with the lowest relative standard deviation values (30% cutoff) were processed with a t-test between GXP-negative and GXP-positive.

Statistically significant features calculated from a t-test from all 3 clinical visits were used for generating receiver operating characteristic (ROC) curves. The area under the ROC curve (A.U.C) was calculated using the logistic regression model of identified molecules. In our previous study, a feature ranking algorithm, significance analysis of microarrays (SAM), was used to select the most powerful feature that separates the two groups [5]. In this study, SAM was applied to statistically significant features in all 3 clinical visits and the scores calculated from SAM analysis were used to determine the significant of each feature.

Declarations

Acknowledgments

Part of this work was funded by National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R01A/147347 and the Strategic Health Innovation Partnerships (SHIP) of the South African Medical Research Council as a sub-grant from the Bill and Melinda Gates Foundation.

Author Contributions

Experiment design: D.C. W.A.B. R.W. M.M. EBA samples: W.A.B. R.B. M.T. T.J.S. M.H. S.G. D.F.W. EBA sample preparation: N.A.B. Reagents and materials: N.A.B. A.P.D. E.R.C. C.R.H. Mass spectrometry: D.C. Data processing: D.C. D.S. Data analysis: D.C. D.S. Manuscript: D.C. All authors understood the results. All authors approved the manuscript.

Competing Interests Statement

W.A.B. M.M. D.S. D.C. N.A.B. A.P.D. E.R.C. C.R.H. have competing financial interests. An unpublished U.S. Provisional Patent Application assigned to Zeteo Tech, Inc was applied based on this research. M.T. T.J.S. M.H. S.G. D.F.W. R.H. do not have competing financial interests.

References

1. Jansson, B.O. and Larsson, B.T., 1969. Analysis of organic compounds in human breath by gas chromatography-mass spectrometry. *The Journal of laboratory and clinical medicine*, *74*(6), pp.961-966.
2. Chen, D., Bryden, W.A. and McLoughlin, M., 2020. A novel system for the comprehensive collection of nonvolatile molecules from human exhaled breath. *Journal of Breath Research*, *15*(1), p.016001.
3. Chen, D., Devin, A.P., Caton, E.R., Bryden, W.A. and McLoughlin, M., 2021. Capture of Viruses and Microorganisms in Aerosols Using A Newly Designed Collection System: A Proof-of-concept Study. *bioRxiv*.
4. Riley, R.L., Mills, C.C., O'grady, F., Sultan, L.U., Wittstadt, F. and Shivpuri, D.N., 1962. Infectiousness of air from a tuberculosis ward: ultraviolet irradiation of infected air: comparative infectiousness of different patients. *American Review of Respiratory Disease*, *85*(4), pp.511-525.
5. Chen, D., Bryden, W.A. and Wood, R., 2020. Detection of tuberculosis by the Analysis of exhaled Breath particles with High-resolution Mass Spectrometry. *Scientific reports*, *10*(1), pp.1-9.
6. Miekisch, W., Schubert, J.K. and Noeldge-Schomburg, G.F., 2004. Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clinica chimica acta*, *347*(1-2), pp.25-39.
7. Schallschmidt, K., Becker, R., Jung, C., Bremser, W., Walles, T., Neudecker, J., Leschber, G., Frese, S. and Nehls, I., 2016. Comparison of volatile organic compounds from lung cancer patients and healthy controls—Challenges and limitations of an observational study. *Journal of breath research*, *10*(4), p.046007.
8. Hu, Q., Noll, R.J., Li, H., Makarov, A., Hardman, M. and Graham Cooks, R., 2005. The Orbitrap: a new mass spectrometer. *Journal of mass spectrometry*, *40*(4), pp.430-443.
9. Lotfi, M., Hamblin, M.R. and Rezaei, N., 2020. COVID-19: Transmission, prevention, and potential therapeutic opportunities. *Clinica chimica acta*, *508*, pp.254-266.
10. Romieu, I., Barraza-Villarreal, A., Escamilla-Nuñez, C., Almstrand, A.C., Diaz-Sanchez, D., Sly, P.D. and Olin, A.C., 2008. Exhaled breath malondialdehyde as a marker of effect of exposure to air pollution in children with asthma. *Journal of Allergy and Clinical Immunology*, *121*(4), pp.903-909.

11. Beck, O., Olin, A.C. and Mirgorodskaya, E., 2016. Potential of mass spectrometry in developing clinical laboratory biomarkers of nonvolatiles in exhaled breath. *Clinical chemistry*, *62*(1), pp.84-91.
12. Greening, N.J., Larsson, P., Ljungström, E., Siddiqui, S. and Olin, A.C., 2021. Small droplet emission in exhaled breath during different breathing manoeuvres: Implications for clinical lung function testing during COVID-19. *Allergy*, *76*(3), pp.915-917.
13. Scriba, T.J., Fiore-Gartland, A., Penn-Nicholson, A., Mulenga, H., Mbandi, S.K., Borate, B., Mendelsohn, S.C., Hadley, K., Hikuam, C., Kaskar, M. and Musvosvi, M., 2021. Biomarker-guided tuberculosis preventive therapy (CORTIS): a randomised controlled trial. *The Lancet Infectious Diseases*, *21*(3), pp.354-365.
14. Wood, R., Morrow, C., Barry III, C.E., Bryden, W.A., Call, C.J., Hickey, A.J., Rodes, C.E., Scriba, T.J., Blackburn, J., Issarow, C. and Mulder, N., 2016. Real-time investigation of tuberculosis transmission: developing the respiratory aerosol sampling chamber (RASC). *PloS one*, *11*(1), p.e0146658.
15. Patterson, B., Morrow, C., Singh, V., Moosa, A., Gqada, M., Woodward, J., Mizrahi, V., Bryden, W., Call, C., Patel, S. and Warner, D., 2017. Detection of Mycobacterium tuberculosis bacilli in bio-aerosols from untreated TB patients. *Gates open research*, *1*.
16. Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O. and Arita, M., 2015. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature methods*, *12*(6), pp.523-526.
17. Kind, T., Liu, K.H., Lee, D.Y., DeFelice, B., Meissen, J.K. and Fiehn, O., 2013. LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nature methods*, *10*(8), pp.755-758.
18. Pei, P.T.S., Henly, R.S. and Ramachandran, S., 1975. New application of high pressure reversed-phase liquid chromatography in lipids. *Lipids*, *10*(3), pp.152-156.
19. He, Y., Rashan, E.H., Linke, V., Shishkova, E., Hebert, A.S., Jochem, A., Westphall, M.S., Pagliarini, D.J., Overmyer, K.A. and Coon, J.J., 2021. Multi-Omic Single-Shot Technology for Integrated Proteome and Lipidome Analysis. *Analytical Chemistry*, *93*(9), pp.4217-4222.
20. Zeka, A.N., Tasbakan, S. and Cavusoglu, C., 2011. Evaluation of the GeneXpert MTB/RIF assay for rapid diagnosis of tuberculosis and detection of rifampin resistance in pulmonary and extrapulmonary specimens. *Journal of clinical microbiology*, *49*(12), pp.4138-4141.
21. Mann, M., Kumar, C., Zeng, W.F. and Strauss, M.T., 2021. Artificial intelligence for proteomics and biomarker discovery. *Cell Systems*, *12*(8), pp.759-770.
22. Frediani, J.K., Jones, D.P., Tukvadze, N., Uppal, K., Sanikidze, E., Kipiani, M., Tran, V.T., Hebbar, G., Walker, D.I., Kempker, R.R. and Kurani, S.S., 2014. Plasma metabolomics in human pulmonary tuberculosis disease: a pilot study. *PloS one*, *9*(10), p.e108854.
23. Feng, S., Du, Y.Q., Zhang, L., Zhang, L., Feng, R.R. and Liu, S.Y., 2015. Analysis of serum metabolic profile by ultra-performance liquid chromatography-mass spectrometry for biomarkers discovery: application in a pilot study to discriminate patients with tuberculosis. *Chinese medical journal*, *128*(2), p.159.

24. Weiner 3rd, J., Parida, S.K., Maertzdorf, J., Black, G.F., Repsilber, D., Telaar, A., Mohney, R.P., Arndt-Sullivan, C., Ganoza, C.A., Fae, K.C. and Walzl, G., 2012. Biomarkers of inflammation, immunosuppression and stress are revealed by metabolomic profiling of tuberculosis patients. *PloS one*, *7*(7), p.e40221.

Tables

Table 1 is not available with this version

Figures

Figure 1

Figure 1. CORTIS participant information and workflow for identification of NOCs from EBA samples. Numbers of CORTIS study participants who tested negative (GXP-negative) or positive (GXP-positive) with GeneXpert MTB/RIF are shown. EBA samples were collected as previously described [5]. A dual-mobile phase extraction method, utilizing 50% ACN and then 70% IPA, enabled extraction of small metabolites and lipids from the EBA samples. Extracted molecules were processed with LC-MS/MS and identified via database matching. Relative standard deviation (RSD) of each identified molecules was calculated in each group. Molecules with <30% RSD were used for t-test between GXP negative and positive of each visit. Molecules showing statistical differences in all visits were used to generate ROC curve and AUC calculation. A feature ranking algorithm was used to show the important of identified molecules.

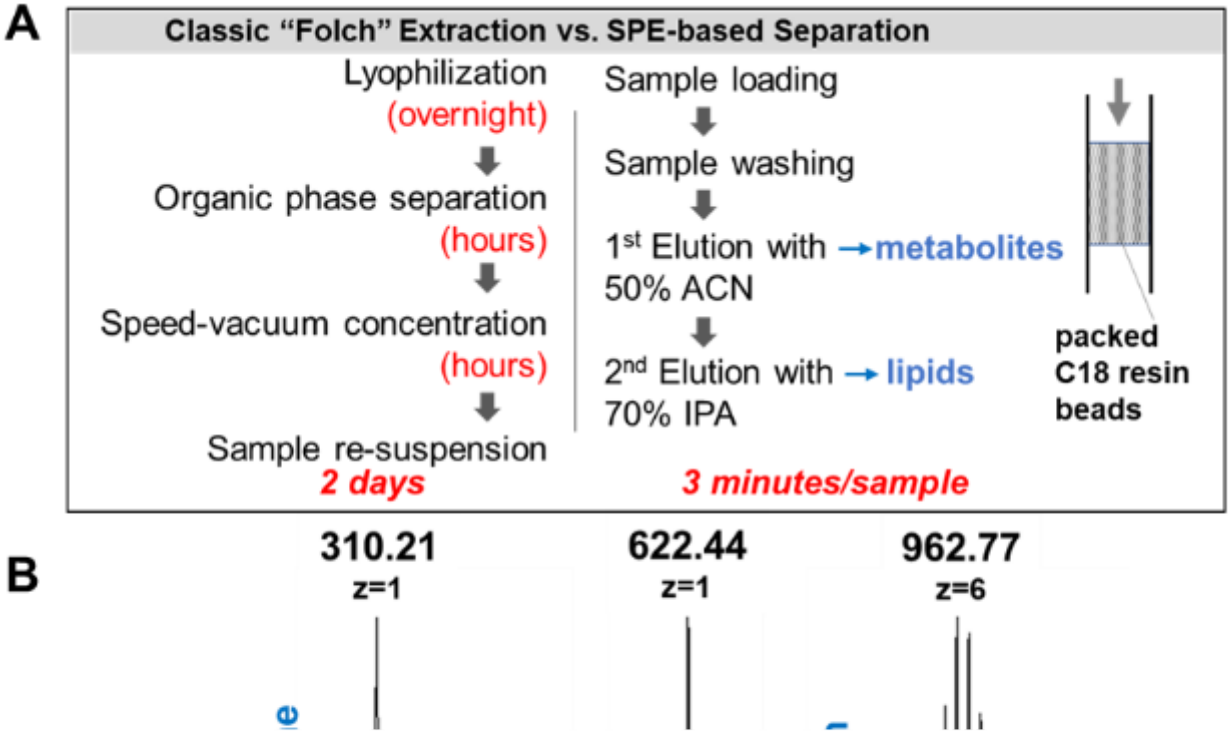

Figure 2

Figure 2: Method development for the separation of small molecules and lipids from one sample using duel mobile phase solvents. (A) Comparison between the classic Folch method and the solid-phase extraction (SPE)-based method developed in this work which incorporates dual-mobile phase extraction. (B) Mass spectra of three representative molecules used to validate the dual-mobile phase extraction method. (C) Recorded ion intensities of three representative molecules in either washing or eluting solutions.

Figure 3

Figure 3. Overview of molecular profiles of EBA samples with liquid chromatography and high-resolution MS. (A) Total ion chromatograms of control samples and EBA samples. (B) Representative total ion chromatograms of small metabolites from GXP-positive and GXP-negative participants. (C) Feature distributions of small metabolites from three visits of GXP-positive and GXP-negative participants. (D) Feature distributions of lipids in three visits of GXP-positive and GXP-negative participants. (E) Dynamic range of small metabolites detected across all study participants. (F) Dynamic range of lipids detected across all study participants.

Figure 4

Figure 4. Overview of study participant correlation and individual feature deviations. (A) Global correlation heatmap of study participants. Distribution of relative standard deviation values of metabolites and lipids in (B-C) GXP-negative and (D-E) GXP-positive study participants.

Figure 5

Figure 5. Identification of definitive features in GXP-negative and GXP-positive groups. (A) Numbers of metabolites and lipids identified in each study group at each visit. (B) The five most identified metabolites and lipids identified in (B) GXP-negative and (C) GXP-positive study participants. Blue, green, and orange circles indicate the numbers of identified molecules at each visit.

Figure 6

Figure 6. Statistical differences in molecular profiles for GXP-negative and GXP-positive groups at each visit. (A) Correlation scattering of metabolites and lipids of GXP-positive/GXP-negative fold changes in pairwise combinations of three different visits. Volcano plots of GXP-positive/GXP-negative fold change and t-test p values for (B) metabolites and (C) lipids at each visit. Red dots indicate that the identified molecules were higher in the GXP-positive group while green dots indicate lower expression.

Figure 7

Figure 7. Significance analysis of metabolites and lipids differentiating GXP-negative and GXP-positive groups. Standard error distribution of (A) metabolites and (B) lipids that were significantly different in GXP-negative and GXP-positive groups across all three visits. Red crosses indicate the mean values. (C) AUC values for logistic regression models using identified molecules. (D) Correlation heatmap of metabolites and lipids based on Pearson correlation coefficient analysis. (E) Feature ranking analysis of metabolites and lipids based on significance analysis of microarrays (SAM) algorithm. Red dots indicate increased and green dots indicate decreasing in GXP-positive. G6P: d-glucosamine-6-phosphate; AMP: adenosine monophosphate; BA: butanoic acid; PA: palmitic acid; NAM: n-acetylmuramic acid; PHE: phenylalanine; HA: hyocholic acid; GLN: glutamine; iPEN: isopentenyladenine; PS: phosphatidylserine; Cer: ceramides; PI: phosphatidylinositol; DG: diacylglycerol; PC: phosphatidylcholine; TG: triglycerides.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SuppleTable01.xlsx