The effect of sex and HIV on the TB-induced altered sputum metabolome

D Beukes

orcid.org 0000-0003-1987-4480

Dissertation accepted in partial fulfilment of the requirements for the degree *Master of Science in Biochemistry* at the North-West University

Supervisor:	Dr I du Preez
Co-supervisor :	Prof DT Loots
Co-supervisor:	Dr M van Wyk

Graduation July 2023

13128531

Acknowledgements

תהילה לאל במרומים

Glory be to God in the highest

2 Samuel 22 v 7-10 & 17-20

"IN MY DISTRESS I CALLED TO THE LORD;

I CALLED OUT TO MY GOD.

FROM HIS TEMPLE HE HEARD MY VOICE;

MY CRY CAME TO HIS EARS.

THE EARTH TREMBLED AND QUAKED,

THE FOUNDATIONS OF THE HEAVENS SHOOK;

THEY TREMBLED BECAUSE HE WAS ANGRY.

SMOKE ROSE FROM HIS NOSTRILS;

CONSUMING FIRE CAME FROM HIS MOUTH,

BURNING COALS BLAZED OUT OF IT.

HE PARTED THE HEAVENS AND CAME DOWN;

DARK CLOUDS WERE UNDER HIS FEET.

"**HE REACHED DOWN FROM ON HIGH AND TOOK HOLD OF ME**;

HE DREW ME OUT OF DEEP WATERS.

HE RESCUED ME FROM MY POWERFUL ENEMY,

FROM MY FOES, WHO WERE TOO STRONG FOR ME.

THEY CONFRONTED ME IN THE DAY OF MY DISASTER,

BUT THE LORD WAS MY SUPPORT.

HE BROUGHT ME OUT INTO A SPACIOUS PLACE;

HE RESCUED ME BECAUSE HE DELIGHTED IN ME.

DADDY, MOMMY, MATTEO, MILAN, LARA, LEEANUS, DILLION, UNCLE RONNIE, AUNTY MIONE and all my closest friends.

My study leader, Dr Ilse du Preez.

Thank you for all the motivation, continued support and understanding. Your prayers carried me through.

Summary

Various studies have identified TB-induced metabolome variations. However, in most of these studies, the unaccounted-for variation between individual patients is noteworthy. To investigate one potential cause of this variation, we evaluated the effect of sex on the sputum metabolomes of TB and TB/HIV co-infected individuals. Untargeted GCxGC/TOF-MS analyses were applied to the sputum of 31 TB+ and 197 TB- individuals. Univariate statistical methods were used to identify metabolites that differ significantly between TB+ and TB- individuals within each sex (male and female) 1) irrespective of HIV status and 2) for HIV+ patients.

Ignoring HIV status, 21 compounds were significantly different between the TB+ and TB- individuals within the female subgroup (11% lipids; 10% carbohydrates; 1% amino acids, 5% other and 73% unannotated), and six within the male subgroup (20% lipids; 40% carbohydrates; 6% amino acids, 7% other and 27% unannotated). For the HIV+ patients (TB+ vs TB-), a total of 125 compounds were significant within the female subgroup (16% lipids; 8% carbohydrates; 12% amino acids, 6% organic acids, 8% other and 50% unannotated), and 44 within the male subgroup (17% lipids; 2% carbohydrates; 14% amino acids related, 8% organic acids, 9% other and 50% unannotated). One annotated compound, 1-oleoyl lysophosphatidic acid, was consistently identified as a TB marker, irrespective of sex or HIV status.

In this cohort, sex played a definite role in how TB and TB/HIV infection perturbs the metabolome. Our findings, therefore, highlight the need for better stratified cohorts in metabolomics studies.

Keywords: *tuberculosis; sputum; HIV/TB co-infection; GCxGC-TOFMS; metabolomics*

Please note that large parts of this dissertation, including the summary, was accepted for publication as a scientific paper in Metabolomics journal (impact factor = 4.7).

Table of contents

List of abbreviations

Abbreviation	Meaning	Abbreviation	Meaning
BSTFA	N,O-bis(trimethylsilyl)trifluoroacetamide	**GC-MS**	Gas chromatography mass spectrometry
CE	Capillary electrophoresis	**GCxGC-TOFMS**	Two-dimensional gas chromatography time-of-flight mass spectrometry
CF	Cystic fibrosis	**HIV**	Human immunodeficiency virus
COPD	Chronic obstructive pulmonary disease	**HREC**	Human Research Ethics Committee
CT	Computerized tomography	**HRMAS**	High-resolution magic-angle sample spinning
CV	Coefficient of variation	**Hz**	Hertz
Da	Dalton	**IL-4**	Interleukin-4
DESI	Desorption electrospray ionization	**IL-5**	Interleukin-5
ECDC	European Centre for Disease Prevention and Control	**IL-13**	Interleukin-13
ELISA	Enzyme-linked immunosorbent assay	**IRIS**	Immune reconstitution inflammatory syndrome
ES	Effect size	**LC**	Liquid chromatography
ESI	Electrospray ionization	**LC-MS**	Liquid chromatography mass spectrometry
ETC	Electron transport chain	**LC-MS/MS**	Liquid chromatography-tandem mass spectrometry
eV	Electron volts	**LPA**	Lysophosphatidic acid
FAME	fatty acyl methyl ester	**LPA-1**	Lysophosphatidic acid receptor 1
FC	Fold change	**LPC**	Lysophosphatidylcholine
FEV_1	Forced expiratory volume in 1 second	**m**	meter
FIE-MS	Flow Infusion Electrospray FIE-MS	**MADD**	Multiple acyl-CoA dehydrogenase deficiency
FTIR	Fourier transform infrared spectroscopy	**MALDI-TOF**	Matrix-assisted laser desorption/ionization-time of flight
FVC	Forced vital capacity	**mg**	Milligram
g	Gravitational force	**min**	minutes
GA type II	Glutaric aciduria type II	**mL**	**Millilitre**
GC	Gas chromatography	**mm**	**Millimetre**

Abbreviation	Meaning	Abbreviation	Meaning
MRS	Magnetic Resonance Spectroscopy	**QC**	Quality control
MSEA	Metabolite set enrichment analysis	**ROC-AUC**	Receiver operating characteristic-Area under curve
Mtb	*Mycobacterium tuberculosis*	**s**	seconds
m/z	Mass-to-charge-ratio	**SCFA**	Short chain fatty acid
n	Denotes counting numbers	**SCLC**	Small Cell Lung Cancer
NSCLC	Non-small cell lung cancer	**TB**	Tuberculosis
NTM	Nontuberculous mycobacterial	**Th2**	T helper 2
NMP	National Metabolomics Platform	**TIC**	Total ion chromatogram
NMR	Nuclear magnetic resonance	**MS/MS**	Tandem mass spectrometry
NWU	North-West University	**TMCS**	Trimethylsilyl chloride
OPLS-DA	Orthogonal projections to latent structure-discriminant analysis	**uL**	Micro-litre
OS	Oxidative stress	**um**	Micro-meter
PUFA	Polyunsaturated fatty acids	**UHPLC-MS/MS**	Ultra-high-performance liquid chromatography-tandem mass spectrometry
POC	Point of care	**UPLC-MS**	Ultra-high-performance liquid chromatography-mass spectrometry
TOF-MS	Time-of-flight-mass-spectrometry	**UPLC-qTOF**	Ultra-high performance liquid chromatography quadrupole time-of-flight mass spectrometry
PAF	Platelet activating factor	**UPLC-TQMS**	Ultra-pressure liquid chromatography triple quadrupole mass spectrometry.
PEx	Pulmonary exacerbations	**V**	Volts
PCA	Principal component analysis	**WHO**	World Health Organization
PC	Phosphocholine	**2D**	Two-dimensional
PLA_2	Phospholipase A_2	**4D**	Four-dimensional
PLS-DA	Partial least squares-discriminant analysis	**°C**	Degree Celsius

Chapter 1: Preface

1.1 Background, motivation and problem statement

Tuberculosis (TB), caused by *Mycobacterium tuberculosis (M. tuberculosis)*, still poses a threat, not only to health, but also to socio-economics, particularly in low and middle income countries. New TB diagnoses are estimated at 10 million, with 1.2 million deaths among human immunodeficiency virus (HIV)-negative TB patients and 208 000 deaths in HIV and TB co-infected individuals (WHO, 2020). Moreover, a difference in TB prevalence exists between males and females, with males having a 1.6 times higher prevalence than females (World Health Organization, 2016).

In the fight against TB, various research groups are applying the increasingly popular research approach known as metabolomics to investigate various aspects of the diseases, ranging from basic biology to the identification of disease biomarkers and improved treatment regimens (Du Preez et al., 2019). The cohorts used in these studies are, however, complex, including patients with large individual variation (age, sex, etc.) and an array of environmental factors such as comorbidities, eating habits, smoking, etc. (Koen et al., 2016). Previous reports indicate that many of these internal and external factors can perturb the human metabolome, even in participants considered to be healthy (Darst et al., 2019, Li et al., 2018, Krumsiek et al., 2015).

We therefore hypothesize that genetic factors and comorbidities, such as sex and HIV status, could have a concomitant effect on the host metabolome variations induced by active TB, but it is still unclear on what specifically the impact might turn out to be. Although the collection of sputum can be considered invasive, especially in HIV co-infected individuals, it originates directly from the source of infection, and therefore reflects variations in the host metabolome due to the active disease state in addition to the bacterial infection (Du Preez et al., 2019).

Sputum is a viscous matrix complicating the analysis thereof. Our research group has developed a procedure specifically focused on the GCxGC-TOFMS analysis of sputum samples which, in summary, includes a homogenization step before performing the extraction (Schoeman et al., 2012). When applying this technique to TB patient samples, the team was able to identify several altered metabolites characterising changes induced by active TB in the sputum metabolome, better describing the disease and also establishing several new hypotheses (Du Preez & Loots, 2013). Many samples

were, however, considered outliers, and excluded from the final datasets. The reason for the divergent metabolite profiles was not apparent, highlighting the need for the proposed sub-cohort investigation in this study. More specific or personalised TB metabolome profiles can lead to a better understanding of the disease which could ultimately aid in the development of personalised TB diagnostic and treatment approaches. The goal when studying these obtained TB profiles is the discovery of unique disease biomarkers, irrespective of sex and HIV status, that can be applied for developing rapid point of care tests, thereby accelerating diagnosis and treatment.

1.2 Aim and objectives

1.2.1 Aim

To determine the influence of sex (genetic) and HIV-status (comorbidity) on the sputum metabolome of pulmonary TB patients.

1.2.2 Objectives

1. To evaluate method performance of a previously developed total metabolome extraction method and untargeted two-dimensional gas chromatography time-of-flight mass spectrometry (GCxGC-TOFMS) analysis of sputum samples.
2. To analyse all samples (TB+ and TB-) using the method from objective 1, and to apply this data to determine the metabolic variation when comparing the sample groups (TB+ and TB-), irrespective of sub-cohort classifications.
3. To determine the effect of sex and HIV-status on the TB-induced metabolome change identified in objective 2, using univariate and multivariate statistical analyses.

1.3 Structure of the dissertation and research outputs

This dissertation complies with the requirements of the North-West University (NWU) for the completion of an MSc degree (Biochemistry) in article format. For this reason, each chapter has an introduction, body, and reference list. **Chapter 1** gives a general introduction to the study by describing the problem statement, aims and objectives. A summary of the content of each chapter and the contribution of co-authors and co-workers is also provided here. **Chapter 2** provides an

overview of the literature required to contextualise the subsequent chapters. This chapter was written in the format of a mini review paper, focussing on the importance and relevance of sputum as a sample matrix in metabolomics research, for future publication purposes. **Chapter 3** describes the overall study design and the materials and methods used in greater detail. Parts of **Chapter 3** was submitted for publication as an original paper and presented as a poster at an international conference:

- Beukes, D., Van Reenen, M., Loots DT., Du Preez, I. The effect of sex on the sputum metabolome of TB and TB/HIV co-infected individuals: An untargeted GCxGC-TOFMS study. Accepted for publication in *Metabolomics* - 10 May 2023.

- Beukes, D., Van Reenen, M., Loots DT., Du Preez, I. The effect of sex on the sputum metabolome of TB and TB/HIV co-infected individuals. Presented as poster at the 18th Annual Conference of the Metabolomics Society: Metabolomics 2022: Valencia, Spain, 19-23 June 2022.

Chapter 4 reports the results of the study and provides a thorough discussion of the findings. The final chapter, **Chapter 5**, provides a summary of the main outcomes and conclusions drawn from the study, culminating in considerations for future applications.

1.4 Author contributions

The contributions of the main author of this dissertation (Ms Derylize Beukes), co-authors and co-workers are given in **Table 1.1**.

The following statement from the main author, supervisor and co-supervisors confirms their roles in the study and gives their permission that the data and conclusions drawn from this study may be used as part of this dissertation: *I declare that my role in this study, as indicated in **Table 1.1**, is a representation of my actual contribution, and I hereby give my consent that this work may be published as part of the M.Sc. dissertation of Derylize Beukes.*

 *Supervisor* **Dr I Du Preez**	 *Co-Supervisor* **Prof DT Loots**	 *Co-Supervisor* **Dr M van Wyk** (Publishing under van Reenen)	 *MSc candidate* **D Beukes**

***Table 1.1**: Researchers that contributed and were involved in this study.*

Co-author/Co-worker	Contribution
Derylize Beukes **(B.Sc. Hons. Biochemistry)** ***MSc Candidate***	**Conceptualised and designed the study together with the supervisor and co-supervisors. Performed the preparation of sputum samples for GCxGC-TOFMS analyses. Performed GCxGC-TOFMS analysis, including processing, and quantifying the data. Performed statistical analysis, drafted and compiled the original drafts of the dissertation and all associated documents, including the scientific publication and poster.**
Dr Ilse Du Preez **(Ph.D. Biochemistry)** ***Supervisor***	**Conceptualised, coordinated and supervised all aspects of the study, including study design, planning, execution, and writing of the dissertation and all associated documents, including the scientific publications and poster**
Prof Du Toit Loots **(Ph.D. Biochemistry)** ***Co-supervisor***	**Critically evaluated and advised on the study design, and execution of the study, including the compilation of the dissertation and scientific publication. Managed the sample collection and ethical aspects.**
Dr Mari van Reenen	**Critically evaluated and advised on the study design and execution of the study, including the compilation**

(Ph.D. Statistics) ***Co-supervisor***	**of the dissertation and scientific publication. Supervised statistical analyses and result interpretation.**

1.5 References

Darst, B.F., Koscik, R.L, Hogan, K.J., Johnson, S.C., Engelman, C.D. 2019. Longitudinal plasma metabolomics of aging and sex. *Aging*, 11(4):1262-1282.

Du Preez, I. and Loots, D.T. 2013. New sputum metabolite markers implicating adaptations of the host to Mycobacterium tuberculosis, and vice versa. *Tuberculosis*, 93(3): 330–337.

Du Preez, I., Luies, L., Loots, DT. 2019. The application of metabolomics toward pulmonary tuberculosis research. *Tuberculosis*, 115:126-139.

Koen, N., Du Preez, I., Loots, du T. 2016. Metabolomics and Personalized Medicine. *Adv Protein Chem Struct Biol*. 102:53-78.

Krumsiek, J., Mittelstrass, K., Do, K.T., Stückler, F., Ried, J., Adamski, J., *et al.* 2015. Gender-specific pathway differences in the human serum metabolome. *Metabolomics*, 11(6):1815-1833.

Li, Z., Zhang, Y., Hu, T., Likhodii, S., Sun, G., Zhai, G., *et al.* 2018. Differential metabolomics analysis allows characterization of diversity of metabolite networks between males and females. *PLoSONE*, 13(11):e0207775.

Schoeman, J.C., Du Preez, I., Loots, D.T. 2012. A comparison of four sputum pre-extraction preparation methods for identifying and characterising Mycobacterium tuberculosis using GCxGC-TOFMS metabolomics. *J Microbiol Methods*, 91:301–311.

World Health Organization. 2016. Global Tuberculosis Report 2016. https://apps.who.int/iris/bitstream/handle/10665/250441/9789241565394-eng.pdf?sequence=1 Accessed 27.07.2021

World Health Organization. 2020. Global tuberculosis report. https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf Accessed 15 June 2021.

Chapter 2: Literature review

2.1 Introduction

TB is recognised as one of the leading causes of morbidity and mortality globally. In 2019, an estimated 10.0 million people fell ill with active TB. Men (aged ≥15 years) accounted for 56% of individuals who developed TB in 2019, whereas women (aged ≥15 years) accounted for 32% and children (aged <15 years) for 12% (World Health Organization, 2020). Various factors are hypothesized to explain the sex gap, including biological differences in the disease presentation and distinct access to health care (Ingersoll, 2017; World Health Organization, 2005). In addition, it is well known that women and men react differently to stress and have divergent coping abilities (Bassuk & Manson, 2015; Leening *et al.*, 2014). Moreover, of all active TB cases recorded globally in 2019, 8.2% were among people living with HIV (World Health Organization, 2020). Since as early as the 1980's, HIV infection has been considered one of the most important predisposing factors for developing active TB from a latent infection (Zumla *et al.*, 2000). Inversely, TB is one of the leading causes of death among people living with HIV (Center of Disease Control, 2016). The goal of this MSc was to determine if sex and HIV status have prominent effects on the sputum metabolome profiles of pulmonary TB patients, to ultimately aid in the development of personalised diagnostic and therapeutic approaches. In accordance, this chapter provides a comprehensive background on metabolomics as a research field and the analytical apparatus typically used in these investigations; the application of sputum as a sample matrix in metabolomics studies; and the known metabolome variations related to sex and HIV status in healthy individuals.

2.2 Metabolomics

Metabolomics is a research approach involving the quantitative measurement of metabolic changes in a living system in response to a genetic variation or physiological stimuli (Dunn *et al.*, 2005; Collino *et al.*, 2013). These metabolic changes are measured via the identification, quantification, and characterisation of the metabolome, the latter of which includes all small molecules (< 1500 Da) such as metabolic intermediates, hormones, and other signalling molecules, as well as secondary metabolites found in a specific cell, organ, organism or biofluid. Metabolomics can, therefore, be used to characterise a specific metabolic phenotype related to an external or internal perturbation, such

as, age, sex, or a disease state (Wishart *et al.*, 2007). The recent application of different metabolomics methodologies has led to the identification of metabolites/metabolite profiles which better characterise a variety of phenotypes, thereby envisaging biomedical applications such as diagnostics, monitoring of disease progression and treatment efficiency, and the development of personalised treatment plans (Gonzales-Covarrubias *et al.*, 2022).

Metabolomics can be approached in an untargeted or targeted manner. Per definition, untargeted metabolomics is the comprehensive analysis of all measurable analyte features in a sample. In this case, features are typically annotated based on the similarity of their spectra with those in spectral libraries and quantified relative to an internal standard. In contrast, targeted metabolomics is the measurement of predefined groups of chemically characterised metabolites (Ribbenstedt, Ziarrusta & Benskin, 2018). Annotation is based on the similarity to spectra obtained from the analyses of compound standards and may or may not include retention time matching. Absolute quantification is applied in targeted analysis by including internal standards (isotopes) and/or calibration curves for each compound. The untargeted approach is widely used for the potential discovery of new metabolite variations and is therefore regarded as an exploratory or "hypothesis generating" approach. In contrast, targeted analyses are applied when specific compounds, compound class or metabolic pathways are of interest, often based on preceding untargeted results, hypotheses generated from literature searches or known disease-induced metabolic variations (i.e. diagnostics of inborn metabolic disorders). Both targeted and untargeted metabolomics have been successfully implemented in the past few years to better characterise numerous aspects of TB using an array of sample matrices, analytical apparatuses and statistical methodologies (Du Preez *et al.*, 2019).

2.2.1 Analytical techniques used in metabolomics

A range of analytical instruments can be used for metabolomics applications. The most widely used techniques are chromatography-based, including gas (GC) and liquid chromatography (LC), which are usually coupled to a mass spectrometer (MS). Capillary electrophoresis (CE) and nuclear magnetic resonance (NMR) are also popular (Kaddurah-Daouk *et al.*, 2008). Each of these analytical platforms comes with a set of advantages and disadvantages, however, the preferred platform depends mainly on the application (untargeted or targeted) and the sample matrix (Emwas *et al.*, 2015).

Comprehensive GC- and LC- untargeted metabolomics analysis is performed using mostly TOF-based mass analyzers due to their ability to acquire full mass range spectra without sacrificing speed or sensitivity which makes it an excellent choice for qualitative and quantitative analyses across a wide dynamic range to then better characterize complex mixtures (Lei *et al.*, 2011). When focusing on the different analytical platforms available for both targeted and untargeted approaches, **Table 2.1** compares the performance of NMR and GC or LC-MS-based applications for metabolomics.

***Table 2.1**: Summary of the advantages and disadvantages of mass spectrometry (MS) vs nuclear magnetic resonance (NMR) spectroscopy for metabolomics applications (Adapted from Emwas et al. 2019)*

Performance parameter	**Nuclear magnetic resonance (NMR)**	**Mass Spectrometry (MS) (Either GC-MS or LC-MS)**
Reproducibility	High reproducibility	Less reproducible
Sensitivity	Intrinsically low but can be improved with multiple scans (time), higher magnet field strength, cryo-cooled and microprobes, and hyperpolarization methods	High sensitivity: metabolites with nanomolar concentrations can be readily detected
Selectivity	Mostly applied for nonselective analysis. Peak overlaps from multiple detected metabolites pose major challenges	Highly selective
Sample measurement	Measures all metabolites at a detectable concentration in one run	Different ionization methods are required to maximize the number of detected metabolites
Sample preparation	Minimal sample preparation	More extensive sample preparation. Requires sample derivatization for in GC analyses
Sample recovery	Non-destructive - the sample can be recovered and analysed again with a different method	Destructive; sample cannot be recovered
Quantitative analysis	NMR is inherently quantitative as the signal intensity is directly proportional to the metabolite concentrations and number of nuclei in the molecule	The intensity of the MS line is often not correlated with metabolite concentrations as the ionization efficiency is also a determining factor.
Fluxomics analysis	Permits both in vitro and in vivo metabolic flux analysis. Inherently quantitative so enables precise quantification of precursors and products. Mapping of stable isotope locations and incorporating points in molecules is very easy	Can be used for fluxomics analysis; however, the destructive nature means it is somewhat more limited than NMR-based fluxomics. In vivo fluxomics is not possible with MS, and isotope mapping is more difficult.

Tissue samples	Can detect metabolites in tissue samples using high-resolution magic-angle sample spinning (HRMAS) NMR	Although some MALDI-TOF approaches can be used to detect metabolites in tissue samples, these approaches are still far from being routine.
Number of detectable metabolites	Depending on spectral resolution, usually less than 200 metabolites can be unambiguously detected and identified in one measurement	Using different MS techniques, it is possible to detect thousands of different metabolites and identify several hundred
Targeted analysis	Can be used for both targeted and untargeted analysis, but it is not commonly used for targeted analyses.	Can be used for targeted and untargeted analysis, depending on the mass analyser (e.g. TOF or MS/MS)
In vivo studies	Using magnetic resonance spectroscopy (MRS), in vivo investigation can be carried out most often using nuclei such as ^{1}H and ^{31}P.	Although desorption electrospray ionization (DESI) may be a useful way to analyse tissue samples during surgery, MS is not used for in vivo metabolomics studies.

The minimal and non-destructive nature of sample preparation, the capacity to quantify metabolites, and the superior level of reproducibility, makes NMR the preferred platform for long-term or large-scale clinical metabolomics research projects. The advantages listed are, however, often outbidded due to the superior detection limits of LC-MS and GC-MS (Emwas *et al.*, 2019). For targeted analyses and routine diagnostic applications quadrupole (single and triple) mass analysers are typically used. For comprehensive untargeted GC- and LC- metabolomics analysis, time-of-flight (TOF)-based mass analysers are preferred due to their ability to acquire full mass range spectra without sacrificing speed or sensitivity.

In this MSc, sputum, collected directly from the site of infection, the lungs, was the sample matrix of choice. Despite sputum being used as a matrix for TB diagnostics, it is considered a complex sample type, requiring extensive processing and homogenisation prior to metabolomics analysis (Schoeman *et al.*, 2012). For this reason, we chose to analyse these extracts on the highly advanced, comprehensive two-dimensional (2D) GCxGC-TOFMS. GCxGC has become a revolutionary tool capable of high-resolution separation of metabolites in complex mixtures (Dimandja, 2004; Eiserbeck *et al.*, 2011, 2012; Merritt and Hayes, 1994; Ventura *et al.*, 2007). GCxGC established the visualisation of 2D analytical separation anticipated some 25 years ago (Giddings, 1984). The resolution power of GCxGC is achieved by linking two GC columns, consisting of different stationary phases (e.g., nonpolar and polar), via a cryogenic dual-stage modulator to concentrate the fraction compounds eluting from the first GC column within a certain modulation period, to a fine band. This fraction is then transferred into a second GC column where it is separated based on varying chromatographic selectivity (Adahchour *et al.*, 2008). The separation technique is regarded as comprehensive, since all the

fractions collected from the first column are transferred to a second separation dimension (**Figure 2.1**). This technique differs from other multidimensional techniques, for example, heart-cut (e.g., Sciarrone *et al.*, 2010), which only transfers specific fractions of the sample for further separation. The addition of the second dimension enhances the peak capacity of a chromatogram by as much as 20-fold as all compounds in a sample can be distributed within an area spanned by the first and second dimension, as opposed to a single dimension. Because of its fast acquisition rate capabilities, TOFMS is ideal for delineation of narrow second dimension peaks.

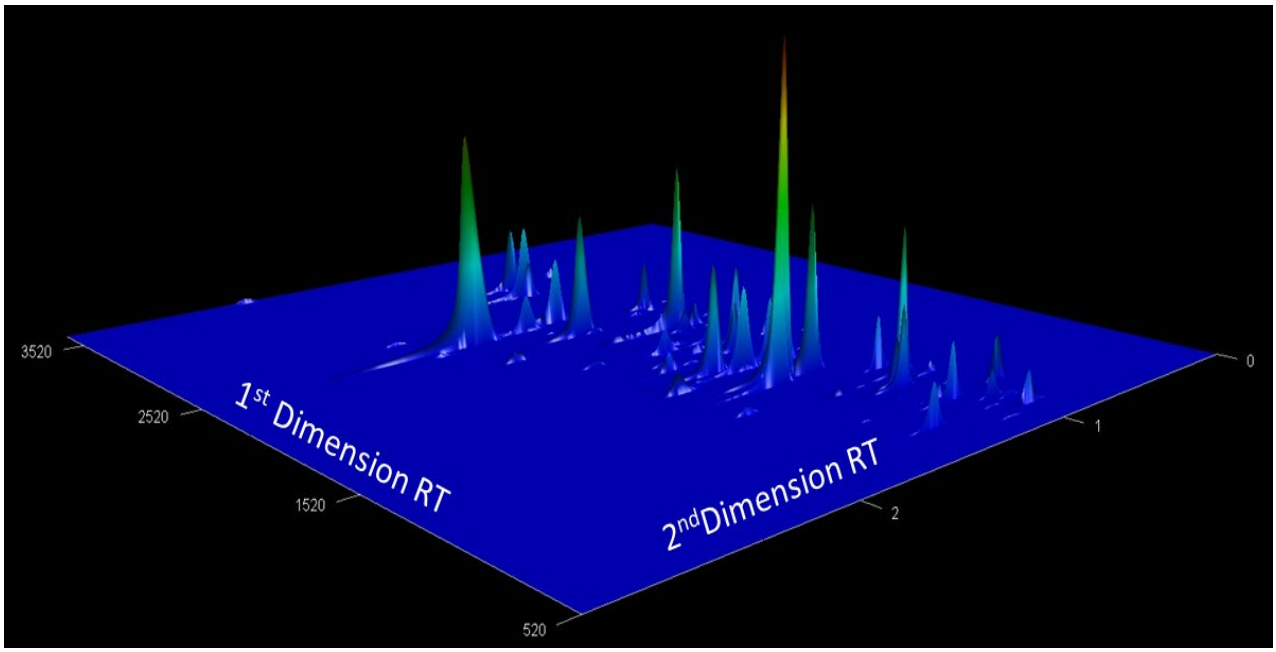


***Figure 2.1**: A two-dimensional total ion chromatogram (TIC) representation example which exhibits the two retention times obtained for the chromatogram of each compound.*

The comprehensive, untargeted analysis of sputum can lead to the detection and identification of compounds from numerous compound classes, representing most biologically relevant metabolic pathways. Various multivariate and univariate statistical techniques can then be employed to identify variation between experimental groups, such as TB+ and TB- individuals, leading to a better understanding of the underlying mechanisms of the perturbation being investigated.

2.3 Sputum as a research and clinical specimen

Sputum, located in the lungs in increased quantities in common lung-related disease states, such as cystic fibrosis (CF), bronchial asthma, lung cancer and TB, consists of two major components, one being the thicker mucous material (distinctly different to saliva), which consists of mucus, whole cells, and cellular materials. The second component is serous fluid, which does not contain cells and is heavily contaminated with saliva. Sputum is produced by the cells that line the airway and is normally secreted into the lower airways of the respiratory tract, working with cilia to trap and remove foreign substances or pathogens (Voynow & Rubin, 2009).

Some of the limitations and important aspects to be considered during sputum collection include:

- Collection variability: The quality and quantity of sputum collected can vary widely depending on the individual and the method of collection. This can affect the accuracy and reproducibility of research results.
- Contamination: Sputum samples can easily be contaminated with saliva, which can affect the accuracy of the results. It is important to ensure that sputum is collected properly to minimize contamination.
- Inability to collect: Some individuals may not be able to produce sputum or may find it difficult to expectorate. This can limit the usefulness of sputum as a diagnostic or research specimen.
- Limited utility for certain pathogens: Sputum may not be the best specimen for detecting certain respiratory pathogens, such as viruses or *Mycobacterium tuberculosis*, which may require other types of specimens like nasopharyngeal swabs or bronchoalveolar lavage.

All the limitations mentioned above was considered and as sample collection was not performed by our group, it could not be intricately controlled by us per se, but the general pathology lab collection procedure for standard TB testing was followed. All subjects included in this study visited a clinic to specifically be tested for some undiagnosed lung related disease, thus sputum production was likely not a problem. The study purposefully wanted to collect sputum, since this is the current chosen matrix for testing for TB disease in South African clinics, hence we didn't consider additional sample matrices such as nasopharyngeal swabs or bronchoalveolar lavage. Sputum smear microscopy could still be a helpful screening tool in cases where very little sputum is collected, or determination of the specific strain is the most important, but culture confirmation is still a requirement. In a previous study, it was found that the yield of sputum is higher in patients with chronic obstructive pulmonary

disease (COPD) and asthma compared to healthy controls (Fahy *et al.*, 2001). The increased quantities of sputum (sputum hyperproduction) in various lung disease states could in part be attributed to extensive lung damage, which further adds to additional mucus build-up in the lungs. Even though the exact mechanism for mucus hyperproduction in TB has not yet been elaborated, Cohn *et al.* (2002) were able to connect this occurrence to the mechanism of T helper 2 (Th2) lymphocytes in the absence of interleukin-4 (IL-4), IL-5, eosinophils, and mast cells but in conjunction with IL-4 receptor alpha signalling. The study established that Th2 cells in the absence of IL-13 were not able to encourage mucus production, even where airway inflammation is present. IL-13 was shown to function directly on goblet cells (specialised epithelial cells responsible for producing mucus), providing an explanation as to why inflammation of the airways results in mucus hyperproduction in asthma disease states. Various other studies detected increased activity in the abovementioned cytokines (Lienhart *et al.*, 2002; Harris *et al.*, 2007), implying that the proposed mucus hyperproduction mechanism in asthma disease might be similar in TB (Luies and Du Preez 2020).

Typical sputum processing involves the separation of sputum from serous fluid, followed by the analysis of cell viability, and measurement of total cell counts. The presence of inflammatory mediators and cells in sputum can be determined by means of microscopy or other techniques.

Even though induced sputum acquisition has been performed since the early 1950's, the procedure has only been standardised in the past two decades in an effort to reduce variability and adverse effects in these individuals. **Figure 2.2** (below) illustrates a typical protocol for the induction of sputum (adapted from Lacy *et al.*, 2005). In short, it entails the issuing of nebulized hypertonic saline at selected increasing doses, which is then to be inhaled by the patient.

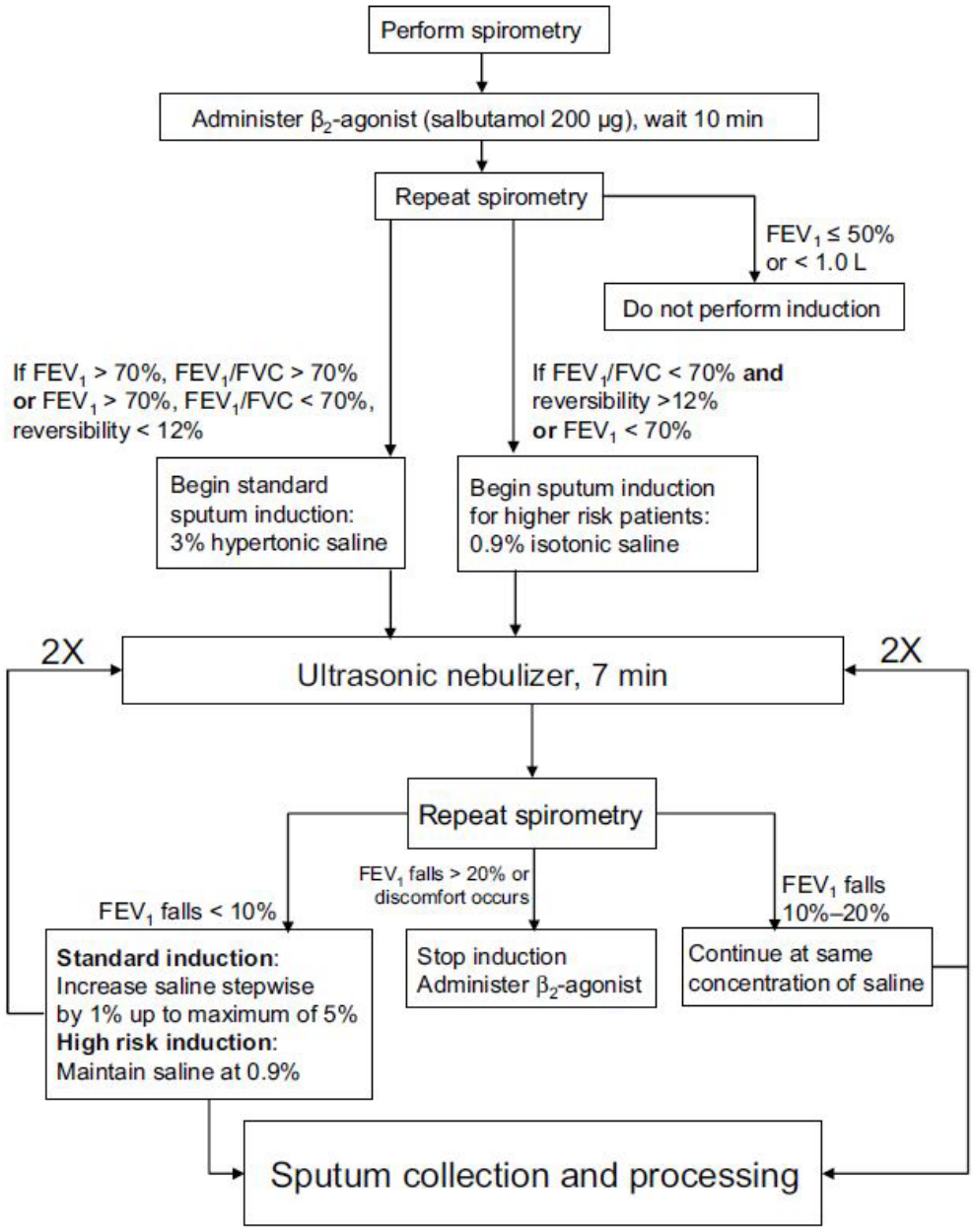


Figure 2.2*: Example of a procedure followed for sputum induction as done for sample collection and consequent testing (adapted from Lacy et al, 2005). Abbreviations: Forced expiratory volume in 1 second (FEV_1) Forced vital capacity(FVC)*

The standardised induction procedure promoted the field of sputum research and consequently elaborated on our understanding of airway diseases such as asthma and COPD (Fahy *et al.*, 1993; Keatings & Barnes 1997; Wielders and Dekhuijzen 1997; Rutgers *et al.*, 2001). The specific degree of airway inflammation at the time of sputum collection can easily be determined by the presence and prevalence of cellular markers, measurements which are often critical in diagnosing and predicting

patient responses to treatment. Cell counts tend to correlate with respiratory physiology data and so may contain valuable information about the progression of a particular disease state (Lacy *et al.*, 2005). In addition, the method provides highly reproducible measurements, which makes it a preferred matrix for the diagnosis of specific airway diseases such as TB (Lacy *et al.*, 2005).

Robert Koch demonstrated the first microscopic staining method for detecting *M. tuberculosis* from sputum in 1882 (Dorman, 2010). Sputum smear microscopy is currently still used as a point-of-care (POC) method for TB diagnosis in endemic countries, the ease and speed of the technique outweighing its poor sensitivity (±60%) (Steingart *et al.*, 2006). This diagnostic method entails the collection of sputum and then "smearing" it thinly on a glass slide prior to staining with a dye that binds to mycobacteria specifically, making the bacteria easier to visualize and identify under a microscope (ECDC, 2018). Sputum smear microscopy, however, has significant limitations, one of which is a lack of sensitivity when the bacterial load is less than 10 000 organisms/mL. The outcome is also hampered in patients with extrapulmonary TB, paediatric TB, and patients co-infected with HIV (Desikan, 2013).

Although the current "gold standard" for TB diagnostics, bacteriological culture, provides the required sensitivity, it takes several weeks to obtain results, requiring well-equipped laboratories and trained personnel (Adler *et al.*, 2005). This extended turn-around time results in delayed diagnosis and treatment, as well as continued transmission and the risk of developing drug resistance (Storla *et al.*, 2008). The application of newer research fields, such as metabolomics, might lead to the identification of novel, more sensitive and specific TB biomarkers, with the capacity to elucidate previously unknown disease mechanisms, and potential to develop improved point of care (POC) TB devices.

2.4 Application of sputum in metabolomics research

Metabolomics studies conducted on sputum are limited. To date, most of these investigations have focussed on the characterisation of COPD, bronchial asthma, CF, lung cancer and TB. **Table 2.2** gives a broad overview of these studies, including the analytical platform used, metabolite markers identified, and statistical methods applied. **Table 2.2** was compiled to investigate whether the same type of metabolites identified in TB cases correlate with those in some of the other lung diseases mentioned above. The summary also highlights metabolites which were identified for more than one study (bold text).

***Table 2.2**: An overview of metabolomics studies using sputum as a sample matrix. Metabolites identified as markers in more than one study, are indicated in bold. Studies are listed in chronological order.*

Reference	Disease investigated	Analytical apparatus used	Disease markers	Statistical methods
Ahmed *et al.* (2022)	Lung cancer	LC-QTOF-MS	Glucose, adenosine monophosphate, N1, N12- diacetylspermine	MFC; Benjamini–Hochberg multiple correction method; Mann–Whitney paired test and Bonferroni test
Breen *et al.* (2022)	CF	LC-MS/MS, UHPLC-QTOF/MS[+], UHPLC-MS/MS with QHESI-II	C_{18} ceramides, **2-methylcitrate**, tryptophan associated metabolites, serine, threonine, **histidine**, **lysine**, tryptophan, leucylglycine, phenylalanylalanine, valylglycine,	sPLS-DA
Chandra *et al.* (2022)	TB	RP/UPLC-MS/MS[+], RP/UPLC-MS/MS[-], HILIC/UPLC-MS/MS[+]	Cholestenone(4-cholesten-3-one), methylsuccinate, methylcitrate, and 2-aminoadipate, methylsuccinate, methylmalonate,**2-methylcitrate**, 4,5-9,10-diseco-3-hydroxy-5,9,17-tri-oxoandrosta-1(10), 2-diene-4-oic acid (DSHA)	PCA; Random Forest; Box plots; ANOVA
Liu *et al.* (2022)	Bronchial asthma	UHPLC-QTOF/MS[+]	Adenosine 5′-monophosphate, allantoin, nicotinamide, histamine, **histidine**, 5-L-glutamyl-L-alanine, nicotinamide, dihydrothymine, L-leucine, L-phenylalanine, alanyl-leucine, phenylalanyl-serine, phenylalanylphenylalanine, glyceric acid, taurine, dihydrothymine, tyramine, L-glutamate, threoninyl-phenylalanine, taurine, glycerophosphocholine**, heptadecanoic acid**, **oleic acid**, dodecanoic acid	OPLS-DA; pathway topology enrichment analysis; Logistic and negative binomial regression models
Zhu *et al.* (2020)	COPD	LC-MS/MS, ELISA	Superoxide dismutase, myeloperoxidase, 8-iso-prostaglandin F2α	PCA; OPLS-DA; Pathway enrichment analysis
Tian *et al.* (2017)	Bronchial asthma	UHPLC-QTOF/MS[+]	1-Hexadecanoyl-sn-glycerol, Glycerol 1-stearate, Sphingosine, PC (16:0/16:0) 1-Hexadecanoyl-sn-glycerol, glutamic acid-proline , proline-arginine , L-arginine, phenylalanine-histidine, **N-Acetyl-D-glucosamine**, Gamma-L-Glutamyl-L-valine, Glutamic acid-Proline, lysine-proline, phenylalanine-isoleucine, phenylalanine-histidine, proline-valine, urocanic acid, tyrosine-proline, histidine-proline, tyrosine-proline, N-Acetyl-D-glucosamine, lysine-phenylalanine, threonine-phenylalanine, L-Citrulline, arginine-phenylalanine, adenine, phenylalanine-tyrosine, phenylalanine-glutamine, tyrosine-alanine, phenylalanine-serine	PCA; OPLS-DA; t-test
Tian *et al.* (2017)	Bronchial asthma	UHPLC-QTOF/MS[-]	Cytidine 2′,3′-cyclic phosphate, 1-Hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phospho-(1′-rac-glycerol), 1-Octadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoserine, 1-stearoyl-sn-glycero-3-phosphocholine, Thymine, p-Chlorophenylalanine, Phenylpyruvate, Phosphoenolpyruvate, Urocanic acid, Adenine, Gamma-L-Glutamyl-L-valine, Thymidine	PCA; OPLS-DA; t-test
Reference	**Disease investigated**	**Analytical apparatus used**	**Disease markers**	**Statistical methods**

Ahmed *et al.* (2016)	Lung cancer	^{1}H-NMR	**N-acetyl sugars,** glycoprotein, **propionate**, **lysine**, formate,methanol, ethanol, **acetate**, acetone	Fisher's exact test; Wilcoxon two-sample test
Cameron *et al.* (2016)	Lung cancer	FIE-MS and GC-MS	Putrescine, N,N,N-Trimethylethenaminium, glycerophospholipids of the cardiolipin(PC) class, isobutyl decanoate, diethyl glutarate, hexanal, cysteic acid, hydroxypyruvic acid, and the cholesterol ester with an acyl group CE (22:5(4Z,7Z,10Z,13Z,16Z)), gangliosideGM1 (18:1/12:0)	PCA; HCA with heat maps; RF; ROC-AUC
O'Shea *et al.* (2016)	Lung cancer	FIE-MS	Phenylacetic acid, L-fucose, caprylic acid, acetic acid, propionic acid, glycine	PCA; Welch t-tests
Quin *et al.* (2016)	CF	LC-MS/MS	Platelet activating factor, inflammatory lipids, monacylglycerophosphocholine lipid, ceramide	Bray-Curtis distance matrix (ANOSIM R and PerMANOVA F); ANOVA and Tukey's test with Bonferroni correction
Ghorbani *et al.* (2015)	CF	GC-FID	**Acetate**, **propionate**, butyrate (Short Chain Fatty Acids)	Paired nonparametric t-tests; linear regressions for correlation analyses
Saude *et al.* (2004)	CF	^{1}H-NMR	3-chlorotyrosine, 3-bromotyrosine and 3,5-dibromotyrosine	Forward and backward stepwise regression analyses; Correlation matrix; Two-sided nonparametric Mann-Whitney U-test
Telenga *et al.* (2014)	COPD	LC-Q-TOF MS/MS (+ESI and -ESI)	Sphingolipids, 28 ceramides, 11 dihydroceramides, 19 phytoceramides, 36 sphingomyelins, and 74 glycosphingolipids, solanesol, solanesyl palmitate, solanesyl stearate, bombiprenone, geranylfarnesylacetone and farnesylfarnesylacetone	Mann-Whitney unpaired analysis with Benjamini Hochberg false discovery rate correction; Mann-Whitney *U* test; Wilcoxon signed rank test; Spearman rank correlation test

Reference	Disease investigated	Analytical apparatus used	Disease markers	Statistical methods
Du Preez *et al.* (2013)	TB	GCxGC-TOFMS	à-D-glucopyranose-2-acetylamino-2-deoxy, à-D-glucopyranoside á-L-mannopyranose, à-D-mannopyranoside, D-galactose-6-deoxy, à-D-galactopyranose, D-glucosamine, **N-acetyl-glucosamine**, Methyl-17-methyl-octadecanoic acid, 10-heptadecenoic acid, **oleic acid**, nonadecanoic acid, 2-deoxy-D-erythro-pentitol), D-citramalic acid, D-gluconic acid d-lactone, glutaric acid, sebacic acid, thane, butanal g-aminobutyric acid (GABA), 3,4-dihydroxybutanoic acid, normetanephrine	PCA; PLS-DA; t-test; FC
Grasemann *et al.* (2012)	CF	LC/MS/MS	Polyamine, **putrescine,** spermidine, L-ornithine, acetyl choline, spermine,	Standard error of the mean (SEM), Spearmans test, Paired t-test, ANOVA, Tukey-Krammer test
Olivier *et al.* (2012)	TB	GC-MS	Hexadecanoic acid, **oleic acid**, tuberculostearic acid, palmitoleic acid, benzenepropanoicacid, **heptadecanoic acid**, hexadecane, hexacosanoic acid, tetracosanoic acid, eicosanoic acid, tetradecanoic acid, docosanoic acid,isopropyl-tetradecanoic acid, nervonic acid, squalene, erucic acid	PCA; PLS-DA
Yang *et al.* (2012)	CF	LC-MS/MS	9-oxo-11R,15S-dihydroxy-5Z,13E-prostadienoic acid; 9S,15S-dihydroxy-11-oxo-5Z,13E-prostadienoic acid; 9S,11,15S-trihydroxy-thromboxa-5Z,13E-dien-1-oic acid;5S,12R-dihydroxy-6Z,8E,10E,14Z-eicosatetraenoic acid; 5S,12R-dihydroxy-6E,8E,10E,14Z-eicosatetraenoic acid; 5S,12R,20-trihydroxy-6Z,8E,10E,14Z-eicosatetraenoic acid; 5S,12R-dihydroxy-6Z,8E,10E,14Z-eicosatetraene-1,20-dioic acid; 20-hydroxy-5Z,8Z,11Z,14Z-eicosatetraenoic acid;15S-hydroxy-5Z,8Z,11Z,13E-eicosatetraenoic acid; 11S-hydroxy-5Z,8Z,11E,14Z-eicosatetraenoic acid;12S-hydroxy-5Z,8Z,10E,14Z-eicosatetraenoic acid; 8S-hydroxy-5Z,9E,11Z,14Z-eicosatetraenoic acid; 9S-hydroxy-5Z,7E,11Z,14Z-eicosatetraenoic acid; 5S-hydroxy-6E,8Z,11Z,14Z-eicosatetraenoic acid; 14R,15S-epoxy-5Z,8Z,11Z-eicosatrienoic acid;11,12-epoxy-5Z,8Z,14Z-eicosatrienoic acid;8,9-epoxy-5Z,11Z,14Z-eicosatrienoic acid; 5,6-epoxy-8Z,11Z,14Z-eicosatrienoic acid; diols; 5S,12R,18R-trihydroxy-6Z,8E,10E,14Z,16E-eicosapentaenoic acid;11-oxo-5Z,9,12E,14E-prostatetraenoic acid;5S,6R,15S-trihydroxy-7E,9E,11Z,13E-eicosatetraenoic acid	PLS-DA
Lewis *et al.* (2010)	Lung cancer	FTIR	Protein, nucleic acid, glycogen levels.	HCA; PCA

Compounds in bold have been identified as markers in more than one study.

List of abbreviations:

UHPLC-MS/MS: Ultrahigh performance liquid chromatography-tandem mass spectrometry; Q-TOF: Quadrupole time-of-flight mass spectrometry; OPLS-DA: Orthogonal projections to latent structures discriminant analysis; LC: Liquid chromatography; GC: Gas chromatography; MS: Mass spectrometry; COPD: Chronic obstructive pulmonary disease; sPLS: Sparse partial least squares; PCA: Principal component analysis; ANOVA: Analysis of variance; NMR: Nuclear magnetic resonance; FID: Flame ionization detector; ESI: Electrospray ionization; MFC: Median fold change; RP: Reverse phase; HILIC: Hydrophilic interaction liquid chromatography; ANOSIM: Analysis of similarities; PerMANOVA: Permutational multivariate analysis of variance; HCA: Hierarchical clustering; RF: Random Forest; FC: Fold change; ROC-AUC: Receiver operating characteristic area under the curve; C: Carbon; FTIR: Fourier transform infrared.

2.4.1 Chronic obstructive pulmonary disease

COPD is the most common pulmonary disease found worldwide and probably best characterised by classical symptoms, including difficulty breathing, chronic cough and wheezing combined with irreversible damage caused to the lungs once the disease is set into motion (Zhu *et al.*, 2020). Several research groups have investigated COPD from sputum, using metabolomics methodologies:

Telenga *et al.* (2014) applied untargeted lipidomics to identify variations between smokers with COPD and non-smokers with COPD. They found that concentrations of glycerophospholipids and fatty acids in sputum were significantly lower in smokers with COPD in comparison to the non-smoking cohort. Zhu *et al.* (2020) showed the potential value of sputum metabolomics and the correlation between glycerophospholipid metabolites and oxidative stress products for predicting the severity of COPD. Esther *et al.* (2022) evaluated patients with COPD to predict disease severity and prospective exacerbations. The following pathways were found to be involved in mucus hydration: adenosine metabolism, methionine salvage and oxidative stress (their oxidative stress findings in line with what Zhu et al. found). Elevated sialic acid and hypoxanthine levels were associated with a shorter time to disease progression or severity and improved the predictive accuracy of models of future exacerbations.

Both Telenga *et al.* (2014) and Zhu *et al.* (2020) identified glycerophospholipids as one of the metabolic classes of interest for predicting the severity of COPD. Pathways involved in oxidative stress were also identified as significant for COPD characterisation by both Zhu *et al.* (2020) and Esther *et al.* (2022).

2.4.2 Lung cancer

Lung cancer is the most common cancer in the world and is responsible for 1.3 million deaths annually (World Health Organization, 2020). Doctors primarily rely on three tools for diagnosing lung cancer, namely: an X-ray of the lungs, a computed tomography (CT) lung scan, and a bronchoscopy. These methods have failed to improve early detection rates but have improved overall detectability of the disease (Sutedja, 2003). Metabolomics research has been applied by several groups to identify lung cancer induced variations to the sputum metabolome:

Lewis *et al.* (2010) conducted applied Fourier transform infrared spectroscopy (FTIR; a non-invasive technology that can detect a change of functional group in tissue or cell molecules) to sputum and

demonstrated that the technique could be used as a non-invasive method to identify biomarkers for lung cancer. Increased glycogen levels in the sputum of cancer patients were evident, while significant changes in proteins and nucleic acids were also observed. The method shows potential, having sufficient sensitivity and specificity to be used as a non-invasive, cost-effective, and high-throughput method for screening.

Cameron *et al.* (2016) performed metabolomics profiling using GC-MS and Flow Infusion Electrospray (FIE)-MS and identified putrescine, cysteic acid and ganglioside GM1 (18:1/12:0), the latter being a glycosphingolipid linked to a single sialic acid via its sugar group, as biomarkers with potential clinical application in lung cancer. Prospects would entail using the obtained biomarkers to detect different stages of lung cancer and the histological subtypes.

Ahmed *et al.* (2016) detected the absence of glucose, as well as decreased concentrations of N-acetyl sugars, propionate, lysine, glycoprotein, acetate and formate, in the sputum of lung cancer patients compared to that of patients in the benign state of the disease. The absence of glucose combined with the decrease in methanol concentration are metabolic biomarkers that allow for a ^{1}H MRS (magnetic resonance spectroscopy) screening method for earlier detection, routine screening for recurrence, and ultimately, to monitor the treatment response.

Ahmed *et al.* (2022) performed a pilot study where the differences in sputum and breath condensate before and after surgical resection of early-stage non-small cell lung cancer (NSCLC) was investigated. They reported significantly increased levels of glucose, adenosine monophosphate and N1, N12-diacetylspermine in sputum post-surgical resection. These alterations in the sputum metabolism form a foundation for further elucidating biomarkers for early detection of lung cancer and could serve as a non-invasive diagnostic tool (after validation of the findings via larger clinical trial studies).

O'Shea investigated sputum metabolomes of lung cancer patients and identified various organic acid and amino acid metabolite markers (eg. glycine and propionic acid) that differed significantly between patients with Non-small cell lung cancer NSCLC and small cell lung cancer (SCLC) indicating that sputum metabolic profiling has potential for screening of lung cancer.

In both studies conducted by Ahmed *et al.* (2016 & 2022) the group demonstrated an upregulation of glucose in sputum post-surgical resection. The N-acetyl sugar metabolites was also found to be one of the most significant metabolite classes characterising lung cancer (Ahmed *et al.*, 2016), which correlates with similar findings for bronchial asthma (Tian *et al.*, 2017) and TB (Du Preez *et al.*, 2013). This shows overlapping sputum metabolome variations across various lung diseases, and it could indicate that some of these compounds may possibly represent the general lung pathological state and/or lung damage, and is not specific to a disease.

2.4.3. Cystic Fibrosis

CF is a genetic disease that causes respiratory failure and various other complications such as the formation of scar tissue (fibrosis) and the development of cysts in the lungs. More than 1000 new CF cases are diagnosed globally each year (Wetmore *et al.*, 2010).

The sputum metabolome has been studied more extensively for CF than any other disease (**Table 2.2**):

Yang *et al.* (2012) performed a sputum oxylipins extraction for adult CF patients and identified a broad range of both proinflammatory and anti-inflammatory lipid mediators, never previously reported in CF respiratory tract secretions. All subjects exhibiting the presence of anti-inflammatory lipid mediator resolving E1 had notably superior lung function compared to individuals without. Broad metabolic profiling of regulatory lipid mediators in CF sputum could lead to a more comprehensive understanding of the molecular mechanism involved in CF respiratory tract inflammatory pathophysiology.

Grasemann *et al.* (2012) tested if L-ornithine derived polyamines are detectable in CF airways and whether they play any role in airway pathophysiology. The study found that sputum putrescine and spermidine levels were similar in stable CF patients and healthy controls, while spermine were significantly higher in the patients. Changes in the spermine levels in the sputum of patients undergoing treatment matched the alterations detected in L-ornithine.

Saude *et al.* (2004) designed their study around the theory that higher levels of modified tyrosine residues in the airways of individuals suffering with CF might be detectable by NMR. They therefore compared the production of tyrosine residues in sputum samples from control subjects to CF patients,

discovering that chlorinated and brominated tyrosine residues were present in sputum CF samples but not in the control group. Their findings confirmed that NMR is a useful tool for detecting biological markers of inflammatory airway disease using sputum. In addition, neutrophil and eosinophil activation can be detected in CF using this specific analytical platform, with neutrophil counts corresponding well with the presence of 3-chlorotyrosine.

Ghorbani *et al.* (2015) was able to identify a positive correlation between the established marker of CF airway inflammation, namely the number of neutrophils present in sputum, and the total sputum short chain fatty acids (SCFAs) in stable CF patients (excluding patients with pulmonary exacerbation). They then studied the effects of SCFAs on airway inflammatory responses using epithelial cell lines and primary cell cultures and described how SCFAs contribute to CF-specific adaptations to airway inflammation and infection.

Quin *et al.* (2016) found that the chemistry of sputum samples over the longitudinal profiles of patients are more comparable than unpaired patients with matching clinical states, concluding that there is value in tracking an individual while they experience numerous exacerbations (worsening pulmonary inflammatory episodes). This will provide better information regarding inflammatory processes/mechanisms prior to and during a CFPE (CF Pulmonary Exacerbation) event. The most significant elevated metabolites identified during such events were platelet activating factor and a related monacylglycerophosphocholine lipid. Related lipids were increased during intermittent pulmonary exacerbations (Pex), while ceramides were increased during the treatment phase of disease progression. The metabolome of CF sputum was confirmed to be largely patient specific, supporting a personalized outlook on molecular detection of the onset of Pex.

Breen *et al.* (2022) identified several divergent metabolites when comparing nontuberculous mycobacterial (NTM) infected CF cases to that of controls, the latter of which presented with lower tryptophan-related branched-chain amino acid metabolites and elevated phospholipid metabolism compounds. Variations in sputum 2-methylcitrate and certain ceramides were also observed, which have the potential to be used to identify risk factors and therapeutic targets for NTM infections in CF patients.

Regarding biomarkers identified in sputum for diagnosing CF Pex, putrescine was consistently higher in CF Pex individuals compared to healthy control and stable CF samples. Sputum nitrate was persistently decreased in CF Pex compared to stable CF samples in two separate studies, furthermore, lactic acid was found to possess diagnostic potential (Nguyen *et al.*, 2022).

2.4.4. Asthma

Asthma is a chronic respiratory disease characterised by inflammation and narrowing of the airways, resulting in recurring episodes of wheezing, coughing, and difficulty breathing. Various metabolites associated with asthma aid in the characterisation of the underlying disease mechanisms and were identified in different sample matrices i.e., serum, plasma, urine, local tissue, exhaled breath condensate, bronchoalveolar lavage fluid and stool (Wang *et al.*, 2021). In accordance, metabolomics studies have also investigated asthma using sputum as the sample matrix:

Tian *et al.* (2017) showed that sputum metabolomics can discern between patients with asthma and healthy subjects, identifying the most affected pathways as glycolysis, glycerophospholipid metabolism, inositol phosphate metabolism, and gluconeogenic metabolic pathways. Their findings prove that sputum metabolomics can be utilized not only for the early diagnosis of asthma but also for predicting the severity of symptoms.

Liu *et al.* (2022) indicated that various inflammatory asthma phenotypes have distinct sputum metabolic profiles. Pathway topology analysis showed how histidine-, glycerophospholipid-, nicotinate- and nicotinamide-, linoleic acid metabolism including phenylalanine, tyrosine and tryptophan biosynthesis played a prominent role in the pathogenesis of various asthma phenotypes. They also uncovered evidence of the predictive potential of adenosine 5'-monophosphate and nicotinamide for severe asthma exacerbation progression.

Interestingly, studies have identified significant perturbations in metabolic pathways such as phenylalanine, tyrosine, and tryptophan biosynthesis, as well as valine, leucine, and isoleucine degradation in both asthma (Tian et al., 2017; Liu et al., 2022) and cystic fibrosis (Breen et al., 2022).

These shared metabolic pathways suggest potential common mechanisms between lung diseases such as CF, TB and asthma for example, highlighting the importance of investigating metabolic profiles across different respiratory diseases.

2.4.5 Tuberculosis

Up to 2019, only nine studies have applied metabolomics methodologies to TB patient sputum samples (summarised in Du Preez *et al.*, 2019), most of which focused specifically on the analyses of lipids. In 2022, Chandra *et al.* successfully differentiated individuals with TB and TB-negative individuals presenting with TB-like symptoms, based on their cholestenone levels. The authors hypothesized that the cholestenone build-up in TB-positive cases may point to a substitute role for cholestenone in pathogenesis and serve as potential clinically functional biomarker of TB infection. Only two TB-related studies (Schoeman *et al.*, 2012; Du Preez & Loots, 2013) have applied an untargeted approach to sputum, both by our group, as discussed in the problem statement section. The outcome of these studies led to -several scientific contributions, including: 1) confirmation of the previously proposed citramalate cycle in *M. tuberculosis*; 2) the interaction of this cycle with an up-regulated glyoxylate cycle during pulmonary *M. tuberculosis* infection; 3) the increased utilisation of fatty acids and glutamate as alternative carbon sources in *M. tuberculosis* during pulmonary infection; 4) an alternative mechanism by which the host produces hydrogen peroxide via glucose oxidation, in order to eliminate the bacterial infection; 5) inhibition of the electron transport chain (ETC) due to pronounced oxidative stress during an active TB disease state, resulting in increased concentrations of various neurotransmitters and other metabolites previously associated with an inborn error of metabolism (MADD/GA type II); and 6) elevated concentrations of neurotransmitters potentially explaining several symptoms associated with TB.

2.5 Metabolome variations related to sex and HIV in healthy individuals

None of the metabolomics studies done to date (on any sample matrix) have investigated the effect of sub-cohorts based on sex and HIV status on the identified TB-induced metabolome variations, although these factors independently have been shown to influence metabolite profiles of healthy individuals. Darst *et al.* (2019), for example, indicated that 56,8% (n=623) of the metabolites identified in healthy human plasma showed correlations of 63,4% (n=695) with sex. In an earlier study, a clear variance between the metabolic profiles of healthy men and women was described using metabolomics as the principal tool of investigation. Phosphatidylcholines' and in acylcarnitines' classes showed significant differences between males and females (Li *et al.*, 2018). Sex-specific pathways in the human serum metabolome have also been elucidated by Krumsiek *et al.* (2015). In this study, the

entire class of amino acids as well as the entire super-pathway of carbohydrates (specifically glycolysis, gluconeogenesis, pyruvate, fructose, mannose, galactose, starch, and sucrose metabolisms) were significantly higher in males. Regarding lipids, androsterone sulfate derivatives were significantly higher in males, whereas glycerol and myristoleate were elevated in females. In the nucleotide metabolite class, the purine/urate metabolism was significantly higher in males (Krumsiek *et al.*, 2015). According to our knowledge, sex-related variations specifically, have not been investigated in the sputum metabolome to date. When investigating sex as a confounding variable in the sputum metabolome of lung cancer patients, O'Shea *et al.* (2016) identified no significant effect.

Due to the natural variation that exists within TB cohorts, whether said variation is due to diet, cofactors such as smoking or non-smoking individuals, or even then factors like sex and HIV status, separation between TB positive and TB negative groups remain difficult to achieve. This study is not aimed on investigating how sex and HIV status for example impacts the sputum metabolome in the context of TB infection, but if such an impact or direct relationship is identified, it would add great value to better understand the impact of these factors on the overall TB metabolome.

Regarding HIV, various studies have applied metabolomics as a tool to investigate different aspects of HIV, using diverse sample matrices (Liebenberg *et al.*, 2021). Cassol *et al.* (2013), for example, could differentiate the untargeted plasma metabolic profiles of HIV patients on antiretroviral therapy and healthy controls, identifying increased bile acids and acylcarnitines and decreased sulfated steroids, polyunsaturated fatty acids, and lysophosphocholine (LPC) in the patient samples, comparatively. No metabolomics study to date has, however, used sputum as the sample matrix of choice to investigate HIV. In relation, Cribbs *et al.* (2014), could differentiate the metabolome profiles of bronchoalveolar lavage fluid collected from HIV infected individuals and healthy controls, however, most of the identified features could not be annotated, while others represented a disturbance in the lung microbiome.

In this study, we investigated the effect of sex and HIV status on the TB-induced sputum metabolome, using an untargeted GCxGC-TOFMS metabolomics approach. These outcomes could assist in optimising the study designs of future metabolomics investigations for TB disease characterisation.

2.6 References

Adahchour, M., Beens, J., Brinkman, U.A.T. 2008. Recent developments in the application of comprehensive two-dimensional gas chromatography. *Journal of Chromatography. A,* 1186:67-108.

Adler, H., Straub, C. & Frei, R. 2005. Comparison of BacT/ALERT 3D, LOWENSTEIN-JENSEN medium and MIDDLEBROOK 7H10/7H11 biplate for recovering mycobacteria from clinical specimens. *European Journal of Clinical Microbiology & Infectious Diseases*, 24:499–500.

Ahmed, N., Bezabeh, T., Ijare, O. B., Myers, R., Alomran, R., Aliani, M., *et al.* 2016. Metabolic Signatures of Lung Cancer in Sputum and Exhaled Breath Condensate Detected by 1H Magnetic Resonance Spectroscopy: A Feasibility Study. *Magnetic resonance insights*, 9, 29–35.

Ahmed, N., Kidane, B., Wang, L., Nugent, Z., Moldovan, N., McElrea, A., *et al.* 2022. Metabolic Alterations in Sputum and Exhaled Breath Condensate of Early Stage Non-Small Cell Lung Cancer Patients After Surgical Resection: A Pilot Study. *Frontiers in oncology,* 12, 874964.

Barnes, P.J., Chowdhury, B., Kharitonov, S.A., Magnussen, H., Page, C.P., Postma, D., Saetta, M. 2006. Pulmonary biomarkers in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med,* 174:6-14.

Breen, P., Zimbric, M., Opron, K., & Caverly, L. J. 2022. Sputum Metabolites Associated with Nontuberculous Mycobacterial Infection in Cystic Fibrosis. *mSphere,* 7(3), e0010422.

Cameron, S. J., Lewis, K. E., Beckmann, M., Allison, G. G., Ghosal, R., Lewis, P. D., & Mur, L. A. 2016. The metabolomic detection of lung cancer biomarkers in sputum. *Lung cancer*, *94*, 88–95.

Chandra, P., Coullon, Afarwak, M., H., Goss, C.W., Philips, J.A. 2022. Macrophage global metabolomics identifies cholestenone as host/pathogen cometabolite present in human Mycobacterium tuberculosis infection. *J Clin Invest.*, 132(3):e152509.

Cohn, L., Whittaker, L., Niu, N., Homer, R.J. 2002. Cytokine regulation of mucus production in a model of allergic asthma, p 201–220. *Wiley Online Library*, Hoboken, NJ.

Collino, S., Martin, F.P.J., Rezzi, S. 2013. Clinical metabolomics paves the way towards future healthcare strategies. *Br J Clin Pharmacol*, 75:619–29.Dimandja, J.D. 2004. Comprehensive 2-D GC provides high-performance separations in terms of selectivity, sensitivity, speed, and structure. *Anal. Chem*., 76, 9, 167 A–174 A

Cowburn, A. S., Condliffe, A. M., Farahi, N., Summers, C., & Chilvers, E. R. 2008. Advances in neutrophil biology: clinical implications. *Chest*, *134*(3), 606–612.

Cystic Fibrosis Foundation Patient Registry (CFFPR) 2010 Annual Data Report (2011) Bethesda, Maryland: Cystic Fibrosis Foundation. Available: https://www.cff.org/intro-cf/about-cystic-fibrosis. Accessed 9 December 2022.

Desikan, P. 2013. Sputum smear microscopy in tuberculosis: Is it still relevant? *Indian Journal of Medical Research*, 137(3): 442–444.

Dorman, S.E. 2010. New Diagnostic Tests for Tuberculosis: Bench, Bedside, and Beyond. *Clinical Infectious Diseases*, 50 (Supplement 3): S173–177

Dunn, W.B., & Ellis, D.I. 2005. Metabolomics: Current analytical platforms and methodologies. *Trends in Analytical Chemistry*, 24(4): 285–294.

Du Preez, I. and Loots, D.T. 2013. New sputum metabolite markers implicating adaptations of the host to Mycobacterium tuberculosis, and vice versa. Tuberculosis, 93(3): 330–337.

Eiserbeck, C., Nelson, R.J, Grice., K., Curiale., J., Reddy, C.M., Raiteri, P. 2011. Separation of 18α(H)-, 18β(H)-oleanane and lupane by comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, Volume 1218, 32: 5549-5553.

Eiserbeck, C., Nelson, R., Grice, K., Curiale, J., Reddy, C. 2012. Comparison of GC–MS, GC–MRM-MS, and GC × GC to characterise higher plant biomarkers in Tertiary oils and rock extracts. *Geochimica et Cosmochimica Acta*, 87:299-322.

Emwas, A. H., Roy, R., McKay, R. T., Tenori, L., Saccenti, E., Gowda, G. A. N., *et al.* (2019). NMR Spectroscopy for Metabolomics Research. *Metabolites*, 9(7), 123.

Esther, C.R., Jr, O'Neal, W.K., Anderson, W.H., Kesimer, M., Ceppe, A., Doerschuk, C.M., *et al.* 2022. Identification of Sputum Biomarkers Predictive of Pulmonary Exacerbations in COPD. *Chest*, 161(5), 1239–1249.

European Centre for Disease Prevention and Control (ECDC). Handbook on tuberculosis laboratory diagnostic methods in the European Union – Updated 2018. Stockholm: ECDC; 2018. https://www.ecdc.europa.eu/sites/portal/files/documents/TB-handbook-updated-2018.pdf

Fahy, J. V., Liu, J., Wong, H., & Boushey, H. A. 1993. Cellular and biochemical analysis of induced sputum from asthmatic and from healthy subjects. *The American review of respiratory disease*, *147*(5), 1126–1131.

Fahy, J. V., Boushey, H. A., Lazarus, S. C., Mauger, E. A., Cherniack, R. M., Chinchilli, V. M., *et al.* (2001). Safety and reproducibility of sputum induction in asthmatic subjects in a multicenter study. *Am J Respir Crit Care Med*, 163(6), 1470–1475.

Gibson, P. G., Dolovich, J., Denburg, J., Ramsdale, E. H., & Hargreave, F. E. 1989. Chronic cough: eosinophilic bronchitis without asthma. *Lancet*, *1*(8651), 1346–1348.

Giddings, J.C. 1984. Two-dimensional separations – concept and promise. *Analytical Chemistry*, 56 (1984), pp. 1258-1270.

Gonzalez-Covarrubias, V., Martínez-Martínez. E., Del Bosque-Plata, L. 2022. The Potential of Metabolomics in Biomedical Applications. *Metabolites*. 19;12(2):194.

Grasemann, H., Shehnaz, D., Enomoto, M., Leadley, M., Belik, J., & Ratjen, F. 2012. L-ornithine derived polyamines in cystic fibrosis airways. *PloS one*, 7(10), e46618.

Ghorbani, P., Santhakumar, P., Hu, Q., Djiadeu, P., Wolever, T. M., Palaniyar, N., & Grasemann, H. 2015. Short-chain fatty acids affect cystic fibrosis airway inflammation and bacterial growth. *The European respiratory journal*, 46(4), 1033–1045.

Haince, J.-F., Joubert, P., Bach, H., Ahmed Bux, R., Tappia, P.S., Ramjiawan, B. 2022. Metabolomic Fingerprinting for the Detection of Early-Stage Lung Cancer: From the Genome to the Metabolome. *Int. J. Mol. Sci.,* 23, 1215.

Harris, J., De Haro, S.A., Master, S.S., Keane, J., Roberts, E.A., Delgado, M., Deretic, V. 2007. T helper 2 cytokines inhibit autophagic control of intracellular Mycobacterium tuberculosis. *Immunity,* 27:505–517.

Jaiyesimi, O.A., McAvoy, A.C., Fogg, D.N., & Garg, N. 2021. Metabolomic profiling of Burkholderia cenocepacia in synthetic cystic fibrosis sputum medium reveals nutrient environment-specific production of virulence factors. *Scientific reports*, 11(1), 21419.

Jayaram, L., Parameswaran, K., Sears, M. R., & Hargreave, F. E. 2000. Induced sputum cell counts: their usefulness in clinical practice. *The European respiratory journal*, *16*(1), 150–158.

Kaddurah-Daouk, R., Kristal, B.S., Weinshilboum, R.M. 2008. Metabolomics: a global biochemical approach to drug response and disease. *Annual review of pharmacology and toxicology*, 48(1), pp.653-683.

Keatings, V. M., & Barnes, P. J. 1997. Granulocyte activation markers in induced sputum: comparison between chronic obstructive pulmonary disease, asthma, and normal subjects. *American journal of respiratory and critical care medicine*, *155*(2), 449–453.

Lacy, P., Lee, L.J., Vethanayagam, D. 2005. Sputum analysis in diagnosis and management of obstructive airway diseases. *Therapeutics and Clinical Risk Management,* 1(3) 169–179.

Lienhardt, C., Azzurri, A., Amedei, A., Fielding, K., Sillah, J., Sow, O.Y., *et al.* 2002. Active tuberculosis in Africa is associated with reduced Th1 and increased Th2 activity in vivo. *Eur J Immunol*, 32:1605–1613.

Liou, T. G., Adler, F. R., Fitzsimmons, S. C., Cahill, B. C., Hibbs, J. R., & Marshall, B. C. 2001. Predictive 5-year survivorship model of cystic fibrosis. *American journal of epidemiology*, *153*(4), 345–352.

Liou, T. G., Adler, F. R., Keogh, R. H., Li, Y., Jensen, J. L., Walsh, W., Packer, K., *et al.* 2012. Sputum biomarkers and the prediction of clinical outcomes in patients with cystic fibrosis. *PloS one*, *7*(8), e42748.

Liu, Y., Zhang, X., Zhang, L., Oliver, B. G., Wang, H. G., Liu, Z. P., *et al.* 2022. Sputum Metabolomic Profiling Reveals Metabolic Pathways and Signatures Associated With Inflammatory Phenotypes in Patients With Asthma. *Allergy, asthma & immunology research,* 14(4), 393–411.

Lei, Z., Huhman, D. V., Sumner, L. W. 2011. Mass Spectrometry Strategies in Metabolomics. *Journal of Biological Chemistry*, Volume 286(29): 25435-25442.

Lewis, P.D., Lewis, K.E., Ghosal, R., Bayliss, S., Lloyd, A.J., Wills, J., *et al.* 2010. Evaluation of FTIR spectroscopy as a diagnostic tool for lung cancer using sputum. *BMC Cancer* 10: 640.

Luies, L. and Du Preez, I. 2020. The Echo of Pulmonary Tuberculosis: Mechanisms of Clinical Symptoms and Other Disease-Induced Systemic Complications. *Clinical Microbiology Reviews*, 33(4).

Merritt, D.A. & Hayes, J.M. 1994. Factors Controlling Precision and Accuracy in Isotope-Ratio-Monitoring Mass Spectrometry. *Anal. Chem.*, 66, 2336– 2347

Nguyen, A. V., Haas, D., Bouchard, M., & Quon, B. S. 2022. Metabolomic Biomarkers to Predict and Diagnose Cystic Fibrosis Pulmonary Exacerbations: A Systematic Review. *Frontiers in pediatrics*, 10, 896439.

Olivier, I., & Loots, duT. 2012. A metabolomics approach to characterise and identify various Mycobacterium species. *Journal of microbiological methods*, 88(3), 419–426.

Pizzichini, E., Pizzichini, M. M., Efthimiadis, A., Evans, S., Morris, M. M., Squillace, D., *et al.* 1996. Indices of airway inflammation in induced sputum: reproducibility and validity of cell and fluid-phase measurements. *American journal of respiratory and critical care medicine*, *154*(2 Pt 1), 308–317.

Pizzichini, M. M., Pizzichini, E., Clelland, L., Efthimiadis, A., Mahony, J., Dolovich, J., & Hargreave, F. E. 1997. Sputum in severe exacerbations of asthma: kinetics of inflammatory indices after prednisone treatment. *American journal of respiratory and critical care medicine*, *155*(5), 1501–1508.

Quinn, R. A., Lim, Y. W., Mak, T. D., Whiteson, K., Furlan, M., Conrad, D., Rohwer, F., & Dorrestein, P. 2016. Metabolomics of pulmonary exacerbations reveals the personalized nature of cystic fibrosis disease. *PeerJ*, 4, e2174.

Ribbenstedt, A., Ziarrusta, H., & Benskin, J. P. 2018. Development, characterization and comparisons of targeted and non-targeted metabolomics methods. *PloS one*, 13(11), e0207082.

Rutgers, S.R., Timens, W., Kauffman, H.F., & Postma, D.S. 2001. Markers of active airway inflammation and remodelling in chronic obstructive pulmonary disease. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*, *31*(2), 193–205.

Saude, E. J., Lacy, P., Musat-Marcu, S., Mayes, D. C., Bagu, J., Man, S. F., Sykes, B. D., & Moqbel, R. 2004. NMR analysis of neutrophil activation in sputum samples from patients with cystic fibrosis. *Magnetic resonance in medicine*, 52(4), 807–814.

Sciarrone, D., Ragonese, C., Carnovale, C., Piperno, A., Dugo, P., Dugo, G., Mondello, L. 2010. Evaluation of tea tree oil quality and ascaridole: A deep study by means of chiral and multi heart-cuts multidimensional gas chromatography system coupled to mass spectrometry detection. *Journal of Chromatography.* A, 1217:6422-6427.

Singh, D., Edwards, L., Tal-Singer, R., Rennard, S. 2010. Sputum neutrophils as a biomarker in COPD: findings from the ECLIPSE study. *Respir Res*. 15;11(1):77.

Spahn, J.D. 2007. Asthma biomarkers in sputum. Immunol. *Allergy Clin. North Am.*, 27, 387–399.

Steingart, K.R., NG, V., Henry, M., Hopewell, P.C, Ramsay, A, Cunningham, J., *et al.* 2006. Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. *Lancet Infectious Diseases*, 6(10): 664–674.

Storla, D.G., Yimer, S. & Bjune, G. A. 2008. A systematic review of delay in the diagnosis and treatment of tuberculosis. *BMC Public Health*, 8: 15.

Sutedja, G. 2003. New techniques for early detection of lung cancer. *Eur. Respir. J.* 39: 57s–66s

Tian, M., Chen, M., Bao, Y. L., Xu, C. D., Qin, Q. Z., Zhang, W. X., He, Y. T., & Shao, Q. 2017. Sputum metabolomic profiling of bronchial asthma based on quadruple time-of-flight mass spectrometry. International journal of clinical and experimental pathology, 10(10), 10363–10373.

Telenga ED, Hoffmann RF, Ruben TK, Hoonhorst SJ, Willemse BW, van Oosterhout AJ, *et al.* 2014. Untargeted lipidomic analysis in chronic obstructive pulmonary disease. Uncovering sphingolipids. *Am J Respir Crit Care Med.* 190(2):155–64.

Ventura, G.T., Kenig, F., Reddy, C.M., Schieber, J., Frysinger, G.S., Nelson, R.K., *et al.* 2007. Molecular evidence of Late Archean archaea and the presence of a subsurface hydrothermal biosphere. *Proceedings of the National Academy of Sciences*, 104 (36) 14260-14265.

Voynow, J. A., & Rubin, B. K. 2009. Mucins, mucus, and sputum. *Chest*, 135(2), 505–512.

Wang C, Jiang S, Zhang S, Ouyang Z, Wang G, Wang F. 2021. Research Progress of Metabolomics in Asthma. *Metabolites,* 11(9):567.

Wielders, P.L., & Dekhuijzen, P.N. 1997. Disease monitoring in chronic obstructive pulmonary disease: is there a role for biomarkers?. *The European respiratory journal*, *10*(11), 2443–2445.

Wetmore, D. R., Joseloff, E., Pilewski, J., Lee, D. P., Lawton, K. A., Mitchell, M. W., *et al.* 2010. Metabolomic profiling reveals biochemical pathways and biomarkers associated with pathogenesis in cystic fibrosis cells. *The Journal of biological chemistry*, *285*(40), 30516–30522.

WHO, Cancer Factsheet, WHO Fact Sheets, Number 2970. (2013). http://www.who.int/mediacentre/factsheets/fs297/en/. Accessed 6 December 2022.

Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., *et al.* 2007. HMDB: the human metabolome database. *Nucleic Acids Res*, 35:D521–6.

World Health Organization. 2020. International Agency for Research on Cancer (IARC). World Cancer Report 2020. https://publications.iarc.fr/586. Accessed 30.04.2023.

Yang, P., Chan, D., Felix, E., Madden, T., Klein, R. D., Shureiqi, I., Chen, X., Dannenberg, A. J., & Newman, R. A. 2006. Determination of endogenous tissue inflammation profiles by LC/MS/MS: COX- and LOX-derived bioactive lipids. *Prostaglandins, leukotrienes, and essential fatty acids*, 75(6), 385–395.

Yu, L., Li, K., Zhang, X. 2017. Next-generation metabolomics in lung cancer diagnosis, treatment and precision medicine: mini review. *Oncotarget.,* 8(70):115774-115786.

Zhao, J., Schloss, P.D., Kalikin, L.M., Carmody, L.A., Foster, B.K., Petrosino, J.F., *et al.* 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc Nat Acad Sci*, 109:5809–14.

Zhao, J., Evans, C. R., Carmody, L. A., & LiPuma, J. J. 2015. Impact of storage conditions on metabolite profiles of sputum samples from persons with cystic fibrosis. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*, *14*(4), 468–473.

Zhu, T., Li, S., Wang, J., Liu, C., Gao, L., Zeng, Y., Mao, R., Cui, B., Ji, H., Chen, Z. 2020. Induced sputum metabolomic profiles and oxidative stress are associated with chronic obstructive pulmonary disease (COPD) severity: potential use for predictive, preventive, and personalized medicine. *EPMA J.* 4;11(4):645-659.

Chapter 3: Materials and Methods

3.1 Introduction

To reach the set objectives of the study, samples were collected from TB+ and TB- individuals via a diagnostic pathology laboratory. A previously developed, untargeted GCxGC-TOFMS method was evaluated and applied to the collected sputum samples and the obtained data statistically analysed to determine the metabolic variation induced by TB, irrespective of sub-cohort classifications. Hereafter, the effect of sex and HIV-status on the TB-induced metabolome was investigated.

3.2 Experimental design

To establish the method as fit-for-purpose, the analytical and extraction precision was determined. Analytical precision gives an indication of how repeatable the analytical instrument can detect and measure the analyte features from the same extract. Good analytical precision minimises variations in the data due to instrument instability. Extraction precision indicates the repeatability of the extraction method, where the same analyst prepares several aliquots from the same sample using identical reagents and consumables. Favourable extraction precision is indicative of the extraction efficiency and consistency of the preparation method, that is, confirming the method can be applied to samples in an adequately repeatable manner. These measures, in combination, stipulate the competency of the analyst (MSc candidate in this case) in the method, spanning both the sample preparation procedure as well as operation of the analytical instrument (GCxGC-TOFMS).

The patient sputum samples collected were randomly assigned to 19 analytical batches, which were consecutively extracted and analysed on the GCxGC-TOFMS by the same analyst (MSc candidate). Raw mass spectral data, obtained from all samples in all analytical batches, were processed simultaneously and aligned to obtain a single data matrix. Established data processing was performed on this data matrix prior to statistical analysis. Next, to reach the aim of the study, the processed data of the individuals were assigned to six subgroups, including: all participants, all female participants, all male participants, all HIV+ participants, HIV+ female participants, and HIV+ male participants (**Figure 3.1**). For each of the subgroups, the statistically significant metabolite profiles between TB+ and TB- individuals were shortlisted for interpretation. These marker lists were then cross-compared to

identify the similarities and differences between the TB-induced metabolome variations in the different subgroups.

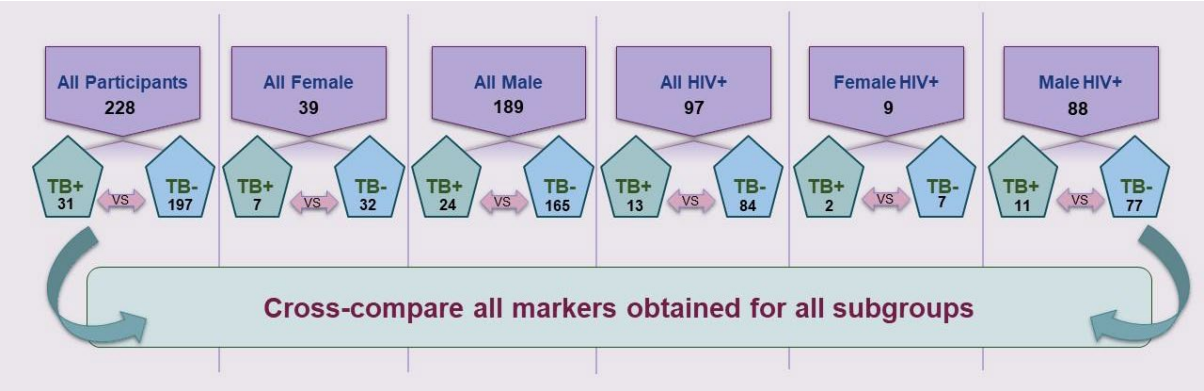


Figure 3.1*: Experimental design and overview of the composition of the participants in each sample subgroup. Numerical values indicate the total number of participants per subgroup, and double-headed arrows indicate the cohort comparisons used in the statistical analyses.*

3.3 Materials and methods

3.3.1 Ethical approval

Ethical approval for the larger study, of which this MSc is a sub-study, was submitted and granted by the Ethics committee of the North-West University, NWU HREC approval number: NWU-00127-11-A1-03.

3.3.2 Sample collection

Sputum samples from patients suspected of having TB, based on a medical assessment of the symptoms associated with the disease, were collected at facilities from all over South Africa, using standard sputum collection procedures. Samples were transported then on ice to a centralised national diagnostic/pathology laboratory, where standard diagnostic procedures, including both Ziehl-Neelsen staining and bacteriological culture, were performed. After the required amount of sample material was removed for the completion of individual diagnostic tests, the remnant proportion of these samples were frozen (-80°C) and transported to the North-West University, Centre for Human Metabolomics, for metabolomics analyses pertaining to this study. No additional inclusion or

exclusion criteria were enforced, that is, all samples collected were included in the analysis. Anonymity was safeguarded by assigning a unique code to each sample prior to transport. Clinical information obtained regarding these anonymous samples included: HIV status (only where the patient requested for an HIV test to be performed), age at time of collection and sex.

3.3.3 Quality control samples

Due to the limited sample volumes available for the study, one TB- sample, with a comparatively higher volume, was used to determine analytical and extraction precision. After homogenisation with a 45% ethanol solution (Schoeman *et al.*, 2012), this sample was divided into 750uL aliquots (equivalent to 250uL sputum) and dried overnight in a solvent evaporator. As an indication of analytical precision, one aliquot was extracted and derivatized, and this extract was consecutively injected on the GCxGC-TOFMS six times. To determine the extraction precision, five aliquots were extracted by the same analyst (MSc candidate) and analysed on the GCxGC-TOFMS.

To determine the severity of and, if necessary correct for, analytical variation within and between the 19 batches, a pooled QC sample was compiled. For this purpose, 24 of the TB-negative samples were randomly selected (not enough TB positive sample volume was available and thus no batch correction of the test samples, using QC samples can be performed due to this) and equal volumes of these samples were pooled as follows: 250µl from each of the selected TB-negative samples (total volume: 6ml sputum) were combined and then homogenized using a 45% ethanol solution (Schoeman *et al.*, 2012). The master QC sample was then divided into 750uL aliquots (equivalent to 250uL pooled sputum) and dried overnight in a solvent evaporator. One QC aliquot was extracted with each of the 19 experimental batches (about 12 samples per batch) and injected at the beginning, middle and end of each batch.

3.3.4 Sample extraction and derivatization

Prior to extraction, 250 µL of each patient sample was homogenized using ethanol (Schoeman *et al.*, 2012) and dried in a solvent evaporator. The precision aliquots, pooled QC aliquots and homogenized experimental samples were subjected to a whole metabolome extraction method described by Schoeman et al., 2012.. After the method performance was evaluated and found sufficiently precise,

experimental samples were randomly analysed across subgroups, in 19 batches. In short, 50 µL of the internal standard, 3-phenylbutyric acid (0.525 mg mL^{-1}), was added to 250 µL of the sample. Hereafter, 1 mL of an extraction solvent mixture, consisting of ultra-pure Burdick & Jackson brand chloroform:methanol:water (1:3:1), was added to the tubes (Honeywell International Inc., Muskegon, MI, USA). The extraction was performed using a MM 400 mixer mill (Retsch GmbH & co. KG, Haan, Germany) at a frequency of 30 Hz, for 5 min, after the addition of a 3 mm tungsten carbide bead to each sample tube to increase the extraction efficacy. Following centrifugation (4 °C at 21 952 x g for 10 min), the supernatant (including the organic and water phases, excluding the pellet) was collected, transferred to a GC-MS sample vial, and dried under a light stream of nitrogen. Following this, samples were derivatized using 50 µL of methoxyamine hydrochloride in pyridine (15 mg/mL) (Merck, Darmstadt, Germany) at 50°C for 90 min, followed by 40 µL of N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) (Sigma-Aldrich, St. Louis, MO, USA) with 1 % TMCS at 60°C for 60 min. The extracts were transferred to a 0.1 mL insert in a sample vial and capped prior to GCxGC-TOFMS analysis.

3.3.5 GCxGC/TOFMS analysis

One microliter of each sample extract was injected (1:5 split ratio) onto a Pegasus 4D GCxGC-TOFMS (Leco Corporation, St. Joseph, MI, USA), which comprises an Agilent 7890A GC (Agilent, Atlanta, GA) coupled to a time of flight mass spectrometer (Leco Corporation, St. Joseph, MI, USA) equipped with a Gerstel Multipurpose Sampler (Gerstel GmbH & co. KG, Eberhard-Gerstel- Platz 1, D-45473 Mülheim an der Ruhr). First dimensional separation was achieved with a Rxi-5Sil MS primary column (29.245 m, 0.25 mm internal diameter, 0.25 µm film thickness) (Restch GmbH & co. KG, Haan, Germany) and a Rxi-17 capillary column (1.400 m, 0.1 mm internal diameter, 0.1 µm film thickness) as the secondary column (Restch GmbH & co. KG, Haan, Germany). The front inlet temperature was held constant at 270°C for the entire run, ensuring rapid vaporization. For the primary oven, an initial GC oven temperature was set at 70°C for 2 min, followed by an initial increase in oven temperature of 4°C/min to a final temperature of 300°C, which was held for 2 min. The secondary column oven temperature was set at 85°C for 2 min, then increased by 4°C/min, until a final temperature of 300°C, at which it was maintained for a further 2 min. The initial temperature of the modulator was 100°C for 2 min, followed by a 4°C/min increase to a final temperature of 310°C held for 9 min. To control the effluent from the primary onto the secondary column, cryomodulation and a hot pulse of nitrogen gas of 0.5 s, every 3 s was used. The acquisition delay for each run was 450 s and the transfer line temperature

was held constant at 270°C, with the ion source temperature maintained at 200°C. The detector voltage was adjusted to 1500 V with the filament bias -70 eV. Spectra were collected from 50–800 m/z at an acquisition rate of 200 spectra per second.

Mass spectral deconvolution, peak alignment, and peak identification were performed using Leco Corporation's ChromaTOF software (version 4.51). Mass spectral deconvolution was performed at a signal-to-noise ratio of 100, with a minimum of three apexing peaks. To eliminate the effect of retention time shifts and to create a data matrix containing the relative abundance of all compounds present in all samples, peaks with similar mass spectra and retention times were aligned using Statistical Compare, a package of ChromaTOF. Mass fragmentation patterns and their respective retention times were screened against commercially available National Institute of Standards and Technology (NIST) spectral libraries (mainlib, replib) for peak annotation, with a similarity setting of at least 80%.

3.4 Statistical data analyses

To determine method performance, Microsoft Excel was used to calculate the coefficient of variance (CV) for each compound (peak areas) detected in the respective analytical and extraction precision experiments. The analytical and extraction precision was deemed acceptable if at least 70% of the compounds had a CV < 50%, as per the inhouse GCxGC-TOFMS procedure.

Statistical analysis of the experimental batch data was performed using MetaboAnalyst (version 5.0) (Pang *et al.*, 2021). Data were normalised to the sample median, log transformed and auto scaled prior to analysis. First, the presence and severity of any batch effect was evaluated, and a summarized view of the data generated using principal component analyses (PCA).

Due to the inequality between subgroup sample sizes (**Figure 3.1**), the usual run of the mill metabolomics statistics was considered inappropriate for this study. Instead, compounds which vary most between the TB+ and TB- participants within each of the subgroups were identified as those with at least a moderate probability of superiority, used here as a robust nonparametric effect size (ES), exceeding 0.64 (Li 2016).

3.5 References

Li, J. C.-H. 2016. Effect size measures in a two-independent-samples case with nonnormal and nonhomogeneous data. *Behavior research methods,* 48(4), 1560–1574.

Pang, Z., Chong, J., Zhou, G., de Lima Morais, D. A., Chang, L., Barrette, M., *et al.* 2021. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic acids research*, 49(W1), W388–W396.

Schoeman, J.C., Du Preez, I., Loots, D.T. 2012. A comparison of four sputum pre-extraction preparation methods for identifying and characterising Mycobacterium tuberculosis using GCxGC-TOFMS metabolomics. *J Microbiol Methods*, 91:301–311.

Chapter 4: Results and Discussion

4.1 Results

4.1.1 Method Performance

4.1.1.1 Analytical and extraction precision

Analytical and extraction precision were acceptable, with 95.4% and 83.5% of all compounds detected having a CV-value below 50%, respectively (**Figure 4.1**).

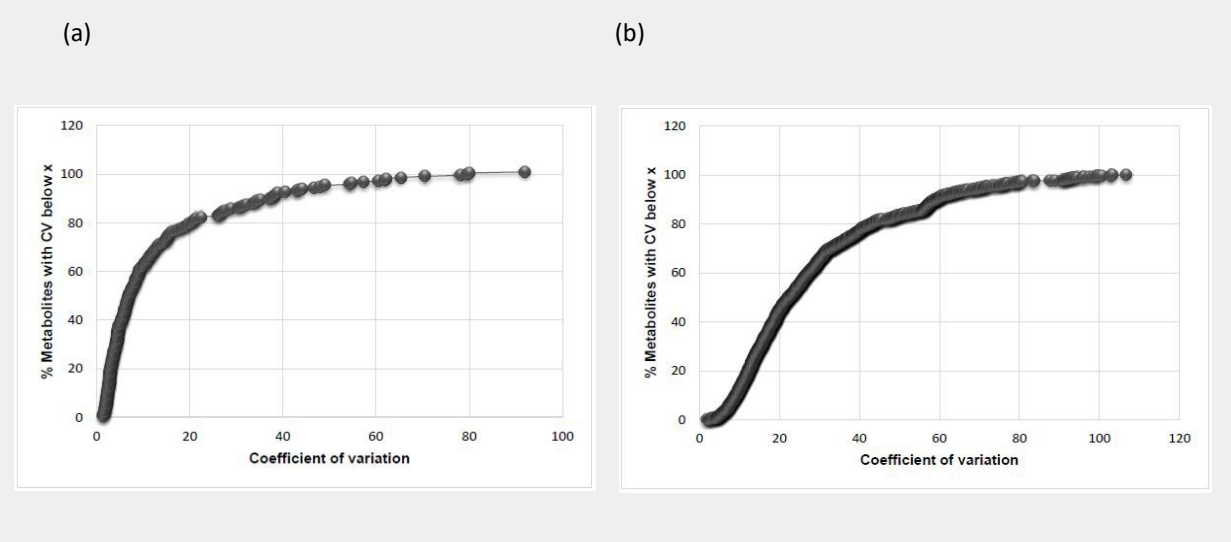


Figure 4.1*: Distribution of the coefficients of variance (CV) values analytical (a) and extraction (b) precision, determined using the areas of all detected compounds.*

4.1.1.2 Batch effect

Batch effect evaluation was performed on 19 quality control (QC) samples (injected at the beginning, middle and end of each batch) using the Combat method (Johnson *et al.* 2007). No prominent batch effect was visible, however, to err on the side of caution, batch correction was performed, but proved inconsequential and was disregarded (**Appendix A Figure S1**). **Figure 4.2** illustrates the total ion chromatograms (TICs) obtained from the pooled QC aliquots, selected from random batches, indicating comparable retention times throughout the analysis.

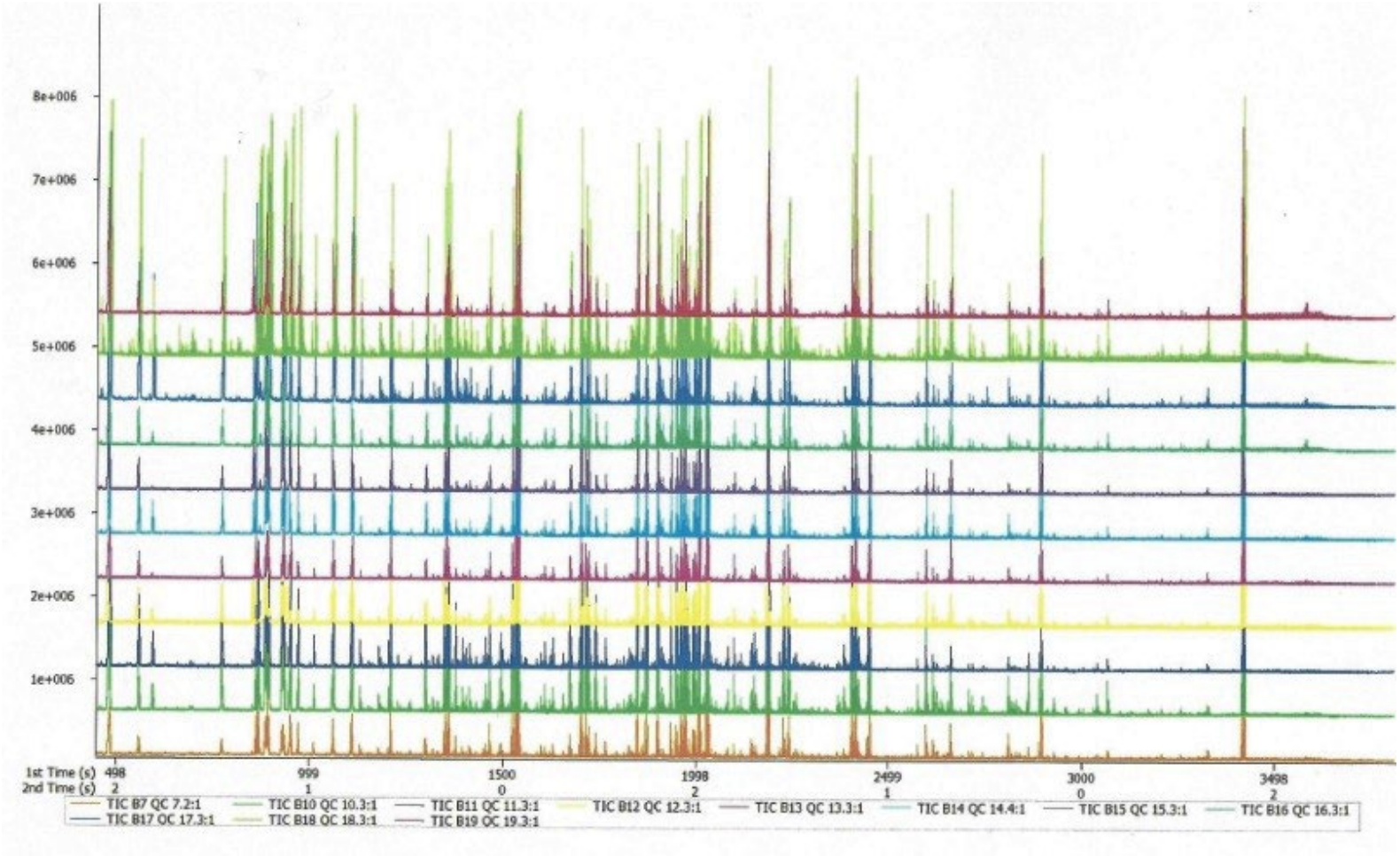


***Figure 4.2**: Total ion chromatograms (TICs) of pooled QC aliquots, analysed across randomly selected batches.*

4.1.2 Population

A total of 228 participants (age 20–60 years), comprising 189 males (24–60 years) and 39 females (20–58 years), were included in the study. The composition of the sample subgroups within the total cohort is stipulated in **Figure 3.1**. Sample sizes and the residual spread between TB+ and TB- for some of the abovementioned subgroups were very small. As previously stated, the samples were originally collected for diagnostic purposes, and, although this cohort is representative of the typical South African TB patient population, the lack of study design in terms of patient recruitment limits the current investigation. Therefore, it should be noted that this is a preliminary study to understand if sex and HIV-status should be considered as confounding factors in future TB metabolomics investigations. Results are not aimed at identifying set diagnostic biomarkers, as is evident from the selection of only a robust ES to describe differences. Expressing metabolic changes in this manner can then guide future power calculations to determine appropriate sample sizes, while identifying differentiating metabolites can guide the choice of future analysis, be it targeted or untargeted.

4.1.3 Metabolic profiles and discussion

In total, 969 compounds were detected, of which 343 could be annotated by spectral comparison to a library of previously injected standards (with a similarity match of at last 80%). Although the remaining 626 compounds could not be annotated, they were still included in statistical analysis, and labelled as "unannotated". Due to the nature of the annotation procedure, some compounds with similar molecular structures and, hence, mass spectra, were given the same label (based on the best library hit). Retention times and unique masses were used to determine if these were indeed two different compounds, and if so, the duplicate name was retained with an index, and concentrations analysed as separate compounds.

Obvious separation between the data obtained from the TB+ and TB- individuals, when including all participants from all subgroups, was not evident (**Figure 4.3**), which is consistent with our previous findings (Du Preez & Loots, 2013). The similarity seen between these larger participant groups is most likely due to the variation in the study population (caused by several contributors such as sex, diet, medication used, etc.), substantiating the need for the evaluation of more homogeneous cohorts.

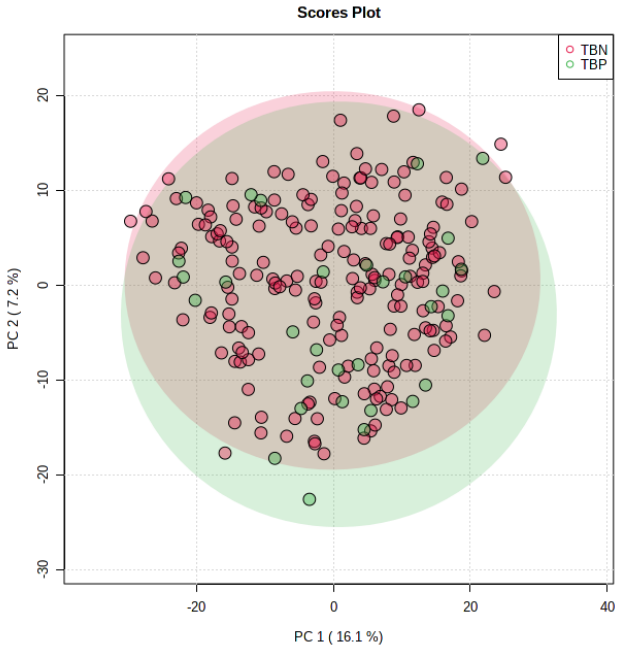


***Figure 4.3**: PCA scores plot for complete dataset (all participants), showing no differentiation between the TB+ (TBP) and TB- (TBN) sample groups.*

Details of the statistically significant compounds identified for each subgroup comparison are given as supplementary info (Appendix A **Table S1** and **Table S2**). A summary of these metabolic variations, according to the compound classes, is given in **Figure 4.4**, with **Table 4.1** stipulating the significant annotated compounds identified in the "all participant" subgroups or the TB/HIV subgroups. For each subgroup comparison (Appendix A **Table S1** and **Table S2**), most of the statistically significant compounds, representing TB-induced perturbations, could not be annotated. When considering the remaining annotated compounds, the dominant classes of compounds identified as important included: lipids, carbohydrates, amino acids, and organic acid-related compounds. Interestingly, the lipids accounted for the largest proportion of TB-induced variation in all subgroups (as high as 30% in the complete dataset, all participants), except in the "all male participant" group dominated by carbohydrates (**Figure 4.4**). Although the main aim of this study was not to identify sex-specific pathway differences, it is interesting to note that Krumsiek *et al.* (2015) previously determined that the entire super-pathway of carbohydrates (specifically including glycolysis, gluconeogenesis, pyruvate, fructose, mannose, galactose, starch, and sucrose metabolisms) was significantly higher in the serum of healthy males compared to females.

When including all participants (Appendix A **Table S1**), 10 compounds were identified as significantly different between the TB+ and TB- groups, of which six could be annotated. In the female subgroup, 74 important compounds (21 annotated) were identified, including mostly lipids and carbohydrates, of which eight overlapped with those identified when including all participants. In the male group, 10 compounds differed significantly between TB+ and TB- participants, and all six of the annotated compounds were also identified in the "all participants" group. Only four compounds (three annotated) were consistently detected as statistically important for differentiating between TB+ and TB- participants whether considering the total cohort, female or male subgroups, these include glycine, 2-hexadecanone, and 1-oleoyl lysophosphatidic acid (LPA) (**Table 4.1**). Although no study to date has identified these compounds as important when characterising TB from sputum, glycine has previously been identified as a potential TB marker in serum (Zhou *et al.*, 2013) and 1-oleoylglycerophosphocholine (lysoPC(p-18:1(9Z)), a direct precursor of 1-oleoyl LPA, was identified as part of a serum biosignature distinguishing TB from other lung diseases (Feng *et al.*, 2015).

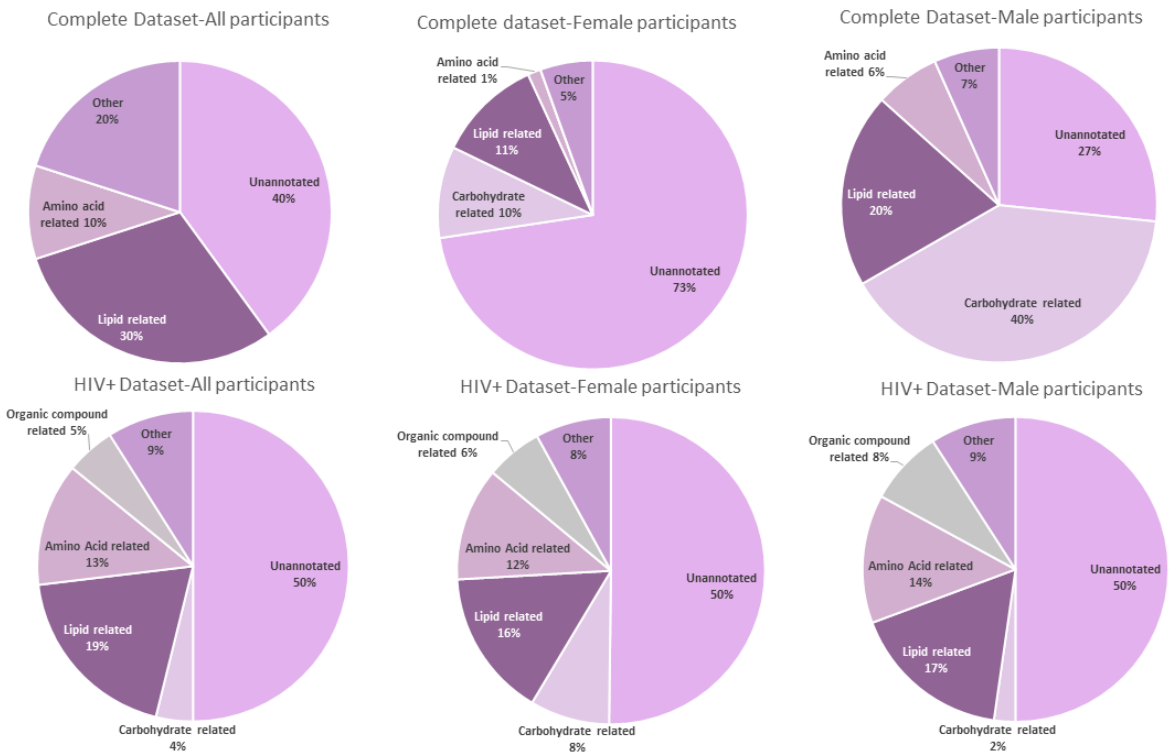


***Figure 4.4**: Representation of the number of compounds per class detected with an ES > 0.64 (as a percentage of the total), for each dataset subgroup.*

In contrast to the relatively low number of varying compounds identified in the total cohort, 391 compounds (of which 121 could be annotated), had an ES > 0.64 when comparing the TB+ and TB- groups, in all HIV+ subgroups (**Table S2**). A total of 104 (39 annotated), mostly lipids and amino acid-related compounds, were identified as statistically important across all participants for the HIV+ cohort (**Table S2**). Furthermore, 337 (125 annotated) and 163 (44 annotated) compounds had an ES > 0.64 in the HIV+ female and HIV+ male groups, respectively. Twenty-five compounds were significant in all TB/HIV+ subgroups (**Table 4.1**), indicating that TB has a more prominent effect on the metabolism in HIV+ individuals, compared to a population where HIV is less prevalent. Although various studies have applied metabolomics as a tool to investigate different aspects of HIV, using diverse sample matrices, metabolomics studies exploring the TB/HIV co-infection are scarce (Liebenberg *et al.*, 2021). In 2019, Silva *et al.* indicated that arachidonic acid and the glycerophospholipid metabolism were altered in the plasma of HIV patients with paradoxical TB-associated immune reconstitution inflammatory syndrome (IRIS), when compared to non-IRIS TB/HIV patients (Silva *et al.*, 2019). In the current study, we also identified both arachidonic acid and 1-oleoyl (LPA), which are both intermediates of the glycerophospholipid metabolism, as significant in all the TB/HIV+ subgroups (**Table 4.1**). Two other studies have identified changes in the tryptophan/kynurenine ratio when comparing HIV+ and HIV- TB patients (Adu-Gyamfi *et al.*, 2017; Collins *et al.*, 2020). Although we detected six amino acid-related compounds characterising TB in the HIV+ subgroup, neither tryptophan nor kynurenine were included in this list.

Interestingly, two of the three annotated compounds identified as important in the “all participants'' group, namely glycine and 2-hexadecanone, were not characteristic of TB when considering only HIV+ subgroups (**Table 4.1**). Nonetheless, two compounds, of which one could be annotated as 1-oleoyl LPA, had an ES > 0.64 across all comparisons irrespective of sex or HIV co-infection status (**Table 4.1**). Although the second compound (hereafter referred to as compound X) detected across all comparisons could not be annotated based on the set criteria of a similarity match > 80%, its mass spectra could be denoted as oleic acid, with a similarity match of 76,4% (Appendix A **Figure S5**). A detailed view of the mean concentrations of compound X in all subgroups is provided as supplementary material (Appendix A **Figure S2**). This compound showed an overall decrease when comparing TB+ to TB- across “all participants”, but an increase in the “all HIV+ participants” TB patients. In addition, compound X was also elevated in the TB patients in both the male subgroups, compared to the TB- cases, but this trend was reversed in both female groups. The 95% confidence

intervals, however, indicate large variation between the individuals within the subgroups, and therefore, no valid conclusions can be drawn from these results without further investigation.

A detailed view of the mean concentrations of the annotated compound, 1-oleoyl LPA, in all subgroups is provided in **Figure 4.5**. These results indicate a significant difference between the TB+ and TB- participants in each subgroup, with decreased concentrations in the TB+ cases, although the male groups had overall higher means.

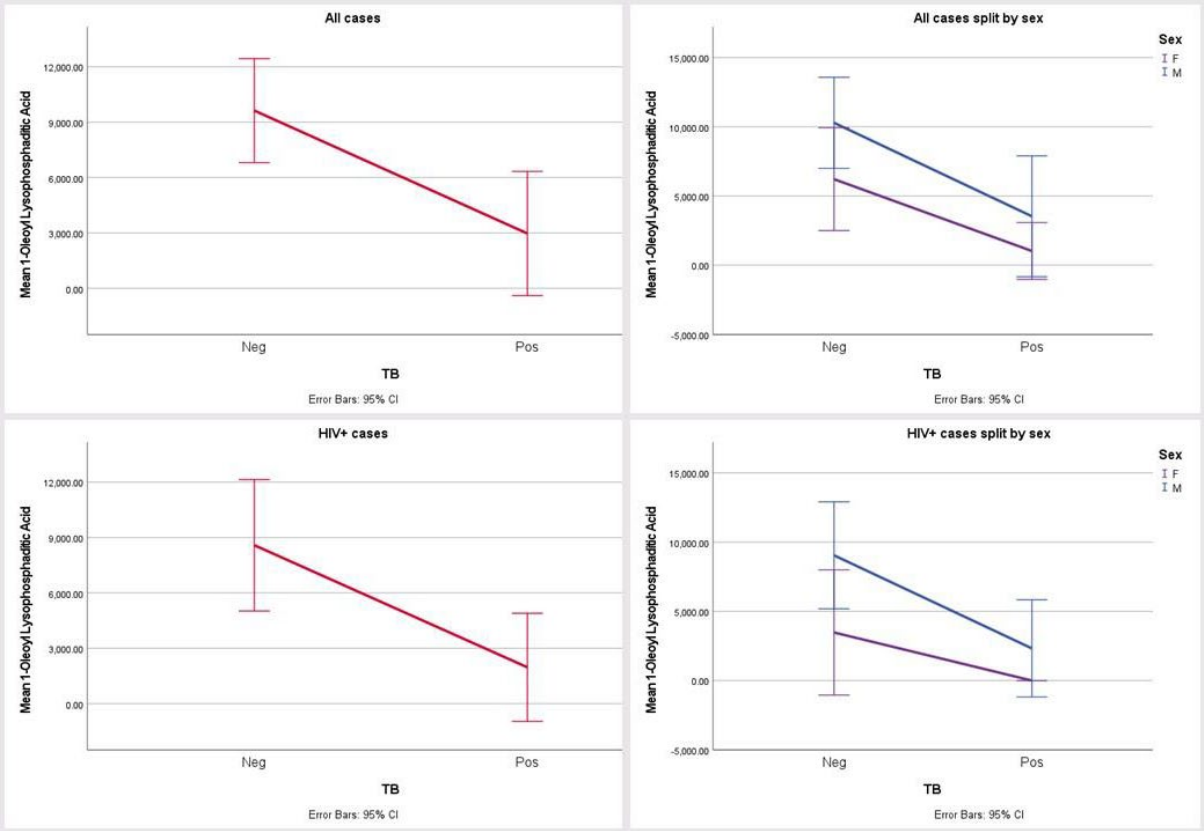


***Figure 4.5**: Mean concentration (with 95% confidence interval displayed as error bars) of the compound, annotated as 1-oleoyl lysophosphatidic acid (LPA), in all participant subgroups.*

It should be noted that the annotation of this compound was based on a comparison of its mass spectrum to those in a spectral library, which showed a similarity match of 82.4% compared to the spectrum of 1-oleoyl LPA (Appendix A **Figure S3**). The compound's mass spectrum was also closely related to that of 1-margaroleoyl LPA acid and 1-palmitoleoyl LPA with similarity matches of 76.5% and 74.0% respectively (Appendix A **Figure S4**). Since, the aim of the method used was to detect

compounds from many compound classes in a single run (untargeted approach) to get an overall indication of the TB-induced variation in the sputum samples, it was not optimised specifically for lipids, and, hence, absolute identification is not possible. We will, therefore, discuss this outcome in terms of LPA in general, irrespective of the fatty acid chain length.

Many previous studies have described changes to the lipid metabolome (lipidome) in TB patients (Han *et al.*, 2021; Chen *et al.*, 2021), including our previous metabolomics investigations done on sputum (Du Preez & Loots, 2013; Schoeman *et al.*, 2012). In addition, it has been shown that, in mice, *M. tuberculosis* uses fatty acids as the main source of carbon (Ghazaei, 2018) and that host lipids are the predominant nutrient sources for these infective bacteria (Han *et al.*, 2021). The lipid class in its entirety is, however, very broad. Our study highlights the potential role of LPA in the pathogenesis of TB and TB/HIV co-infection. LPA is a bioactive phospholipid, made up of a glycerol backbone with a hydroxyl group, a phosphate group, and a fatty acid chain, that is produced during the synthesis of cell membranes and is described as an important extracellular signalling molecule present in all eukaryotic tissues and blood plasma. This compound class also plays an active role in modulating and inducing cell proliferation as well as cell migration, not only during cell development but also in pathological conditions. In some cancers, LPA signalling has been linked to the biological events related to the development of therapy resistance or responses to treatment (Geraldo *et al.*, 2021).

Various studies have specifically linked LPA to lung pathologies. When investigating pulmonary fibrosis in humans, it was shown that LPA-LPA receptor 1 (LPA-LPA1) signalling plays a critical role in the progression of the disease (Tager *et al.*, 2008). Furthermore, during lung injury, increased levels of LPA are produced in bronchoalveolar lavage. The LPA then induces cumulation of fibroblasts and consequent vascular leakage via LPA1, which then advances the progression of fibrosis (Aikawa *et al.*, 2015). When monitoring the plasma lipid levels of TB patients from initial diagnosis to the cured stage, it was concluded that some LPAs might be promising biomarkers for TB, the intervention of lipid metabolism could potentially block energy metabolism and in turn inhibit the cell wall synthesis of *M. tuberculosis* (Chen *et al.*, 2021).

A previous metabolomics study indicated lower abundances of serum amino acids, medium-chain fatty acids and the LPA precursor, lysophosphatidylcholine (LPC) in TB patients compared to healthy participants (Weiner *et al.*, 2012). In addition, they illustrated metabolic profiles with overall decreased phospholipase activity in active TB disease, compared to those individuals with latent

infection. Phospholipase A2 (PLA2) hydrolyses the ester bond at the sn2 position of membrane phospholipids, usually resulting in the release of free fatty acids and lysoglycero-phospolipids, the latter of which are precursors of LPA. PLA2 is essential for inducing inflammation and an immune response (Wesley Burks *et al.*, 2019). Accordingly, in 2020, Han *et al.* applied ultra-high-performance liquid chromatography-tandem mass spectrometry to investigate plasma lipid levels in patients with TB, lung cancer, community-acquired pneumonia, and healthy controls. They found that the plasma phospholipid levels (LPA: bioactive phospholipid) were decreased and speculated that the *M. tuberculosis* infection might be responsible for regulating the lipid metabolism of TB patients by promoting host-assisted bacterial degradation of phospholipids (Han *et al.*, 2021). Lower abundances of LPCs/LPAs in TB patients could mechanistically be related to *M. tuberculosis*-induced macrophage apoptosis by inhibition of PLA2 (Duan *et al.*, 2001)

Interestingly, a significant difference in plasma LPA concentration attributing to sex and age was identified amongst 100 healthy individuals (Michalczyk *et al.*, 2017). However, when investigating sex as a confounding variable in the sputum metabolomes of lung cancer patients, no significant effect was observed, variation in LPA (possibly) overshadows the effect of sex (O'Shea *et al.*, 2016), however, this phenomenon would require further investigation.

***Table 4.1**: Annotated compounds with ES > 0.64 in either all of the 'all participant' subgroups, or all of the HIV+ subgroups*

		Total cohort						HIV					
		All Participants		Female		Male		All Participants		Female		Male	
Class	**Compound***	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**
Carbohydrate related	Glucose (UM:204)	-	-	-	-	-	-	0.66	↓	0.68	↓	0.67	↓
Carbohydrate related	Inosine (UM:230)	-	-	-	-	-	-	0.67	↓	0.86	↓	0.65	↓
Lipid related	1-Palmitoyl Lysophosphatidic acid (UM:299)	-	-	-	-	-	-	0.71	↓	0.86	↓	0.69	↓
Lipid related	**1-Oleoyl Lysophosphatidic acid (UM:357)**	**0.66**	**↓**	**0.73**	**↓**	**0.64**	**↓**	**0.67**	**↓**	**0.86**	**↓**	**0.64**	**↓**
Lipid related	Cholesterol (UM:129)	-	-	-	-	-	-	0.75	↓	0.93	↓	0.74	↓
Lipid related	Eicosane (UM:57)	-	-	-	-	-	-	0.71	↓	0.93	↓	0.68	↓
Lipid related	Heptadecanoic acid (UM:117)	-	-	-	-	-	-	0.66	↓	0.64	↑	0.66	↓
Lipid related	à-Linolenic acid (UM:67)	-	-	-	-	-	-	0.67	↓	0.79	↓	0.65	↓
Lipid related	Arachidonic acid (UM:80)	-	-	-	-	-	-	0.69	↓	0.79	↓	0.67	↓
Lipid related	Oleamide (UM:144)	-	-	-	-	-	-	0.66	↓	0.64	↓	0.66	↓
Lipid related	Palmitaldehyde, dibutyl acetal (UM:57)	-	-	-	-	-	-	0.67	↓	0.79	↓	0.64	↓
Amino acid related	Glycine (UM:102)	0.69	↓	0.68	↓	0.70	↓	-	-	-	-	-	-
Amino Acid related	L-Ornithine (UM:142)	-	-	-	-	-	-	0.65	↓	0.64	↓	0.65	↓
Amino Acid related	L-Valine (UM:144)	-	-	-	-	-	-	0.68	↓	0.79	↓	0.67	↓
Amino Acid related	DL-Ornithine (UM:174)	-	-	-	-	-	-	0.69	↓	0.79	↓	0.66	↓
Amino Acid related	4-Coumaric acid (UM:293)	-	-	-	-	-	-	0.68	↓	0.79	↓	0.65	↓
Amino Acid related	2-Aminomalonic acid (UM:218)	-	-	-	-	-	-	0.67	↓	0.64	↓	0.66	↓
Amino Acid related	N-à-Acetyl-L-Lysine (UM:174)	-	-	-	-	-	-	0.69	↓	0.64	↓	0.71	↓

		Total cohort						HIV					
		All Participants		Female		Male		All Participants		Female		Male	
Class	**Compound***	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**
Organic compound related	Ethanolamine (UM:174)	-	-	-	-	-	-	0.65	↓	0.71	↓	0.65	↓
Organic compound related	Parabanic acid (UM:243)	-	-	-	-	-	-	0.68	↓	0.75	↓	0.67	↓
Organic compound related	2,5-Bis((trimethylsilyl)oxy)pyrazine (UM:241)	-	-	-	-	-	-	0.64	↓	0.79	↓	0.64	↓
Other	Decanamide, N-(2-hydroxyethyl)- (UM:85)	-	-	-	-	-	-	0.68	↓	0.79	↓	0.66	↓
Other	Ethylone (UM:144)	-	-	-	-	-	-	0.66	↓	0.68	↓	0.67	↓
Other	Oxazole, 2-(8Z)-8-heptadecen-1-yl-4,5-dihydro- (UM:98)	-	-	-	-	-	-	0.68	↓	0.71	↓	0.66	↓
Other	2,5-Dimethoxy-4-(n)-propylphenethylamine (UM:174)	-	-	-	-	-	-	0.65	↓	0.68	↑	0.65	↓
Other	3-Chloroamphetamine (UM:116)	-	-	-	-	-	-	0.65	↓	0.71	↓	0.66	↓
Other	2-Hexadecanone (UM:58)	0.69	↓	0.73	↓	0.68	↓	-	-	-	-	-	-

**UM* unique mass

4.2 References

Adu-Gyamfi, C. G., Snyman, T., Hoffmann, C. J., Martinson, N. A., Chaisson, R. E., George, J. A., & Suchard, M. S. 2017. Plasma Indoleamine 2, 3-Dioxygenase, a Biomarker for Tuberculosis in Human Immunodeficiency Virus-Infected Patients. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 65(8), 1356–1358.

Aikawa, S., Hashimoto, T., Kano, K., & Aoki, J. 2015. Lysophosphatidic acid as a lipid mediator with multiple biological actions. *Journal of biochemistry*, 157(2), 81–89.

Chen, J.-X., Han, Y.-S., Zhang, S.-Q., Li, Z.-B., Chen, J., Yi, W.-J., *et al.* 2021. Novel therapeutic evaluation biomarkers of lipid metabolism targets in uncomplicated pulmonary tuberculosis patients. *Signal transduction and targeted therapy*, 6(1), 22.

Collins, J. M., Siddiqa, A., Jones, D. P., Liu, K., Kempker, R. R., Nizam, A., *et al.* 2020. Tryptophan catabolism reflects disease activity in human tuberculosis. *JCI insight*, 5(10), e137131.

Duan, L., Gan, H., Arm, J., & Remold, H. G. 2001. Cytosolic phospholipase A2 participates with TNF-alpha in the induction of apoptosis of human macrophages infected with Mycobacterium tuberculosis H37Ra. *Journal of immunology*, 166(12), 7469–7476.

Du Preez, I. and Loots, D.T. 2013. New sputum metabolite markers implicating adaptations of the host to Mycobacterium tuberculosis, and vice versa. *Tuberculosis*, 93(3): 330–337.

Feng, S., Du, Y.-Q., Zhang, L., Zhang, L., Feng, R.-R., & 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), 159–168.

Ghazaei, C. 2018. and lipids: Insights into molecular mechanisms from persistence to virulence. *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences*, 23, 63.

Geraldo, L. H. M., Spohr, T. C. L. de S., Amaral, R. F. do, Fonseca, A. C. C. da, Garcia, C., Mendes, F. de A., *et al.* 2021. Role of lysophosphatidic acid and its receptors in health and disease: novel therapeutic strategies. *Signal transduction and targeted therapy*, 6(1), 45.

Han, Y.-S., Chen, J.-X., Li, Z.-B., Chen, J., Yi, W.-J., Huang, H., *et al.* 2021. Identification of potential lipid biomarkers for active pulmonary tuberculosis using ultra-high-performance liquid chromatography-tandem mass spectrometry. *Experimental biology and medicine*, 246(4), 387–399.

Johnson, W.E., LI, C. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8(1), 118–127.

Krumsiek J, Mittelstrass K, Do KT, Stückler F, Ried J, Adamski J., *et al.* 2015. Gender-specific pathway differences in the human serum metabolome. *Metabolomics*, 11(6):1815-1833.

Liebenberg, C., Luies, L., & Williams, A. A. 2021. Metabolomics as a Tool to Investigate HIV/TB Co-Infection*. Frontiers in molecular biosciences*, 8, 692823.

Michalczyk, A., Budkowska, M., Dołęgowska, B., Chlubek, D., & Safranow, K. 2017. Lysophosphatidic acid plasma concentrations in healthy subjects: circadian rhythm and associations with demographic, anthropometric and biochemical parameters. *Lipids in health and disease*, 16(1), 140.

O'Shea, K., Cameron, S. J. S., Lewis, K. E., Lu, C., & Mur, L. A. J. 2016. Metabolomic-based biomarker discovery for non-invasive lung cancer screening: A case study. *Biochimica et Biophysica Acta (BBA)* - General Subjects.

Schoeman, J. C., du Preez, I., & Loots, D. T. 2012. A comparison of four sputum pre-extraction preparation methods for identifying and characterising Mycobacterium tuberculosis using GCxGC-TOFMS metabolomics. *Journal of microbiological methods*, 91(2), 301–311.

Silva, C. A. M., Graham, B., Webb, K., Ashton, L. V., Harton, M., Luetkemeyer, A. F., *et al.* 2019. A pilot metabolomics study of tuberculosis immune reconstitution inflammatory syndrome. *International journal of infectious diseases: IJID: official publication of the International Society for Infectious Diseases*, 84, 30–38.

Tager, A. M., LaCamera, P., Shea, B. S., Campanella, G. S., Selman, M., Zhao, Z., *et al.* 2008. The lysophosphatidic acid receptor LPA1 links pulmonary fibrosis to lung injury by mediating fibroblast recruitment and vascular leak. *Nature medicine*, 14(1), 45–54.

Weiner, J., 3rd, Parida, S. K., Maertzdorf, J., Black, G. F., Repsilber, D., Telaar, A., *et al.* 2012. Biomarkers of inflammation, immunosuppression and stress with active disease are revealed by metabolomic profiling of tuberculosis patients. *PloS one*, 7(7), e40221.

Wesley Burks, A., Holgate, S. T., O'Hehir, R. E., Bacharier, L. B., Broide, D. H., Khurana Hershey, G. K., & Stokes Peebles, R., Jr. 2019. Middleton's Allergy E-Book: Principles and Practice. *Elsevier Health Sciences.*

Zhou, A., Ni, J., Xu, Z., Wang, Y., Lu, S., Sha, W., Karakousis, P. C., & Yao, Y. F. 2013. Application of (1)h NMR spectroscopy-based metabolomics to sera of tuberculosis patients. *Journal of proteome research*, 12(10), 4642–4649.

Chapter 5: Conclusions

5.1 Concluding summary

The analysis of sputum is an attractive and effective sample matrix for exploring various airway and lung diseases since it is collected directly from the site of infection. In addition, the use of "sick" controls (TB-, but with similar symptoms) as opposed to healthy controls adds significant value to the outcomes of the study because the identified biomarkers or feature profiles represent a "real life scenario" and have a higher probability of being disease-specific. This approach eliminates the identification of general disease markers, such as those representing inflammation or an immune reaction, for example, thereby increasing the potential future application in diagnostics.

Although sex and HIV status, independently, have been shown to influence metabolite profiles of healthy individuals, no metabolomics studies to date (on any sample matrix) have investigated the effect of these covariates on TB-induced metabolome variations. Here, we indicate that these confounding factors do indeed introduce noteworthy variations in the metabolome's response to infection, and that covariates such as sex should be considered as a potential confounding variable when analysing metabolomics data. This approach could assist in the development of more individualised prognosis, diagnostic, and treatment procedures.

In addition, this study highlights the role that lipid metabolism, in particular LPA, plays in TB pathogenesis. Although LPA was previously identified and described for many lung-related disease states, research on the involvement of this lipid class in TB could be better described since it can shed light on the inner workings of the lipid-based metabolome for the disease which could reveal novel biomarkers for downstream optimization of diagnosis and treatment of TB. Further exploration could potentially lead to a better characterisation of the disease mechanisms and the establishment of metabolite-based diagnostic, disease monitoring, and/or treatment efficacy markers.

There are some limitations regarding this study in that it is representative of a very small cohort, collected from one country. The conclusions and hypotheses generated could therefore be specific to this cohort, and a the study should be validated with a larger-equally distributed TB positive and TB negative patient group, with better described demographics.

5.2 Future prospects

Although this study provided new insights into the pathogenesis of TB, and identified disease-induced perturbations which are not influenced by sex and HIV status, it was exploratory in nature, and paves the way for future investigations. These prospects include:

- The application of this methodology and research approach to a larger, balanced cohort.
- To do similar pilot studies on cohorts collected from various geographical areas in South Africa, Africa and globally.
- To investigate the effect of more potentially confounding factors, such as age, race, chronic medications, or other co-infections, for example, on the TB-induced sputum metabolome.
- The findings of this study also highlight the need for the application of a more targeted lipidomics methodology to thoroughly investigate and understand how the lipid content in sputum varies in patients.
- In accordance with the above, a targeted study investigating LPA should be applied to these samples to confirm the differential markers identified using compound standards. This could aid faster diagnosis and consequent treatment of the disease and provide robust biomarker profiles that can be easily tested for using high-throughput diagnostic methods.
- To repeat the study design using other sample matrices such as urine, blood, or saliva. The latter could be an attractive matrix for the development of point of care devices and to confirm lipid metabolome involvement and how robust it could be as a definitive marker/profile for use in routine diagnostic application.

Overall, this study indicates the importance of well-defined sample cohorts in metabolomics studies. To date, thousands of studies have been done using metabolomics to identify differential metabolites characterising a multitude of diseases. The clinical application of the outcomes of these studies, is, however, scarce, and mostly still limited to the diagnosis of inborn metabolic disorders. We postulate that, by predefining cohorts, and investing potential confounding factors, researchers could perhaps break the cycle and identify metabolic variations (within the lipid metabolome for example) with practical and significant clinical value.

Appendix A: Supplement to Chapter 4

Note: Tables S1, S2 and S3 are included as a supplementary Excel workbook.

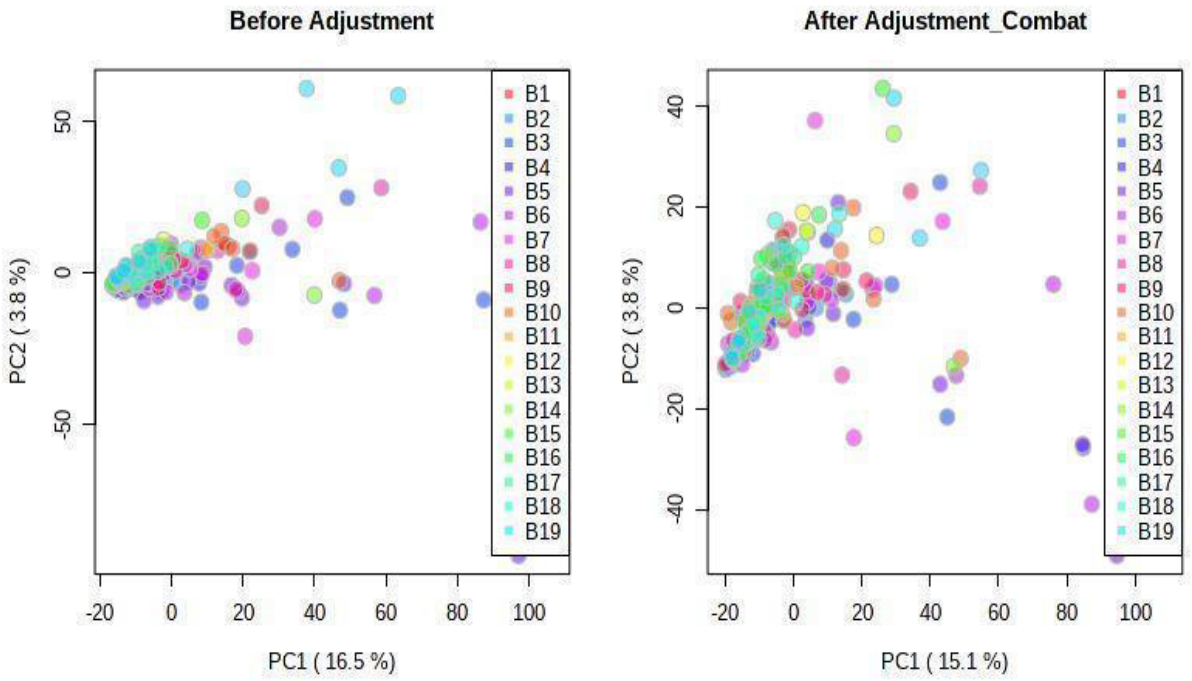


Figure S1*: Principal component analysis (PCA) before and after batch correction using the COMBAT function, indicating no prominent batch effect* (cumulative variation *explained by the first two principal components (PCs) was 20.3 % before and 18.9% thereafter). B = batch.*

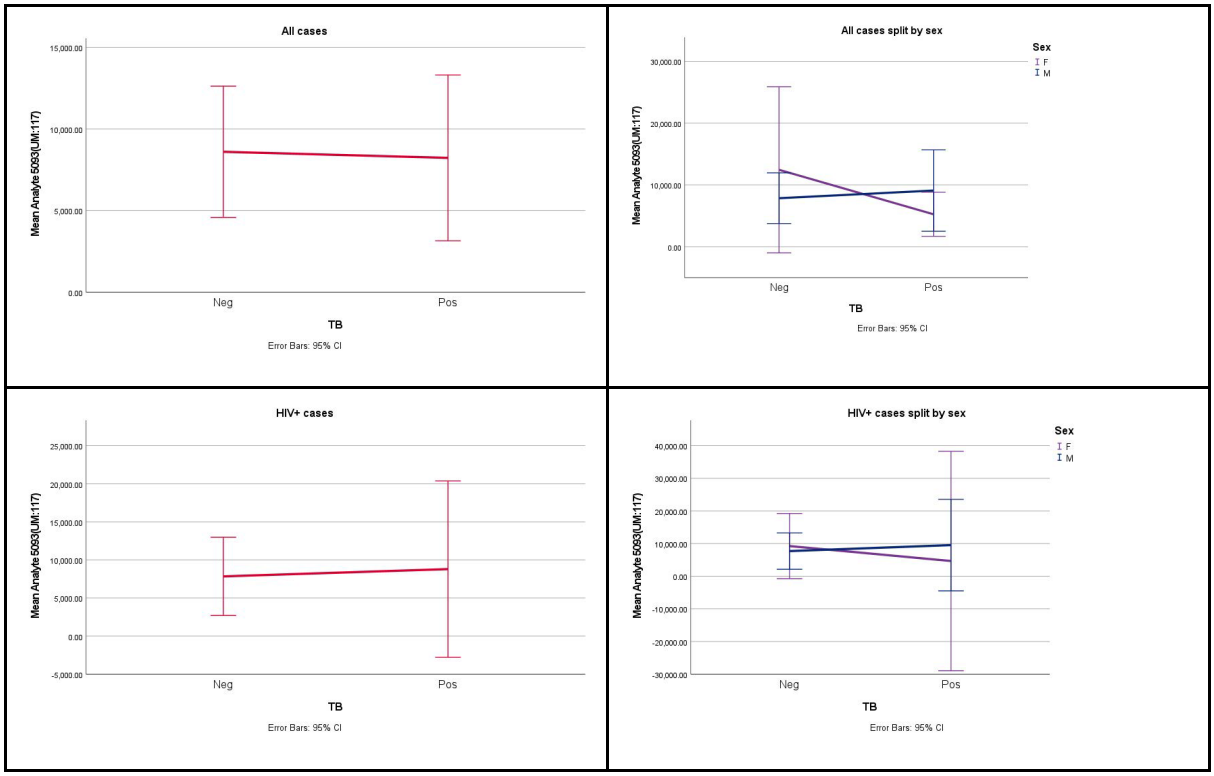


Figure S2*:* Mean concentration (with 95 % confidence interval) of the unannotated compound (compound x) in all participant subgroups.

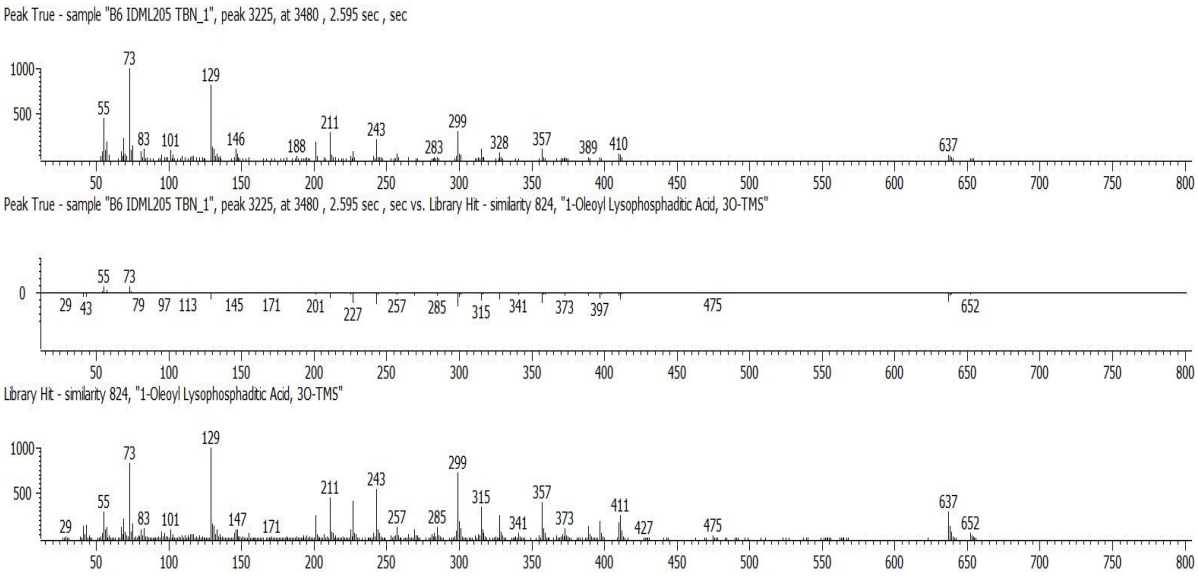


Figure S3: *Mass spectrum of the compound detected as statistically significant in all subgroup comparisons (top), compared to the mass spectrum in the NIST library (bottom), leading to the annotation of the compound as 1-oleoyl lysophosphaditic acid (LPA) based on a 82.4% similarity. The differences between the two mass spectra are given as the middle figure.*

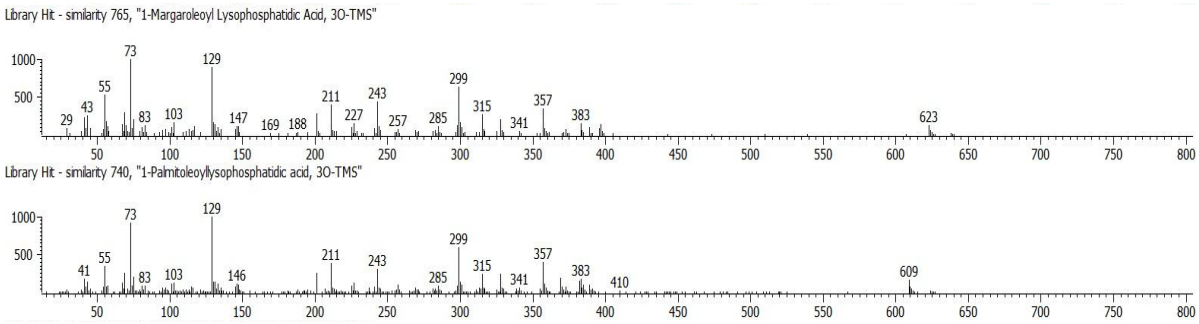


***Figure S4**: NIST library mass spectra of 1-margaroleoyl LPA and 1-palmitoleoyl and LPA.*

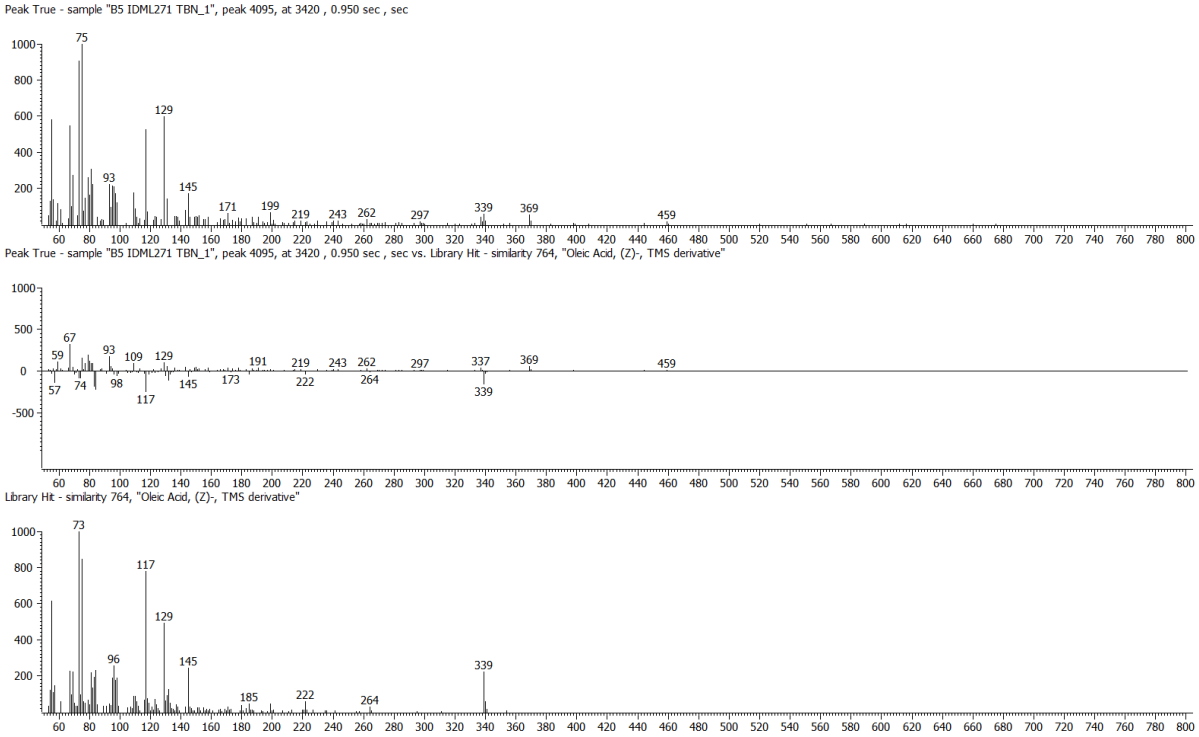
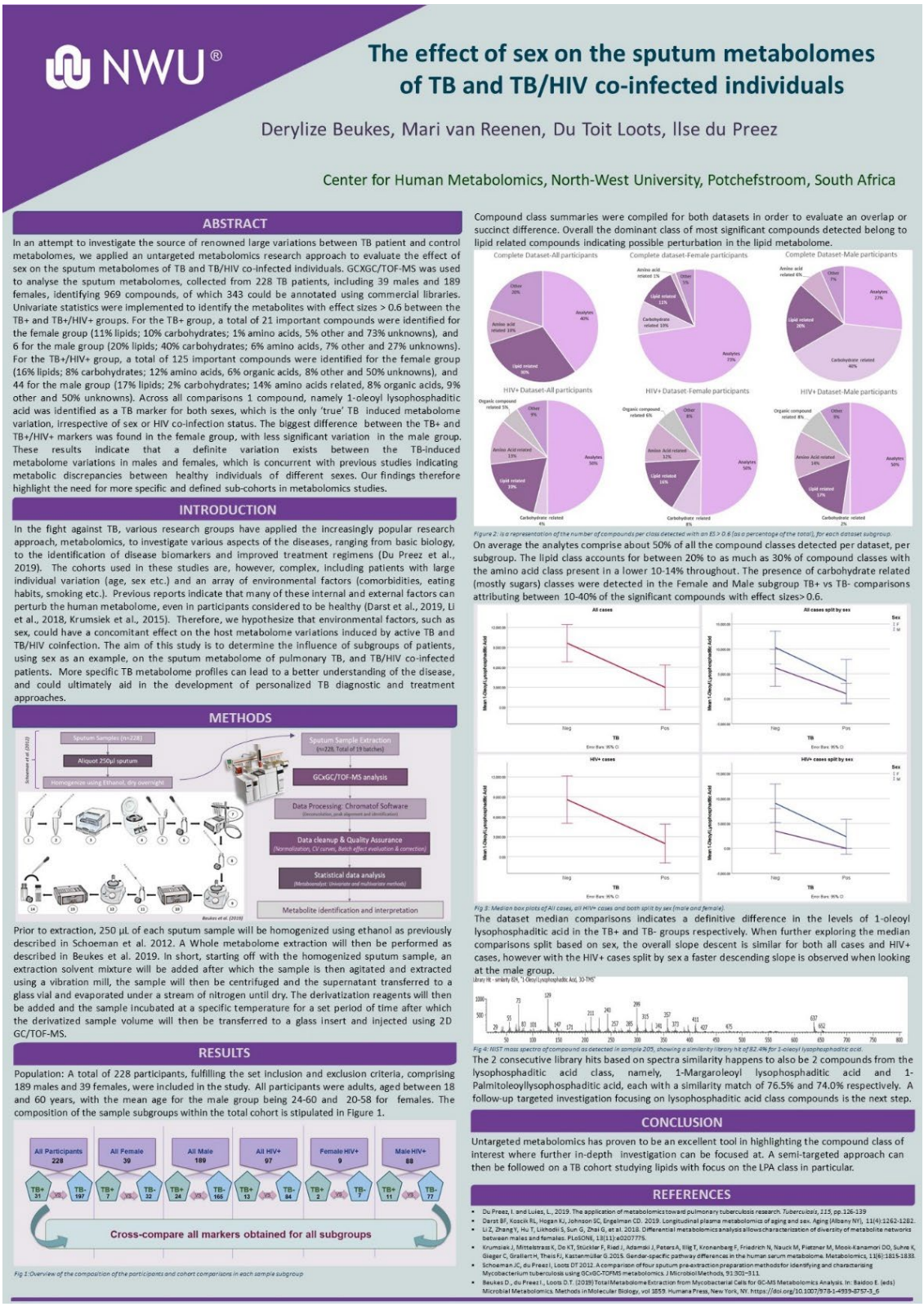


***Figure S5**: Mass spectrum of the unannotated compound detected as statistically significant in all subgroup comparisons (top), compared to the best matched (76,4% similarity) mass spectrum in the NIST library (bottom), leading to the annotation of the compound as oleic acid. The differences between the two mass spectra are given as the middle figure.*

Appendix B: Research outputs

Poster: Beukes, D., Van Reenen, M., Loots DT., Du Preez, I. The effect of sex on the sputum metabolome of TB and TB/HIV co-infected individuals. Presented as poster at the 18th Annual Conference of the Metabolomics Society: Metabolomics 2022: Valencia, Spain, 19-23 June 2022.



Accepted article: Beukes, D., Van Reenen, M., Loots DT., Du Preez, I. Tuberculosis is associated with sputum metabolome variations, irrespective of patient sex or HIV status: An untargeted GCxGC-TOFMS study. Accepted for publication in *Metabolomics* - 10 May 2023.

Tuberculosis is associated with sputum metabolome variations, irrespective of patient sex or HIV status: An untargeted GCxGC-TOFMS study

Abbreviated title: Differential TB sputum metabolome

Derylize Beukes[1], Mari van Reenen[1], Du Toit Loots[1], Ilse du Preez[1]*

[1]Human Metabolomics, North-West University, Potchefstroom, South Africa

***Corresponding author:**
Ilse du Preez
https://orcid.org/0000-0001-5456-3931
Email: ilse.dupreez@nwu.ac.za
Telephone: +27 018 285 2535

Derylize Beukes: 13128531@nwu.ac.za https://orcid.org/0000-0003-1987-4480
Du Toit Loots: dutoit.loots@nwu.ac.za https://orcid.org/0000-0002-0339-6237
Mari van Reenen: 12791733@nwu.ac.za https://orcid.org/0000-0002-5856-3258

Abstract

Introduction Various studies have identified TB-induced metabolome variations. However, in most of these studies, a large degree of variation exists between individual patients.

Objectives To identify differential metabolites for TB, independent of patients' sex or HIV status.

Methods Untargeted GCxGC/TOF-MS analyses were applied to the sputum of 31 TB+ and 197 TB- individuals. Univariate statistics were used to identify metabolites which are significantly different between TB+ and TB- individuals (a) irrespective of HIV status, and (b) with a HIV+ status. Comparisons a and b were repeated for (i) all participants, (ii) males only and (iii) females only.

Results Twenty-one compounds were significantly different between the TB+ and TB- individuals within the female subgroup (11% lipids; 10% carbohydrates; 1% amino acids, 5% other and 73% unannotated), and 6 within the male subgroup (20% lipids; 40% carbohydrates; 6% amino acids, 7% other and 27% unannotated). For the HIV+ patients (TB+ vs TB-), a total of 125 compounds were significant within the female subgroup (16% lipids; 8% carbohydrates; 12% amino acids, 6% organic acids, 8% other and 50% unannotated), and 44 within the male subgroup (17% lipids; 2% carbohydrates; 14% amino acids related, 8% organic acids, 9% other and 50% unannotated). Only one annotated compound, 1-oleoyl lysophosphaditic acid, was consistently identified as a differential metabolite for TB, irrespective of sex or HIV status. The potential clinical application of this compound should be evaluated further.

Conclusions Our findings highlight the importance of considering confounders in metabolomics studies in order to identify unambiguous disease biomarkers.

Keywords tuberculosis; sputum; confounding factors; GCxGC-TOFMS; metabolomics

1 Introduction

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (*Mtb*), still poses a threat, not only to health, but also to socioeconomics, particularly in low- and middle-income countries. An estimated 10 million new cases of TB, 1.2 million deaths among human immunodeficiency virus (HIV)-negative TB patients, and 251 000 deaths in HIV/TB co-infected individuals were reported in 2018 (World Health Organization 2019). Moreover, TB prevalence differs between males and females, with a male to female ratio of 1.6:1 (World Health Organization 2016).

In the fight against TB, metabolomics has become an increasingly popular research approach. This methodology has been used to investigate various aspects of the diseases, from the basic biology to the identification of disease specific biomarkers and improved treatment regimens (Du Preez et al., 2019). In most of these studies, the cohorts are, however, complex due to inter- and intra-individual heterogeneity (e.g., age, sex, and comorbidities). Many of these internal and external factors can perturb the human metabolome, even in participants considered to be healthy. Darst et al. (2019), for example, indicated that 63,4 % (n=695) of the metabolites identified in healthy human plasma correlated to sex. In an earlier study, a clear variation between the phosphatidylcholine and acylcarnitine metabolite classes were identified in males and females (Li et al., 2018). With regards to HIV, Cassol et al. (2013), for instance, could differentiate between the untargeted plasma metabolic profiles of HIV patients on antiretroviral therapy and healthy controls, identifying increased bile acids and acylcarnitines and decreased sulfated steroids, polyunsaturated fatty acids, and lysophosphocholine (LPC) in the patient samples, comparatively.

In accordance, the aim of this study was to identify potential variations in the sputum metabolome of TB patients, when including confounding factors such sex and HIV status. The inclusion of these and other covariances into biomarker studies could enable the identification of more sensitive and specific diagnostic and drug targets. Only one annotated compound, 1-oleoyl lysophosphatidic acid, was consistently identified as characteristic for TB, despite the sex and HIV status of patients. According to our knowledge, metabolomics studies done on sputum are limited, and this is the first attempt to investigate these confounding factors in this matrix. Outcomes therefore also suggest improvements to future metabolomics study designs, not only for TB, but for diseases in general.

2 Materials and methods

2.1 Sample collection

Sputum samples from patients suspected of having TB, based on a medical assessment of the symptoms associated with the disease were sent to a centralized national laboratory, where standard diagnostic procedures, including both Ziehl-Neelsen staining and bacteriological culture were performed, as per

the normal clinical route. After diagnostic testing, the remnant proportion of these samples were frozen (-80 °C) and transported to the North-West University, Centre for Human Metabolomics, for secondary use in research. Samples were received in this manner from March 2009 to May 2010. Anonymity was safeguarded by assigning a unique code to each sample prior to transport. Collected samples were included in the study if they had sufficient volumes for the analyses. Although no additional inclusion or exclusion criteria were considered, some clinical information could be obtained: HIV status (only where the patient requested for an HIV test to be performed), age at time of collection and sex.

2.2 Sample extraction and derivatization

Prior to extraction, 250 µL of each patient sample was homogenized with ethanol (Schoeman et al. 2012) and extracted using a previously described method (Beukes et al., 2019). Samples were randomly analyzed across subgroups, in 19 batches. In short, 50 µL of the internal standard, 3-phenylbutyric acid (0.525 mg mL-1), was added to 250 µL of the homogenised sample. Hereafter, 1 mL of an extraction solvent mixture consisting of chloroform:methanol:water (1:3:1) was added to the tubes (Honeywell International Inc., Muskegon, MI, USA). The extraction was performed using a MM 400 mixer mill (Retsch GmbH & co. KG, Haan, Germany) at a frequency of 30 Hz, for 5 min, after the addition of a 3 mm tungsten carbide bead to each sample tube. Following centrifugation (4 °C at 21 952 x g for 10 min), the supernatant (including the organic and water phases, excluding the pellet) was collected, transferred to a GC-MS sample vial and dried under a light stream of nitrogen. Following this, samples were derivatized using 50 µL of methoxyamine hydrochloride in pyridine (15 mg/mL) (Merck, Darmstadt, Germany) at 50 °C for 90 min, followed by 40 µL of N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) (Sigma-Aldrich, St. Louis, MO, USA) with 1 % TMCS at 60 °C for 60 min. The extracts were transferred to a 0.1 mL insert in a sample vial and capped prior to untargeted GCxGC-TOFMS analysis.

2.3 GCxGC-TOFMS analysis

One microliter of each sample extract was injected (1:5 split ratio) onto a Pegasus 4D GCxGC-TOFMS (Leco Corporation, St. Joseph, MI, USA), which comprises an Agilent 7890A GC (Agilent, Atlanta, GA) coupled to a time of flight mass spectrometer (Leco Corporation, St. Joseph, MI, USA) equipped with a Gerstel Multipurpose Sampler (Gerstel GmbH & co. KG, Eberhard-Gerstel- Platz 1, D-45473 Mülheim an der Ruhr). First dimensional separation was achieved with a Rxi-5Sil MS primary column (29.245 m, 0.25 mm internal diameter, 0.25 µm film thickness) (Restch GmbH & co. KG, Haan, Germany) and a Rxi-17 capillary column (1.400 m, 0.1 mm internal diameter, 0.1 µm film thickness) was fitted as the secondary column (Restch GmbH & co. KG, Haan, Germany). The front inlet

temperature was held at a constant 270 °C for the entire run, ensuring rapid vaporization. For the primary oven, an initial GC oven temperature was set at 70 °C for 2 min followed by an initial increase in oven temperature of 4 °C/min to a final temperature of 300 °C, which was held for 2 min. The secondary column oven temperature was set at 85 °C for 2 min, then increased by 4 °C/min, until a final temperature of 300 °C, at which it was maintained for a further 2 min. The initial temperature of the modulator was 100 °C for 2 min, followed by a 4 °C/min increase to a final temperature of 310 °C held for 9 min. To control the effluent from the primary onto the secondary column, cryomodulation and a hot pulse of nitrogen gas of 0.5 s, every 3 s was used. The acquisition delay for each run was 450 s and the transfer line temperature was held at a constant 270 °C, with the ion source temperature at a constant 200 °C. The detector voltage was adjusted to 1500 V with a filament bias of - 70 eV. Spectra were collected in scan mode from 50 - 800 m/z at an acquisition rate of 200 spectra per second. Note: this was an untargeted approach, with the general aim of separating and detecting as many compounds as possible within a single run (Schoeman et al., 2012, Du Preez & Loots, 2013)

Mass spectral deconvolution, peak alignment and peak identification was performed using Leco Corporation's ChromaTOF software (version 4.51). Mass spectral deconvolution was performed at a signal-to-noise ratio of 100, with a minimum of three apexing peaks. To eliminate the effect of retention time shifts and to create a data matrix containing the relative abundance of all compounds present in all samples, peaks with similar mass spectra and retention times were aligned using Statistical Compare, a package of ChromaTOF. Mass fragmentation patterns and their respective retention times were screened against commercially available National Institute of Standards and Technology (NIST) spectral libraries (mainlib, replib) for peak annotation, with a similarity setting of at least 80%.

2.4 Statistical data analysis

Statistical analyses were performed using MetaboAnalyst (version 5.0) (Pang et al., 2021) and R (R Core Team, 2021). Data were normalized to the sample median, log transformed and auto scaled prior to processing. First, the presence and severity of any batch effect was evaluated, and a summarizing view of the data generated using principal component analyses (PCA). For this purpose, data from all patient samples analysed, grouped according to the batch they were analysed in, were used.

Next, to evaluate the confounding effect of the available clinical information, the processed data were analysed by assigning the participant data to six subgroups, including: all participants (irrespective of HIV status), all female participants, all male participants, all HIV+ participants, HIV+ female participants, and HIV+ male participants (Figure 1). Based on the classification, individual samples could be added to more than one subgroup for statistical analysis.

Considering the limitation of the study design (i.e. the study was not purposefully designed to assess sex as a main factor), we chose to focus solely on effect sizes as a measure to identify differential TB metabolites. The usual multivariate metabolomics statistical approach is considered inappropriate for small sample sizes. In the univariate setting, the reporting of effect sizes along with p-values from hypothesis testing, has become the new standard. Considering that any findings made here will require further investigation, a nonparametric effect size was used to assess the role of sex in more diagnostic terms. The estimated effect sizes reported here will be invaluable for designing future studies. P-values on the other hand, will be of little value as the current sample size does not allow for the generalization of findings, especially after performing the required adjustments for multiple testing. The probability of superiority (*Aw*) effect size measure was selected due to the qualities it possesses. It is a common language measure which improves the accessibility of the information. That is, the interpretation of the measure is straightforward, it is the probability that a case randomly selected from group 1 will have a higher value than a case randomly selected from group 2. *Aw* makes no assumptions regarding normality or homogeneity of variances and is robust to unequal sample sizes. The R package RProbSup (https://cran.r-project.org/package=RProbSup) was used to compute *Aw* given default settings. Compounds with an *Aw* value exceeding 0.64 between TB+ and TB- participants, within each of the subgroups, were shortlisted as informative. An *Aw* value of 0.64 corresponds to a moderate effect. For more information on *Aw*, please refer to Li (2016).

The outputs of the subgroup comparison were then cross-compared to identify those compounds who were identified as differential for TB in all instances. These metabolites were recognized as being characteristic of TB, irrespective of sex or HIV status.

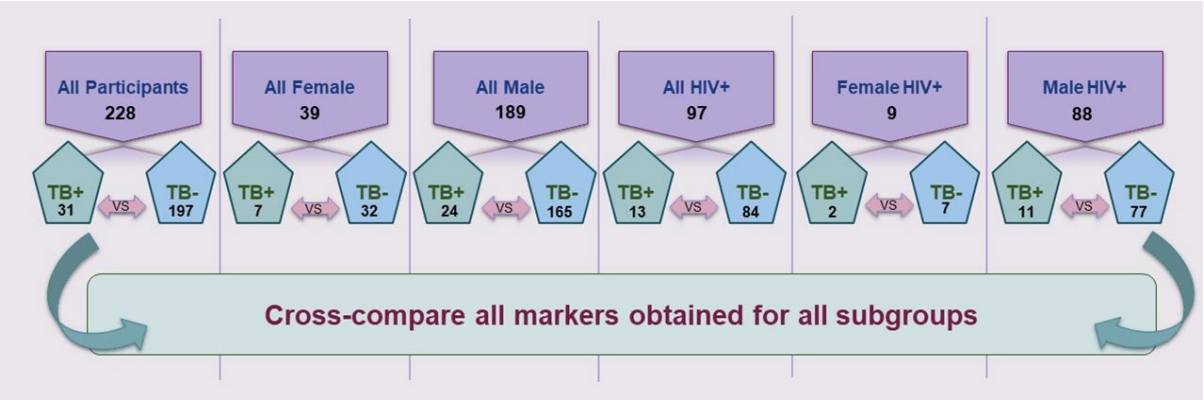


Fig. 1 Overview of the composition of the participants in each sample subgroup. Numerical values indicate the total number of participants per subgroup, and double-headed arrows indicate the cohort comparisons used in the statistical analyses.

3 Results and discussion

3.1 Population

A total of 228 participants, comprising 189 males and 39 females, were included in the study. All participants were adults, aged between 20 and 60 years, with the age range for the male group being 24-60 and 20-58 years for females. The composition of the sample subgroups within the total cohort is stipulated in Figure 1. Sample sizes (especially the female HIV+ group) and the residual spread between TB+ and TB- for some of the abovementioned subgroups were very small.. As previously stated, the samples were originally collected for diagnostic purposes, and, although this cohort is representative of the typical South African TB patient population, the lack of study design in terms of patient recruitment limits the current investigation. Therefore, it should be noted that this is a preliminary study to understand if sex and HIV-status should be considered as confounding factors in future TB metabolomics investigations. The results are not aimed at identifying set diagnostic biomarkers, as is evident from the selection of only a robust ES to identify differentiating compounds. Expressing metabolic changes in this manner can guide future power calculations to determine appropriate sample sizes, while the identified differentiating metabolites are hypothesis generating, and can be used to guide the choice of future, more targeted analysis. The latter should include MRM-based compound identification and absolute quantification using isotopes and calibrators.

3.2 Overview of the data

In total, 969 features were detected, of which 343 could be annotated by spectral comparison to a commercial library compiled of previously injected standards (with a similarity match of at least 80%). Although the remaining 626 compounds could not be annotated, they were still included in all statistical analyses, and classified as 'unannotated'. Due to the nature of the classification procedure, some compounds with similar molecular structures and, hence, mass spectra, were given the same annotation (based on the best library hit). Retention times and unique masses were used to determine if these were indeed two different compounds, and if so, the duplicate name was retained with a code, and concentrations analysed as separate compounds. No prominent batch effect or significant gain difference after batch correction was observed (Supplementary info: Figure S1). The dataset was therefore analysed further without applying a batch effect correction procedure.

Separation between the data obtained from the TB+ and TB- individuals, when including all participants, was not evident (Figure 2). This is consistent with our previous findings (Du Preez & Loots, 2013), and can potentially be ascribed to the inter-individual variation in the study population (sex, comorbidities, etc.).

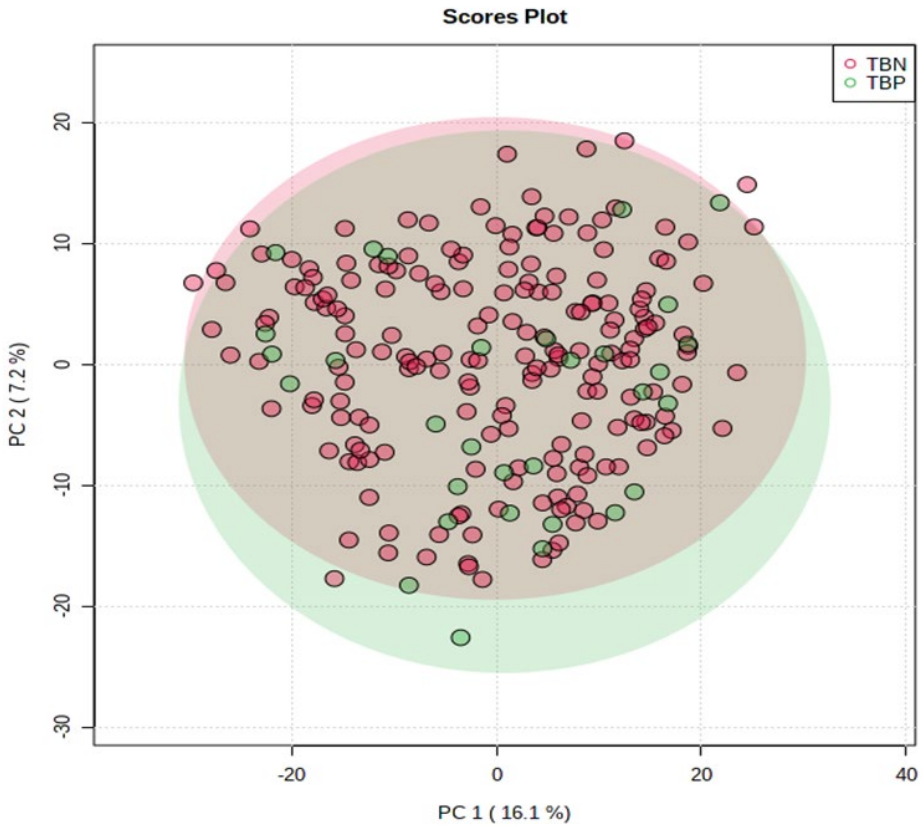


Fig 2 PCA scores plot for complete dataset (all participants), showing no differentiation between the TB+ (TBP) and TB- (TBN) sample groups.

3.3 Subgroup comparisons

Details of the statistically significant compounds identified for each subgroup comparison are given in the supplementary info (Table S1 and Table S2). A summary of these metabolic variations, according to the compound classes, is given in Figure 3. Table 1 stipulates the annotated differential compounds which were either identified in all of the 'all participant' subgroups or all of the TB/HIV+ subgroups. Most differential compounds in all subgroups were lipids, except in the 'all male participant' group, where carbohydrates dominated (Figure 3). Although the main aim of this study was not to identify gender-specific pathway differences, it is interesting to note that Krumsiek et al. (2015) previously determined that the entire super-pathway of carbohydrates (specifically including glycolysis, gluconeogenesis, pyruvate, fructose, mannose, galactose, starch, and sucrose metabolisms) were significantly higher in the serum of healthy males, compared to females.

3.3.1 Differential TB compounds: all male and female participants

When including all participants (Table S1), 10 compounds were identified as significantly different when comparing the TB+ and TB- groups, of which 6 could be annotated. In the female subgroup, 74

significant compounds (21 annotated) were identified, including mostly lipids and carbohydrates, of which 8 were also differential when including all participants. In the male group, 10 compounds were significantly different between the TB+ and TB- participants, and all 6 of the annotated compounds were also differential in the 'all participants' group. Only 4 compounds (3 annotated) were constantly detected as differential when considering all three 'all participant' subgroups, including glycine, 2-hexadecanone, and 1-oleoyl lysophosphatidic acid (LPA) (Table 1). Although these compounds has not yet been identified as sputum TB markers, glycine has been detected as a marker in serum (Zhou et al., 2013) and 1-oleoylglycerophosphocholine (lysoPC(p-18:1(9Z)), a direct precursor of 1-oleoyl LPA, was identified as part of a TB serum biosignature (Feng et al., 2015).

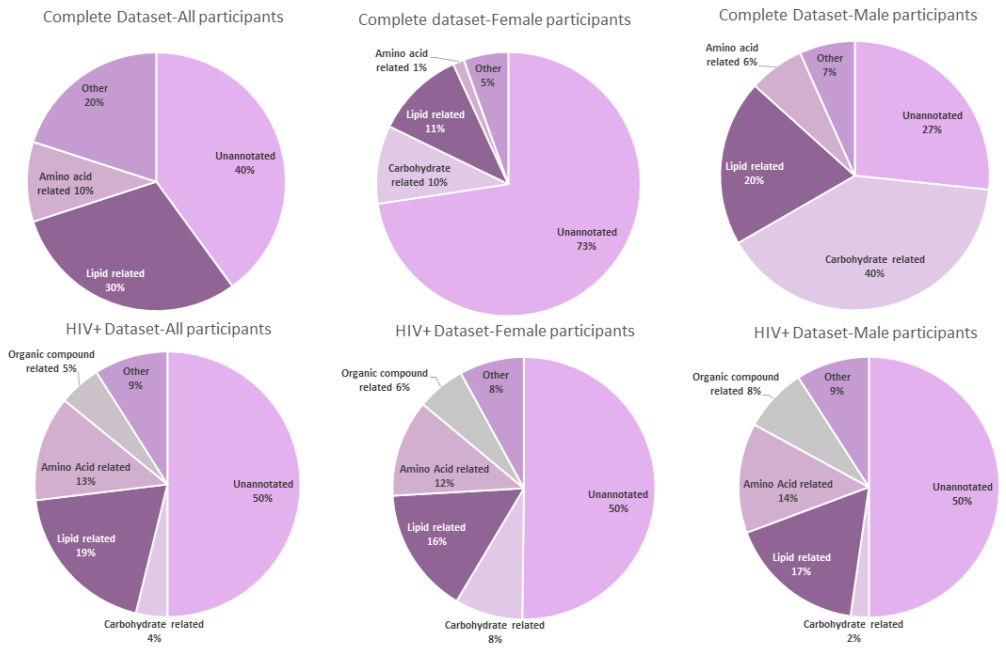


Fig. 3 Representation of the number of compounds per class detected with an ES > 0.64 (as a percentage of the total), for each subgroup comparison of TB+ vs TB-.

3.3.2 Differential TB compounds: HIV+ male and female participants

A total of 104 (39 annotated) compounds, mostly lipids and amino acid related, were identified as differential when including all HIV+ participants (Table S2). Furthermore, 337 (125 annotated) and 163 (44 annotated) compounds were differential in the HIV+ female and HIV+ male subgroups, respectively. Twenty-five compounds were significant when cross-comparing differential compounds identified for the three HIV+ subgroups (Table 1), indicating that TB has a more prominent effect on the metabolism in HIV+ individuals, compared to a population where HIV is less frequent ('all participant' groups). Interestingly, two metabolites, including heptadecanoic acid and 2,5-dimethoxy-4-(n)-propylphenethylamine, were identified as differential for TB in both the female and male HIV+

groups, but with opposite trends. Since the variations between sexes were not the main aim of this study, this was not investigated further.

Although various studies have applied metabolomics as a tool to investigate different aspects of HIV, using diverse sample matrices, metabolomics studies exploring the TB/HIV co-infection are scarce (Liebenberg et al., 2021). In 2019, Silva et al., indicated that arachidonic acid and the glycerophospholipid metabolism were altered in the plasma of HIV patients with paradoxical TB-associated immune reconstitution inflammatory syndrome (IRIS), when compared to non-IRIS TB/HIV patients (Silva et al., 2019). In the current study, we also identified both arachidonic acid and 1-oleoyl LPA, which is an intermediate of the glycerophospholipid metabolism, as differential in all HIV+ subgroups (Table 1). Two other studies have identified changes in the tryptophan/kynurenine ratio when comparing HIV+ and HIV- TB-patients (Adu-Gyamfi et al., 2017; Collins et al., 2020). Although we detected six amino acid related compounds characterising TB in the HIV+ subgroup, neither tryptophan nor kynurenine were included in this list.

3.3.2 Cross-comparison of all markers identified for all subgroups

Two compounds, of which one could be annotated as 1-oleoyl LPA, were identified across all subgroup comparisons (Table 1). Although the second compound (hereafter referred to as compound X) could not be annotated based on the set criteria of a similarity match > 80%, its mass spectra could be denoted as oleic acid, with a similarity match of 76,4% (Supplementary material Figure S4). Boxplots of compound X in all subgroups are provided in the supplementary material (Figure S2). This compound showed an overall decrease in the TB+ cases compared to TB- individuals when including all participants, but an increase in the HIV+ patients. In addition, compound X was higher in the TB patients in both the male subgroups, compared to the TB- cases, but this trend was reversed in both female groups. The standard deviations do, however, indicate large variation between the individuals within the subgroups, and therefore, no valid conclusions can be made from these results without further investigation.

Boxplots of 1-oleoyl LPA, in all subgroups, are provided in Figure 4. These results indicate a significant difference between the TB+ and TB- participants in each subgroup. It should be noted that the annotation of this compound was based on a comparison of its mass spectrum to those in a spectral library, which showed a similarity match of 82.4% compared to the spectrum of 1-oleoyl LPA (supplementary material Figure S3). The compound's mass spectrum was also closely related to that of 1-margaroleoyl LPA acid and 1-palmitoleoyl LPA with similarity matches of 76.5% and 74.0% respectively (supplementary material Figure S4). We will therefore discuss this outcome in terms of LPA in general, irrespective of the fatty acid chain length.

Previous studies have described changes to the lipid metabolome (lipidome) in TB patients (Han et al. 2021;Chen et al. 2021), including our own on sputum (Du Preez and Loots 2013; Schoeman et al. 2012).

In mice, *M.tb* uses fatty acids as the main source of carbon (Ghazaei, 2018) and host lipids are the predominant nutrient source for these infective bacteria (Han et al., 2021). Our current study highlights the potential role of LPA in the pathogenesis of TB and TB/HIV co-infection. LPA is a bioactive phospholipid, consisting of a glycerol backbone with a hydroxyl group, a phosphate group, and a fatty acid chain, produced during the synthesis of cell membranes. LPA is described as an important extracellular signalling molecule present in all eukaryotic tissues and blood plasma. These compounds also play an active role in modulating and inducing cell proliferation and migration, not only during cell development but also in pathological conditions. In some cancers, LPA signalling has been linked to the biological events triggering the development of therapy resistance or responses to treatment (Geraldo et al., 2021).

Various studies have specifically linked LPA to lung pathologies. When investigating pulmonary fibrosis in humans, it was shown that LPA-LPA receptor 1 (LPA-LPA1) signalling plays a critical role in the progression of the disease (Tager et al., 2008). Furthermore, during lung injury, increased levels of LPA are produced in bronchoalveolar lavage. The LPA in turn induces the accumulation of fibroblasts and vascular leakage via LPA_1, advancing the progression of fibrosis (Aikawa et al., 2015). A study monitoring the plasma lipid levels of TB patients from initial diagnosis until cured, indicated that some LPAs show promise as biomarkers for TB. They added that the intervention of lipid metabolism could potentially block energy metabolism and in turn inhibit the cell wall synthesis of *Mtb* (Chen et al., 2021).

A metabolomics study by Weiner et al. (2012) compared TB patients to healthy participants. They found that serum amino acids, medium-chain fatty acids, and the LPA precursor, LPC, were in lower abundance in the patients, comparatively. In addition, the metabolic profile of active TB disease, compared to latent infection, showed decreased phospholipase activity. Phospholipase A_2 (PLA_2) hydrolyses the ester bond at the sn2 position of membrane phospholipids, usually resulting in the release of free fatty acids and lysoglycero-phospolipids, the latter of which are precursors of LPA. PLA_2 is essential for inducing inflammation and plays a key role in the immune response (Wesley Burks et al., 2019). Accordingly, in 2020, Han et al. (2021) applied ultra-high-performance liquid chromatography-tandem mass spectrometry to investigate plasma lipid levels in patients with TB, lung cancer, community-acquired pneumonia, and healthy controls. The study found decreased plasma phospholipid levels (LPA: bioactive phospholipid) and speculated that *Mtb* infection might be responsible for regulating the lipid metabolism of TB patients by promoting host-assisted bacterial degradation of phospholipids (Han et al., 2021). Lower abundances of LPCs/LPAs in TB patients could mechanistically be related to the induction of macrophage apoptosis by *M. tuberculosis* through inhibition of PLA_2 (Duan et al., 2001)

Coincidently, a significant difference in plasma LPA concentration attributing to sex and age was identified amongst 100 healthy individuals (Michalczyk et al., 2017). However, when investigating sex as a confounding variable in the sputum metabolomes of lung cancer patients, no significant effect was observed, indicating that the lung pathology-induced variation in LPA overshadows the effect of sex (O'Shea et al., 2016).

The findings of this study highlight the need for an in depth targeted lipidomics study to thoroughly investigate and understand how lipids, and LPA, contributes to the pathology of TB. The identification of specific LPAs, such as 1-oleolyl LPA in sputum, additionally presents an opportunity for potential biomarker discovery or profiling for diagnosing TB in males or females with or without HIV co-infection. The exact identity of this compound should, however, be confirmed using compound standards and more sensitive MS methods such as MS/MS and ion mobility. In addition, the absolute quantification using isotopes and calibration curves, in a larger cohort, recruited from different endemic countries, would be essential in the validation process.

Table 1 Annotated compounds with ES > 0.64 in either all the 'all participant' subgroups, or all the HIV+ subgroups

		Total cohort						HIV+					
		All Participants		Female		Male		All Participants		Female		Male	
Class	**Compound***	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**	**Effect size**	**Decreased (↓) or increased (↑) in TB+ group**
Carbohydrate related	Glucose (UM:204)	-	-	-	-	-	-	0.66	↓	0.68	↓	0.67	↓
Carbohydrate related	Inosine (UM:230)	-	-	-	-	-	-	0.67	↓	0.86	↓	0.65	↓
Lipid related	1-Palmitoyllysophosphatidic acid (UM:299)	-	-	-	-	-	-	0.71	↓	0.86	↓	0.69	↓
Lipid related	**1-Oleoyl Lysophosphaditic acid (UM:357)**	**0.66**	**↓**	**0.73**	**↓**	**0.64**	**↓**	**0.67**	**↓**	**0.86**	**↓**	**0.64**	**↓**
Lipid related	Cholesterol (UM:129)	-	-	-	-	-	-	0.75	↓	0.93	↓	0.74	↓
Lipid related	Eicosane (UM:57)	-	-	-	-	-	-	0.71	↓	0.93	↓	0.68	↓
Lipid related	Heptadecanoic acid (UM:117)	-	-	-	-	-	-	0.66	↓	0.64	↑	0.66	↓
Lipid related	à-Linolenic acid (UM:67)	-	-	-	-	-	-	0.67	↓	0.79	↓	0.65	↓
Lipid related	Arachidonic acid (UM:80)	-	-	-	-	-	-	0.69	↓	0.79	↓	0.67	↓
Lipid related	Oleamide (UM:144)	-	-	-	-	-	-	0.66	↓	0.64	↓	0.66	↓
Lipid related	Palmitaldehyde, dibutyl acetal (UM:57)	-	-	-	-	-	-	0.67	↓	0.79	↓	0.64	↓
Amino acid related	Glycine (UM:102)	0.69	↓	0.68	↓	0.70	↓	-	-	-	-	-	-
Amino Acid related	L-Ornithine (UM:142)	-	-	-	-	-	-	0.65	↓	0.64	↓	0.65	↓
Amino Acid related	L-Valine (UM:144)	-	-	-	-	-	-	0.68	↓	0.79	↓	0.67	↓

Amino Acid related	DL-Ornithine (UM:174)	-	-	-	-	-	-	0.69	↓	0.79	↓	0.66	↓
Amino Acid related	4-Coumaric acid (UM:293)	-	-	-	-	-	-	0.68	↓	0.79	↓	0.65	↓
Amino Acid related	2-Aminomalonic acid (UM:218)	-	-	-	-	-	-	0.67	↓	0.64	↓	0.66	↓
Amino Acid related	N-à-Acetyl-L-Lysine (UM:174)	-	-	-	-	-	-	0.69	↓	0.64	↓	0.71	↓
Organic compound related	Ethanolamine (UM:174)	-	-	-	-	-	-	0.65	↓	0.71	↓	0.65	↓
Organic compound related	Parabanic acid (UM:243)	-	-	-	-	-	-	0.68	↓	0.75	↓	0.67	↓
Organic compound related	2,5-Bis((trimethylsilyl)oxy)pyrazine (UM:241)	-	-	-	-	-	-	0.64	↓	0.79	↓	0.64	↓
Other	Decanamide, N-(2-hydroxyethyl)- (UM:85)	-	-	-	-	-	-	0.68	↓	0.79	↓	0.66	↓
Other	Ethylone (UM:144)	-	-	-	-	-	-	0.66	↓	0.68	↓	0.67	↓
Other	Oxazole, 2-(8Z)-8-heptadecen-1-yl-4,5-dihydro- (UM:98)	-	-	-	-	-	-	0.68	↓	0.71	↓	0.66	↓
Other	2,5-Dimethoxy-4-(n)-propylphenethylamine (UM:174)	-	-	-	-	-	-	0.65	↓	0.68	↑	0.65	↓
Other	3-Chloroamphetamine (UM:116)	-	-	-	-	-	-	0.65	↓	0.71	↓	0.66	↓
Other	2-Hexadecanone (UM:58)	0.69	↓	0.73	↓	0.68	↓	-	-	-	-	-	-

UM unique mass

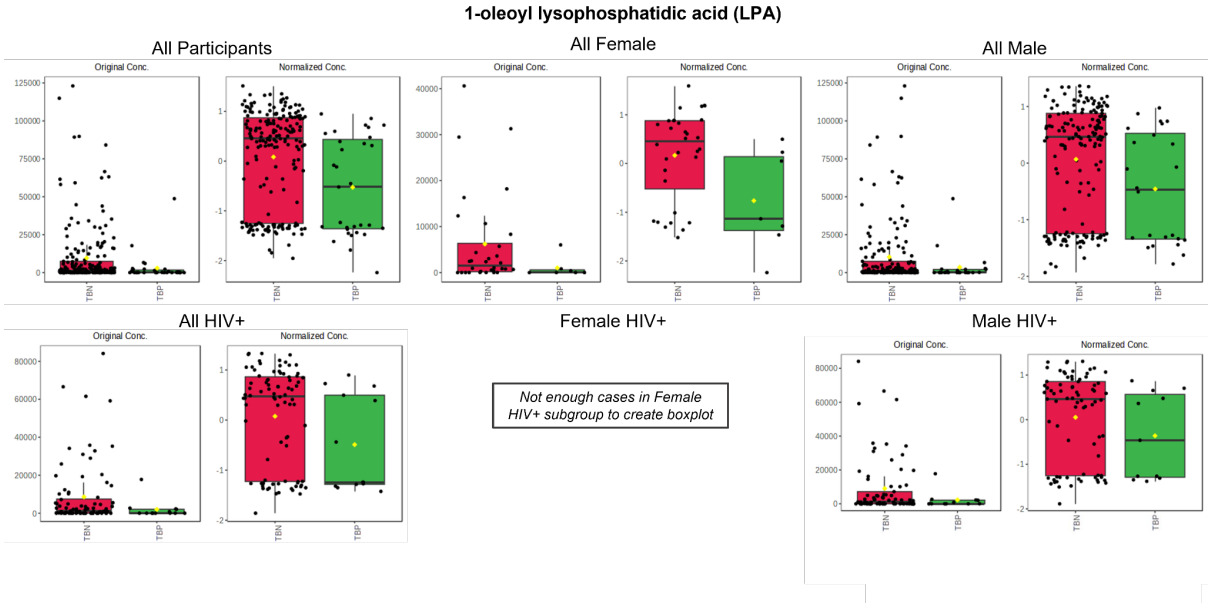


Fig. 4 Boxplots of the compound, annotated as 1-oleoyl lysophosphaditic acid (LPA), in all of the subgroup comparisons, compiled using the original concentrations and normalised concentrations. Due to the small sample size of the female HIV+ group, boxplots could not be created.

5 Conclusion

This is the first study to show that confounding factors can result in major differences in the metabolome's response to *M. tuberculosis* infection. Covariants such as sex and comorbidities should be considered when using metabolomics to explore disease mechanisms. This approach could assist in the development of improved study designs, and therefore, the identification of more specific diagnostic and treatment procedures for TB.

Author contributions

DB and IDP conceived and designed the research. DTL managed the sample collection and ethical aspects. DB conducted the experiments. DB and MVR analysed the data. DB and IDP interpreted the data and wrote the paper. DTL and MVR critically reviewed the paper. All authors read and approved the manuscript.

Funding This study was supported by the Technology Innovation Agency, a national public entity of South Africa.

Data availability Raw data were generated at the Centre for Human Metabolomics, North-West University, South Africa. Data supporting the findings of this study are available from the corresponding author on request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The ethical committee of the North-West University approved this study (NWU-00127-11-A1-03).

References

Adu-Gyamfi, C. G., Snyman, T., Hoffmann, C. J., Martinson, N. A., Chaisson, R. E., George, J. A., & Suchard, M. S. (2017). Plasma Indoleamine 2, 3-Dioxygenase, a Biomarker for Tuberculosis in Human Immunodeficiency Virus-Infected Patients. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, *65*(8), 1356–1358.

Aikawa, S., Hashimoto, T., Kano, K., & Aoki, J. (2015). Lysophosphatidic acid as a lipid mediator with multiple biological actions. *Journal of biochemistry*, *157*(2), 81–89.

Beukes, D., du Preez, I., & Loots, D. T. (2019). Total Metabolome Extraction from Mycobacterial Cells for GC-MS Metabolomics Analysis. *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-4939-8757-3_6

Chen, J.-X., Han, Y.-S., Zhang, S.-Q., Li, Z.-B., Chen, J., Yi, W.-J., et al. (2021). Novel therapeutic evaluation biomarkers of lipid metabolism targets in uncomplicated pulmonary tuberculosis patients. *Signal transduction and targeted therapy*, *6*(1), 22.

Collins, J. M., Siddiqa, A., Jones, D. P., Liu, K., Kempker, R. R., Nizam, A., et al. (2020). Tryptophan catabolism reflects disease activity in human tuberculosis. *JCI Insight*. https://doi.org/10.1172/jci.insight.137131

Darst, B. F., Koscik, R. L., Hogan, K. J., Johnson, S. C., & Engelman, C. D. (2019). Longitudinal plasma metabolomics of aging and sex. *Aging*, *11*(4), 1262–1282.

Duan, L., Gan, H., Arm, J., & Remold, H. G. (2001). Cytosolic phospholipase A2 participates with TNF-alpha in the induction of apoptosis of human macrophages infected with Mycobacterium tuberculosis H37Ra. *Journal of immunology* , *166*(12), 7469–7476.

du Preez, I., & Loots, D. T. (2013). New sputum metabolite markers implicating adaptations of the host to Mycobacterium tuberculosis, and vice versa. *Tuberculosis* , *93*(3), 330–337.

du Preez, I., Luies, L., & Loots, D. T. (2019). The application of metabolomics toward pulmonary tuberculosis research. *Tuberculosis* , *115*, 126–139.

Feng, S., Du, Y.-Q., Zhang, L., Zhang, L., Feng, R.-R., & 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), 159–168.

Geraldo, L. H. M., Spohr, T. C. L. de S., Amaral, R. F. do, Fonseca, A. C. C. da, Garcia, C., Mendes, F. de A., et al. (2021). Role of lysophosphatidic acid and its receptors in health and disease: novel therapeutic strategies. *Signal transduction and targeted therapy*, *6*(1), 45.

Ghazaei, C. (2018). and lipids: Insights into molecular mechanisms from persistence to virulence. *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences*, *23*, 63.

Han, Y.-S., Chen, J.-X., Li, Z.-B., Chen, J., Yi, W.-J., Huang, H., et al. (2021). Identification of potential lipid biomarkers for active pulmonary tuberculosis using ultra-high-performance liquid chromatography-tandem mass spectrometry. *Experimental biology and medicine* , *246*(4), 387–399.

Krumsiek, J., Mittelstrass, K., Do, K. T., Stückler, F., Ried, J., Adamski, J., et al. (2015). Gender-specific pathway differences in the human serum metabolome. *Metabolomics*. https://doi.org/10.1007/s11306-015-0829-0

Liebenberg, C., Luies, L., & Williams, A. A. (2021). Metabolomics as a Tool to Investigate HIV/TB Co-Infection. *Frontiers in molecular biosciences*, *8*, 692823.

Li, J. C.-H. (2016). Effect size measures in a two-independent-samples case with nonnormal and nonhomogeneous data. *Behavior research methods*, *48*(4), 1560–1574.

Li, Z., Zhang, Y., Hu, T., Likhodii, S., Sun, G., Zhai, G., et al. (2018). Differential metabolomics analysis allows characterization of diversity of metabolite networks between males and females. *PloS one*, *13*(11), e0207775.

Michalczyk, A., Budkowska, M., Dołęgowska, B., Chlubek, D., & Safranow, K. (2017). Lysophosphatidic acid plasma concentrations in healthy subjects: circadian rhythm and associations with demographic, anthropometric and biochemical parameters. *Lipids in health and disease*, *16*(1), 140.

O'Shea, K., Cameron, S. J. S., Lewis, K. E., Lu, C., & Mur, L. A. J. (2016). Metabolomic-based biomarker discovery for non-invasive lung cancer screening: A case study. *Biochimica et Biophysica Acta (BBA) - General Subjects*. https://doi.org/10.1016/j.bbagen.2016.07.007

Pang, Z., Chong, J., Zhou, G., de Lima Morais, D. A., Chang, L., Barrette, M., et al. (2021). MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic acids research*, *49*(W1), W388–W396.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Schoeman, J. C., du Preez, I., & Loots, D. T. (2012). A comparison of four sputum pre-extraction preparation methods for identifying and characterising Mycobacterium tuberculosis using GCxGC-TOFMS metabolomics. *Journal of microbiological methods*, *91*(2), 301–311.

Silva, C. A. M., Graham, B., Webb, K., Ashton, L. V., Harton, M., Luetkemeyer, A. F., et al. (2019). A pilot metabolomics study of tuberculosis immune reconstitution inflammatory syndrome. *International journal of infectious diseases: IJID: official publication of the International Society for Infectious Diseases*, *84*, 30–38.

Tager, A. M., LaCamera, P., Shea, B. S., Campanella, G. S., Selman, M., Zhao, Z., et al. (2008). The lysophosphatidic acid receptor LPA1 links pulmonary fibrosis to lung injury by mediating fibroblast recruitment and vascular leak. *Nature medicine*, *14*(1), 45–54.

Weiner, J., 3rd, Parida, S. K., Maertzdorf, J., Black, G. F., Repsilber, D., Telaar, A., et al. (2012). Biomarkers of inflammation, immunosuppression and stress with active disease are revealed by metabolomic profiling of tuberculosis patients. *PloS one*, *7*(7), e40221.

Wesley Burks, A., Holgate, S. T., O'Hehir, R. E., Bacharier, L. B., Broide, D. H., Khurana Hershey, G. K., & Stokes Peebles, R., Jr. (2019). *Middleton's Allergy E-Book: Principles and Practice*. Elsevier Health Sciences.

World Health Organization. (2016). *Global Tuberculosis Report 2016*.

World Health Organization. (2019). *Global Tuberculosis Report 2019*.

Zhou, A., Ni, J., Xu, Z., Wang, Y., Lu, S., Sha, W., et al. (2013). Application of ^{1}H NMR Spectroscopy-Based Metabolomics to Sera of Tuberculosis Patients. *Journal of Proteome Research*. https://doi.org/10.1021/pr4007359