A metabolomics investigation of Tuberculous Meningitis in infants and children

SW Mason

21487855

Thesis submitted for the degree *Philosophiae Doctor* in Biochemistry at the Potchefstroom Campus of the North-West University

Promoter: Prof CJ Reinecke

Co-promoter: Dr RS Solomons

October 2016

This thesis was done through the General Joint Doctorate Agreement between the Vrije Universiteit, Amsterdam and the North-West University (Potchefstroom Campus) in South Africa. The tutoring of the PhD was divided equally between the two institutions through the two respective thesis promoters as stipulated in the agreement, signed on 1 February 2012. The research in this thesis was performed at the Centre for Human Metabolomics, Faculty of Natural Sciences, North-West University (Potchefstroom Campus), South Africa, and the Department of Pediatric Infectious Diseases and Immunology of the Vrije Universiteit Medical Center, Amsterdam, The Netherlands.

The financial assistance of the National Research Foundation (NRF) through their Desmond Tutu Vrije Universiteit Doctoral Scholarship Programme and the Technology and Innovation Agency (TIA) of South Africa towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF or TIA.

PREFACE

Before embarking upon an arduous journey one must be prepared and have reason for it. The journey of writing a thesis is one that is interspersed with trepidations, revelations and insights into academia, a first step on the road towards scholarship, and provides glimpses of the fathomless depth of human knowledge. We start, as is always the case, at the beginning; nervous excitement and naive optimism abound. Out of the gate there is quick realization that this is a marathon, not a sprint. Quietly, almost surreptitiously, one becomes immersed. The qualms of normal life abate, seemingly insignificant in comparison to the journey at hand. Momentum is gained, goals achieved and results of promise obtained. Conversely, there are the occasional setbacks, 'third, fourth, fifth... time lucky' scenarios. Frustrating as they are, negative results raise critical questions and, through renewed experimentation, add to the understanding of the work at hand; building towards the data pool, the foundation on which information is gained. Answering the whys and hows instils a sense of context. Before you know it, the summit is in sight. Years of experience and insights are crammed into what we call a thesis and at the end one is left humble. This is my view on the typical journey of a PhD student. As for the reason for writing a thesis, that is as diverse as each of our genetic codes. For me, it has been for the journey, experiences and wisdom.

My introduction to and appreciation of the science of metabolomics began at the start of my postgraduate career. It involved understanding the complex methodology behind the science, as well as the transition from the mind set of a taught student into one that has begun to grasp the intricacies of the scientific method and the philosophy of science in order to offer relevant insights into proposed biological problems. This beginning ended with the culmination of insights, generated from GC-MS metabolomics data, into the question of acute alcohol abuse in my MSc dissertation. Following this, I began investigating topics of current interest in South Africa related to health and biochemistry. One particular topic stood out — tuberculosis (TB) is a major, and growing, concern in South Africa. Upon reading some of the literature on this disease, it soon became evident that there still exist gaps in knowledge, especially regarding the biochemistry behind the host–pathogen response. Even more so, little is known about the metabolic consequences when the pathogen enters the brain, giving rise to tuberculous meningitis (TBM), a disease that is often fatal in infants and children. To my great good fortune there exists a team of specialists in the Western Cape Province of South Africa, a region where TB is endemic, which closely deals with TBM in its paediatric community. Upon meeting with them, along with experts in TBM from the Vrije University (VU) in the Netherlands, I was given the opportunity to continue my metabolomics journey, now in the field of TBM. This opportunity was afforded by a Desmond Tutu NRF–VU Doctoral Scholarship, which allowed me to travel to the Netherlands, specifically Nijmegen, where I was introduced to the new metabolomics

technology of nuclear magnetic resonance (NMR) spectroscopy by members of the Radboud University Medical Centre. Serendipitously, the Department of Science and Technology in South Africa had, since the beginning of the 2000s, embarked on an extensive new, national investment in biotechnological research and development; one which led to the establishment in 2006 of the Metabolomics Platform by the Technology Innovation Agency, hosted by North-West University (NWU) in Potchefstroom. Thus, access to expertise in the field of metabolomics and TBM, across multiple institutions, both in South Africa and The Netherlands, led to the purposeful and multidisciplinary design of this metabolomics study into TBM in infants and children. Expertise and dilligence have been incorporated into this thesis, which is compiled according to article format as required by the NWU and agreed upon in consultation with the requirements for a thesis at the VU in Amsterdam.

The biological question formulated for this PhD study is "*Can biologically relevant metabolic perturbations be identified in the CSF of infants and children with TBM and are these perturbations reflected in the urine (noninvasive sample collection) as putative biomarkers*?" After much thought and reflection on addressing the biological question, the aims of the investigation (pg. 33) became:

(1) "*Gain new insight(s) into the global metabolite profile close to the site of infection of TBM by analysing CSF using untargeted NMR metabolomics with the intention to obtain a holistic/broad overview of the host's metabolic response to the Mtb infection."*

(2) "*Investigate correlating urine samples by using the more sensitive semi-targeted GC-MS metabolomics approach — unravelling new information on the vast range of excreted metabolites as a result of the perturbations caused by TBM.*"

(3) "*Apply the novel urinary metabolomics information in order to propose a putative biosignature for the noninvasive diagnosis and monitoring of TBM in infants and children from the population group being studied.*"

Technically, the thesis is thus structured in four parts as outlined below.

Part 1 of this thesis consists of an overview of current knowledge about TBM and the role of metabolomics. Chapter 1 begins with a concise overview of the pathological aspects of and existing diagnostic capacity to identify TBM. This is followed by a more incisive overview of the energy-related biochemical aspects of *Mycobacterium tuberculosis* (Mtb), the pathogen responsible for the disease, in the host, from which gaps in existing information are highlighted and articulated as the basis of the biological question for this study. Chapter 2 covers the scientific method of choice, metabolomics, as an overview and in relation to research in TB and meningitis. From this, and the biological question, the aims of this study are articulated, followed

by a general chronological overview of the pre-conceived experimental design for this PhD study and a paper on biosignature identification for respiratory chain deficiency, which involved contributions from me by applying NMR analysis of the biological samples and serves as a template for my study (Smuts *et al.* 2013).

Part 2 begins with a brief overview of NMR metabolomics in research on biofluids, with a more extensive overview given later. In part 2 we introduce the conceptual "astrocyte–microglia lactate shuttle" (AMLS) model, based on the astrocyte–neuron lactate shuttle (ANLS), formulated from the body of work from an untargeted NMR metabolomics analysis of cerebrospinal fluid (CSF) from patients and controls. This AMLS model reflects the host's metabolic response to neuroinflammatory disease, such as TBM. This account of the AMLS model is followed by an overview of TBM in infants and children in a South African context as well as of the recent role NMR metabolomics has played in research on infectious diseases in Africa, particularly TB and meningitis. It further discusses the AMLS model in terms of similar NMR research. The publications yielded from this are Mason *et al.* (2015) and Mason *et al.* (2016a).

Part 3 focuses on semi-targeted gas chromatography–mass spectrometry (GC-MS) metabolomics by first discussing potential sources of error of this highly sensitive method and introduces the novel qualitative method KEMREP for assessing an analyst's ability to produce reliable and repeatable metabolomics data using GC-MS. This is followed by an account of the use of the validated GC-MS method to present a putative metabolic biosignature identified from the urine of TBM-infected infants and children. Scientific publications on this work include Mason *et al.* (2014) and Mason *et al.* (2016c).

Finally, **part 4** serves as a general discussion of the overall results and contributions from this research and a reflection on the aims of the study. This is followed by a brief discussion of questions raised during the course of work on the thesis (Mason *et al.* 2016b) and proposals for future directions that could be taken following this PhD research.

ABSTRACT

This thesis, titled: '*A metabolomics investigation of Tuberculous Meningitis in infants and children',* deals with tuberculous meningitis (TBM), the most severe complication of tuberculosis (TB) and major pandemic of our day. The WHO Global TB Report listed approximately 312 380 new cases of TB in 2013 for South Africa alone, of which up to 10% manifested in the central nervous system — particularly severe as TBM in children. Existing TB tests and diagnostic markers have a low sensitivity and specificity, indicating a lack of valid and specific biomarkers for TBM.

We present here the first comprehensive metabolomics investigation using a homogeneous and well-described TBM infant and children patient group, in a thesis structured into four parts.

Part 1 gives the background on clinical aspects of TBM and a biochemical overview with emphasis on host–pathogen interaction, raising a key biological question: "*Can biologically relevant metabolic perturbations be identified in the cerebrospinal fluid (CSF) of infants and children with TBM and are these perturbations reflected in the urine through putative biomarkers*?" (Chapter 1). The experimental approach to address this question was untargeted proton nuclear magnetic resonance (^{1}H NMR) spectroscopy and semi-targeted gas chromatography–mass spectrometry (GC-MS) metabolomics (Chapter 2).

Part 2 covers the ^{1}H NMR component of the investigation, shown to be a highly repeatable method and useful for an initial, holistic assessment of TBM (Chapter 3). CSF was the biofluid studied, as it is derived from close to the site of TBM infection. The new insight(s) gained from the global CSF metabolite profile (first aim of the study), was expressed as the astrocyte-microglia lactate shuttle (AMLS) hypothesis (Chapter 4). This conceptual AMLS model is further discussed and directives given for hypothesis verification. It is noted that ^{1}H NMR based metabolomics studies offer distinct insight(s) into TB and meningitis (Chapter5).

Part 3 focuses on the GC-MS related aspects of the thesis. Following a brief review (Chapter 6), a new method is described for the qualitative assessment of the precision by which analysts generate a GC-MS metabolomics data matrix — designated as KEMREP (Chapter 7). GC-MS analysis of urine samples from patients and controls revealed a global metabolite profile that characterized TBM (second aim; Chapter 8). The key distinguishing metabolites for TBM were methylcitric, 2-ketoglutaric, quinolinic and 4-hydroxyhippuric acids — SUM-4 — proposed to be a putative diagnostic TBM biosignature (third aim; Chapter 8).

Part 4 discusses the achievements of the thesis in context of the relevant biological and clinical aspects pertaining to TBM (Chapter 9) — addressing the aims. The investigation concludes

(Chapter 10) with perspectives on the limitations and future prospects; illustrated using a targeted ultra-performance liquid chromatography–electrospray ionization–tandem mass spectrometry (UPLC-ESI-MS/MS) method for the determination of the ratio of the L and D enantiomers of lactic acid in CSF samples from TBM patients. This final, follow-up study confirmed that lactic acid in the CSF of TBM cases was only in the L-form, solely a response from the host to the infection, and provided experimental support to the conceptual AMLS model.

Keywords: tuberculous meningitis (TBM); metabolomics; urine; cerebrospinal fluid (CSF); proton nuclear magnetic resonance (^{1}H NMR) spectroscopy; gas chromatography–mass spectrometry (GC-MS); hypothesis; astrocyte-microglia lactate shuttle (AMLS); KEMREP; biosignature; L-lactic acid.

Format: This thesis is presented in article format and meets the requirements set out by North-West University, South Africa and Vrije Universiteit, Amsterdam. Thus the following full, peer reviewed papers forms part of the thesis:

1) Smuts, I., van der Westhuizen, F.H., Louw, R., Mienie, L.J., Engelke, U.F., Wevers, R.A., Mason, S., Koekemoer, G. & Reinecke, C.J. (2013). Disclosure of a putative biosignature for respiratory chain disorders through a metabolomics approach. *Metabolomics*, 9(2):379–391 (Chapter 2.6).
2) Mason, S., van Furth, A.M., Mienie, L.J., Engelke, U.F.H., Wevers, R.A., Solomons, R. & Reinecke, C.J. (2015). A hypothetical astrocyte–microglia lactate shuttle derived from a 1H NMR metabolomics analysis of cerebrospinal fluid from a cohort of South African children with tuberculous meningitis. *Metabolomics*, 11:822–837 (Chapter 4).
3) Mason, S., Reinecke, C.J., Solomons, R. & van Furth, A.M. (2016a). Tuberculous meningitis in infants and children: Insights from nuclear magnetic resonance metabolomics. *South African Journal of Science*, 112(3/4), (http://dx.doi.org/10.17159/sajs.2016/2015008) (Chapter 5).
4) Mason, S., Moutloatse, G.P., van Furth, A.M., Solomons, R., van Reenen, M., Reinecke, C.J. & Koekemoer, G. (2014). KEMREP: A new qualitative method for the assessment of an analyst's ability to generate a metabolomics data matrix by gas chromatography – mass spectrometry. Current *Metabolomics*, 2(1):15–26 (Chapter 7).
5) Mason, S., Tutu van Furth, A.M., Solomons, R., Wevers, R.A., van Reenen, M. & Reinecke, C.J. (2016c). A putative urinary biosignature for diagnosis and follow-up of tuberculous meningitis in children: Outcome of a metabolomics study disclosing host–pathogen responses. *Metabolomics* – Submitted (Chapter 8).
6) Mason, S., Reinecke, C.J., Kulik, W., van Cruchten, A., Solomons, R. & Tutu van Furth, A.M. (2016b). Cerebrospinal fluid in tuberculous meningitis exhibits only the L-enantiomer of lactic acid. *BMC Infectious Diseases* – in press (Chapter 10).

TABLE OF CONTENTS

PREFACE ..... **ii-iv**

ABSTRACT ..... **v-vi**

PART 1: INSIGHTS OFFERED BY PAST TBM RESEARCH AND FUTURE ROLE CURRENTLY OFFERED BY METABOLOMICS

CHAPTER 1 BACKGROUND TO TUBERCULOUS MENINGITIS **2**

1.1 Epidemiology **2**

1.2 Pathogenesis **4**

1.3 Diagnostic capacity **8**

1.4 Biochemistry of TBM **13**

1.4.1 Metabolic aspects of Mtb–host interaction 15

1.4.2 Perspective 20

1.5 Identifying the biological question for this study **21**

CHAPTER 2 METABOLOMICS, AIMS OF THESIS AND EXPERIMENTAL DESIGN **23**

2.1 What is metabolomics? **23**

2.2 Metabolomics workflow **24**

2.3 Metabolomics of TB and meningitis **27**

2.4 Aims of thesis **33**

2.5 Experimental design **33**

2.6 Template paper outlining the first step in biomarker identification **36**

2.6.1 Disclosure of a putative biosignature for respiratory chain disorders through a metabolomics approach (Smuts *et al.* 2013) 37

PART 2: UNTARGETED NMR METABOLOMICS: TOWARDS HYPOTHESIS GENERATION – AN INDUCTIVE ENDEAVOUR INTERTWINED WITH DEDUCTIVE ASPIRATIONS

CHAPTER 3 AN ORIENTATION ON PROTON NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY OF BIOFLUIDS **61**

3.1 Brief theoretical background **61**

3.2 Sample preparation of biofluids for NMR analysis **62**

3.3 Ensuring minimal technical variation 64

3.4 Quality of NMR spectra 65

3.5 Concluding remarks 66

CHAPTER 4 A HYPOTHETICAL ASTROCYTE–MICROGLIA LACTATE SHUTTLE DERIVED FROM A ^{1}H NMR METABOLOMICS ANALYSIS OF CEREBROSPINAL FLUID FROM A COHORT OF SOUTH AFRICAN CHILDREN WITH TUBERCULOUS MENINGITIS (Mason *et al.* 2015) 67

4.1 Abstract 68

4.2 Introduction 68

4.3 Material and methods 71

4.3.1 Experimental design 71

4.3.2 Sample collection, description and storage 73

4.3.3 Sample preparation and ^{1}H NMR spectroscopy 74

4.3.4 Statistical analyses 74

4.4 Results and discussion 75

4.4.1 Data analysis and identification of important variables 75

4.4.2 Brain energy metabolism and the response to TBM 82

4.4.2.1 Metabolic burst 84

4.4.2.2 Amino acids 84

4.4.2.3 Creatinine 85

4.4.2.4 Myo-inositol 85

4.4.2.5 Choline 85

4.4.2.6 Dimethyl sulfone ($DMSO_2$) 85

4.4.3 A hypothetical “astrocyte–microglia lactate shuttle” (AMLS) 86

4.5 Concluding remarks 89

CHAPTER 5 TUBERCULOUS MENINGITIS IN INFANTS AND CHILDREN: INSIGHTS FROM NUCLEAR MAGNETIC RESONANCE METABOLOMICS (Mason *et al.* 2016a) 90

5.1 Summary 91

5.2 Introduction 91

5.3 Pathogenesis and severity of TBM 93

5.4 CSF diagnostic indicators of TBM 95

5.5 NMR metabolomics methodology and applications 96

5.6 NMR metabolomics of TB and TBM 98

5.7 Directives for hypothesis verification 103

5.8 Perspective 106

PART 3: SEMI-TARGETED GC-MS METABOLOMICS: A WALK ON THE ANALYTICAL SIDE OF GC-MS AND APPLICATION TOWARDS A URINARY PROFILE AND BIOSIGNATURE OF TBM

CHAPTER 6 GC-MS: A 'GOLD STANDARD' IN METABOLOMICS 109

6.1 Introduction 109

6.2 Assessing GC-MS and the analyst 111

6.3 Working towards a urinary biosignature for TBM 112

CHAPTER 7 KEMREP: A NEW QUALITATIVE METHOD FOR THE ASSESSMENT OF AN ANALYST'S ABILITY TO GENERATE A METABOLOMICS DATA MATRIX BY GAS CHROMATOGRAPHY–MASS SPECTROMETRY (Mason *et al.* 2014) 113

7.1 Abstract 114

7.2 Introduction 114

7.3 Materials and methods 119

7.3.1 Experimental design 119
7.3.2 Sample selection: background and assumptions 120
7.3.3 GC-MS analysis of organic acids 120
7.3.3 Development of the KEMREP method 122
7.3.3.1 Model description 122
7.3.3.2 Quantitative bounds for smoothed chromatographic profiles 125

7.4 Results and discussion 127

7.5 Conclusions 133

CHAPTER 8 A PUTATIVE URINARY BIOSIGNATURE FOR DIAGNOSIS AND FOLLOW-UP OF TUBERCULOUS MENINGITIS IN CHILDREN: OUTCOME OF A METABOLOMICS STUDY DISCLOSING HOST–PATHOGEN RESPONSES (Mason *et al.* 2016c) 134

8.1 Abstract 135

8.2 Introduction 135

8.3 Materials and methods 138

8.3.1 The study population and sampling 138

8.3.2 Quality control and analytical procedures 139
8.3.3 Statistical analyses 140

8.4 Results and discussion 141
8.4.1 Data generation and analysis 141
8.4.2 Indicators of dysfunctional host metabolism in TBM 144
8.4.2.1 Lipolysis and ketosis 144
8.4.2.2 Perturbed energy metabolism 147
8.4.2.3 Liver damage 147
8.4.3 Indicators of host–microbe response in TBM 148
8.4.3.1 Mtb-induced tryptophan metabolism 148
8.4.3.2 Mtb–host related metabolites 149
8.4.3.3 Gut microbiota catabolism and biotransformation 150
8.4.4 From metabolomics to the clinic 151
8.4.4.1 A biosignature for ketosis 151
8.4.4.2 Exploring a potential Mtb–host biosignature for TBM 152
8.4.4.3 Proposal for a putative Mtb–host biosignature for TBM 157
8.4.4.4 Perspective 158

8.5 Conclusions 159

PART 4: TAKING A STEP BACK AND REFLECTING: A DISCUSSION ON THE TAKE-HOME MESSAGE OF THIS THESIS AND WHERE THIS RESEARCH CAN BE TAKEN

CHAPTER 9 DISCUSSION: INSIGHTS INTO TBM VIA METABOLOMICS 162

9.1 Recapitulation: motivation behind addressing the topic under investigation 162

9.2 Addressing the first aim of the thesis 163
9.2.1 Shifting neuroenergetic paradigms 164
9.2.2 The astrocyte–neuron lactate shuttle (ANLS) hypothesis 165
9.2.3 The brain in crisis 166

9.3 Addressing the second aim of the thesis 170
9.3.1 Beyond homeostasis 171
9.3.2 TBM and allostasis 172
9.3.2.1 Ketosis: a primary host response 172
9.3.2.2 Mtb-related metabolic responses 172

9.4 Addressing the third aim of the thesis 176
9.4.1 Identifying a putative urinary biosignature 176

CHAPTER 10 LIMITATIONS AND FUTURE PROSPECTS 178

10.1 Cerebrospinal fluid in tuberculous meningitis exhibits only the L-enantiomer of lactic acid (Mason *et al.* 2016b) 181

10.1.1 Abstract 182

10.1.2 Background 182

10.1.3 Methods 184

10.1.3.1 Sampling 184

10.1.3.2 Chemicals 185

10.1.3.3 Sample preparation and UPLC-ESI-MS/MS analysis 185

10.1.4 Results 186

10.1.5 Discussion 189

REFERENCES 191

ANNEXURE 1: SUPPLEMENTARY INFORMATION:

A HYPOTHETICAL ASTROCYTE–MICROGLIA LACTATE SHUTTLE DERIVED FROM A ^{1}H NMR METABOLOMICS ANALYSIS OF CEREBROSPINAL FLUID FROM A COHORT OF SOUTH AFRICAN CHILDREN WITH TUBERCULOUS MENINGITIS 235

ANNEXURE 2: SUPPLEMENTARY INFORMATION:

KEMREP: A NEW QUALITATIVE METHOD FOR THE ASSESSMENT OF AN ANALYST'S ABILITY TO GENERATE A METABOLOMICS DATA MATRIX BY GAS CHROMATOGRAPHY–MASS SPECTROMETRY 235

ANNEXURE 3: SUPPLEMENTARY INFORMATION:

A PUTATIVE URINARY BIOSIGNATURE FOR DIAGNOSIS AND FOLLOW-UP OF TUBERCULOUS MENINGITIS IN CHILDREN: OUTCOME OF A METABOLOMICS STUDY DISCLOSING HOST–PATHOGEN RESPONSES 235

ANNEXURE 4: SUPPLEMENTARY INFORMATION:

CEREBROSPINAL FLUID IN TUBERCULOUS MENINGITIS EXHIBITS ONLY THE L-ENANTIOMER OF LACTIC ACID 235

ANNEXURE 5: COPYRIGHT LICENCING AGREEMENTS 295

LAST UPDATED: 08 APRIL 2016

LIST OF TABLES

Table 1-1: Summary of sensitivities of various clinical measures, CSF characteristics and diagnostic markers pertaining to TBM cases as reported by 13 published studies consisting of children only. 10

Table 2-1: Summary of the three inclusion criteria and selected metabolomics data of patients used in this study. 44-45

Table 2-2: Summary of the urinary parameters for the respective controls (22/12) and patients (29/22). 47

Table 2-3: Cross-validation of individual metabolite groups and of the biosignature. 57

Table 2-4: The proposed biosignature. 57

Table 4-1: Quantitative statistical data indicating the important metabolites that discriminate between TBM and non-TBM for both SA_Controls vs TBM and NL_Controls vs TBM cases with dominating metabolites lactate and glucose removed. [The chemical shift, in ppm, of each identified metabolite is given in brackets]. 80-81

Table 4-2: Summarized quantified data of 16 important metabolites discriminating between TBM and controls, compared with normal reference ranges. 83

Table 5-1: Insights offered by nuclear magnetic resonance (NMR)-based metabolomics studies specific to tuberculosis (TB) and meningitis. 97

Table 7-1: Selection of metabolomics-based studies that use technical replicates to assess repeatability and/or reproducibility. 119

Table 8-1: Metabolites that contributed to the separation between the TBM and the control groups, along with their respective statistical significance, ranked according to VIP values. Concentrations of all variables are µmol/mmol creatinine. 143

Table 8-2: Outcomes from the logistic regression models applied to the different experimental groups. 155

LIST OF FIGURES

Figure 1-1: Global incidence rates of TB for 2013 as reported by the World Health Organization (WHO) in the Global Tuberculosis Report 2014 (WHO 2014)..... 3

Figure 1-2: Possible outcomes of Mtb infection (with permission from Young *et al.* 2008)..... 5

Figure 1-3: Life cycle of the granuloma — from formation to lysis (with permission from Russell *et al.* 2010). 7

Figure 1-4: Predicted pathways of propionate metabolism in Mtb as proposed by Savvi *et al.* (2008). 16

Figure 1-5: Side chain β-oxidation by Mtb of cholesterol (adapted from Thomas *et al.* 2011)..... 17

Figure 1-6: Integrated model of routes and regulation in the Mtb citric acid cycle as proposed by Baughn *et al.* (2009). 19

Figure 2-1: Two-dimensional principal component analysis of the controls (indicated by a C and the case number) and patients (indicated by a P and the case number). 53

Figure 2-2: Metabolomics work-flow and cross-validation of metabolite groups and the biosignature. 56

Figure 3-1: Schematic illustration of a precessional orbit of a nucleus whereby the direction of the magnetic field (M) of the charged nucleus returns to its original relaxed state (along z axis). 61

Figure 3-2: Conversion of free induction decay from time domain to frequency domain by means of Fourier transformation (FT). 62

Figure 3-3: NMR spectra for urine (top), serum (middle) and CSF (bottom) taken from healthy individuals and scaled according to the reference peak — TSP, at 0.00 ppm (not shown). 63

Figure 3-4: Coefficients of variation, expressed as a percentage, of 28 pre-selected metabolites from 72 samples measured over a 5-week period. 65

Figure 4-1: Schematic representation of the work flow following data generation, based on statistical pre-processing, data analysis and cross-validation to identify important NMR-derived metabolites of TBM infection. 72

Figure 4-2: ^{1}H 500 MHz NMR spectra of CSF at pH 2.5 scaled according to internal standard peak (TSP) illustrating qualitative differences between single examples of untreated NL_Control (top), treated SA_Control (middle) and TBM cases (bottom). 76

Figure 4-3: PCA scores plot of SA_Controls vs TBM (A) and NL_Controls vs TBM (B), showing natural separation between TBM and non-TBM cases. 77

Figure 4-4: PLS scores plot and corresponding correlation loadings with marked important variables of SA_Control vs TBM (A) and NL_Controls vs TBM (C) reduced cases. The regression prediction plot illustrates that the PLS model classifies the cases correctly in the SA_Control vs TBM (B) and NL_Control vs TBM (D) cases. 79

Figure 4-5: Important metabolites that distinguish between TBM and non-TBM cases based on VIP values >1.0 and significant univariate measures, common to both cases and unique to each case (increase↑/decrease↓ relevant to TBM cases). 82

Figure 5-1: Representation of metabolic pathways of two lactate shuttles within the central nervous system. 101

Figure 5-2: Iterative cycle of knowledge, using the astrocyte–microglia lactate shuttle (AMLS) hypothesis as an example. The existing knowledge forms the basis of inductive reasoning that leads to the formulation of hypotheses, which in turn is followed by the use of deduction to verify these hypotheses and to further existing knowledge. 104

Figure 6-1: Application of smoothing to a raw GC-MS chromatogram (left) to create the density representation (KEMREP-smoothed form) of the chromatogram (right) that is easier to interpret. 111

Figure 7-1: Flow diagram illustrating experimental design focused on analytical procedures of the metabolomics pipeline followed in evaluating KEMREP. 118

Figure 7-2: A: Lineplot of retention time versus the normalized log-transformed data for a single sample (case). B: Accompanying scatterplot of retention times versus the cumulative normalized log-transformed data. The line $H(t)$ represents the linear (or monotone spline) interpolation function. The arrows represent an interpolated retention time. C: Density representation of the chromatogram (i.e. the KEMREP-smoothed chromatogram)..... 123

Figure 7-3: Density estimates for various grid sizes (*G* = 100, 1000, 2000 and 5000) for an individual sample (case). The bandwidth (h), calculated using the method of Sheather and Jones (1991), is indicated in brackets..... 125

Figure 7-4: Overlay of 5 non-normalized chromatograms generated by GC-MS analysis of 5 aliquots of one TBM-infected CSF sample performed on A: GC-MS-1 (398 features) and B: GC-MS-2 (239 features)..... 128

Figure 7-5: KEMREP output of overlays of smoothed chromatograms of 95 common features from 5 repeat aliquots of 1 CSF sample. 129

Figure 7-6: Overlays of density plots of 5 repeat aliquots from a single urine sample. A, B & C: First attempt to generate the GS-MS data by analysts 1, 2 and 3, respectively. D, E & F: Second attempt to generate the GS-MS data by the same analysts, respectively. All analyses were conducted using the identical reagents, SOP and the same GC-MS apparatus..... 132

Figure 8-1: Flow diagram of experimental procedures followed to define the biosignature. 139

Figure 8-2: Plots showing important metabolites reflecting perturbed host metabolism seen in TBM and on the state of ketosis and on the Mtb-host response in different groups of patients and controls. 145-146

Figure 8-3: A global metabolite profile related to TBM..... 153

Figure 8-4: ROC analyses for discriminating TBM, TBM (treated), non-TBM and FMS patients from controls. (a) The discriminator consisted of 6 metabolites (SUM-6) identified by multivariate, univariate and metabolic pathway analyses. Colour code: TBM at admission: red; treated TBM: blue; non-TBM: green; FMS: orange. (b) The comparison between

discriminators SUM-6 (shown in blue) and SUM-4 (shown in red) for TBM, and with discriminator SUM-4 (shown in green) for non-TBM. 157

Figure 9-1: Schematic representation of the ANLS model. Glu = glutamate; Gln = glutamine; GluR = glutamatergic receptor; EAATs = excitatory amino acid transporters; GLUT = glucose transporter; MCTs = monocarboxylate transporters (with permission from Bélanger *et al.* 2011)..... 166

Figure 9-2: The conceptual AMLS model (described in detail in Annexure 1 — Figure A1-2). 168

Figure 9-3: Conceptual model illustrating the dynamic processes and the fine distinction between healthy and diseased states. The progress, or its lack, of TBM can generally be divided into four time intervals (T1–T4). 174-175

Figure 10-1: Representative chromatograms depicting (A) definitive identification of L-lactic acid using the stable isotope (L-lactic acid-d_3); (B) clear differentiation of L and D forms of lactic acid in the spiked sample; (C) in CSF, complete lack of D form of lactic acid with only the L form present; and (D) presence of both L and D forms of lactic acid in urine. 187

Figure 10-2: Scatterplot showing the concentration of lactic acid in the CSF samples and the corresponding CSF glucose (values not reported in text) over different stages of TBM disease (dashed lines indicate respective reference ranges). 188

PART 1: INSIGHTS OFFERED BY PAST TBM RESEARCH AND FUTURE ROLE CURRENTLY OFFERED BY METABOLOMICS

Tuberculous Meningitis. Infant eight months of age. Stage of paralysis, left facial paralysis; left lagophthalmus, bulging fontanelle.

CHAPTER 1 BACKGROUND TO TUBERCULOUS MENINGITIS

1.1 Epidemiology

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (Mtb), is an ancient, persistent disease that remains a deadly issue to this day. TB has become one of the leading causes of death in humans from a single, infectious microorganism with approximately 1.5–2 million individuals dying from it each year (WHO 2014). With the constantly expanding human population and the consequence of immune-compromised individuals, due to the advent of the human immunodeficiency virus *(*HIV), this global epidemic has become a closely monitored pathological scourge, particularly by the World Health Organization (WHO). Global efforts to control and treat this major health threat (through programmes such as the WHO Stop TB Strategy (WHO 2006)) have shown a steady decrease in the incidence rates of this disease (incidence being defined as the number of new and relapsed cases of TB (in all its forms) occurring in a year). However, the lack of definitive biomarkers for the early and objective diagnosis of TB, the realities of health provision in resource-constrained countries and the rate at which the human population is increasing outweighs this noted decrease. This trend results in the number of TB sufferers constantly expanding each year, evident by the fact that the total number of new, global TB cases has increased from 8.0 million to 8.6 million between 2003 and 2012, as documented by the annual WHO Global TB Reports (WHO 2014). TB has been associated historically with poverty and famine, poor living conditions, inadequate health services and overpopulation. These socio-economic factors, linked to the prevalence of this global health burden, are evident by the fact that the highest incidence of TB occurs in the sub-Saharan African region, particularly in South Africa (see Figure 1-1).

TB is most commonly known in its pulmonary form. However, Mtb is not only localized in the lungs, because the systematic spread of the tubercule bacilli can lead to extra-pulmonary forms of TB (EPTB), present in 20% of reported TB cases (Godreuil *et al.* 2007). Central nervous system (CNS) TB accounts for up to an estimated 10% of all EPTB cases (Bhigjee *et al.* 2007; Cherian & Thomas 2011; Rieder *et al.* 1991; Rock *et al.* 2005; CDC 2013). According to the WHO Global TB Report for 2013, there were approximately 312 380 clinically defined new cases of TB in South Africa alone, of which 37 709 (12%) were EPTB, consisting of up to an estimated 3771 cases of CNS-TB. Tuberculous meningitis (TBM), which is not only the most prevalent form of CNS-TB, is also the most common form of

bacterial meningitis (BM) in South African children below the age of 13 years (Wolzak *et al.* 2012), especially in the Western Cape (Donald *et al.* 1996). The pockets of socio-economically deprived communities of this particular region have reported the TB in childhood exceeding 1000/100 000 population (Beyers *et al.* 1996; van Rie *et al.* 1999), which accounts for the unusually high prevalence of TBM.

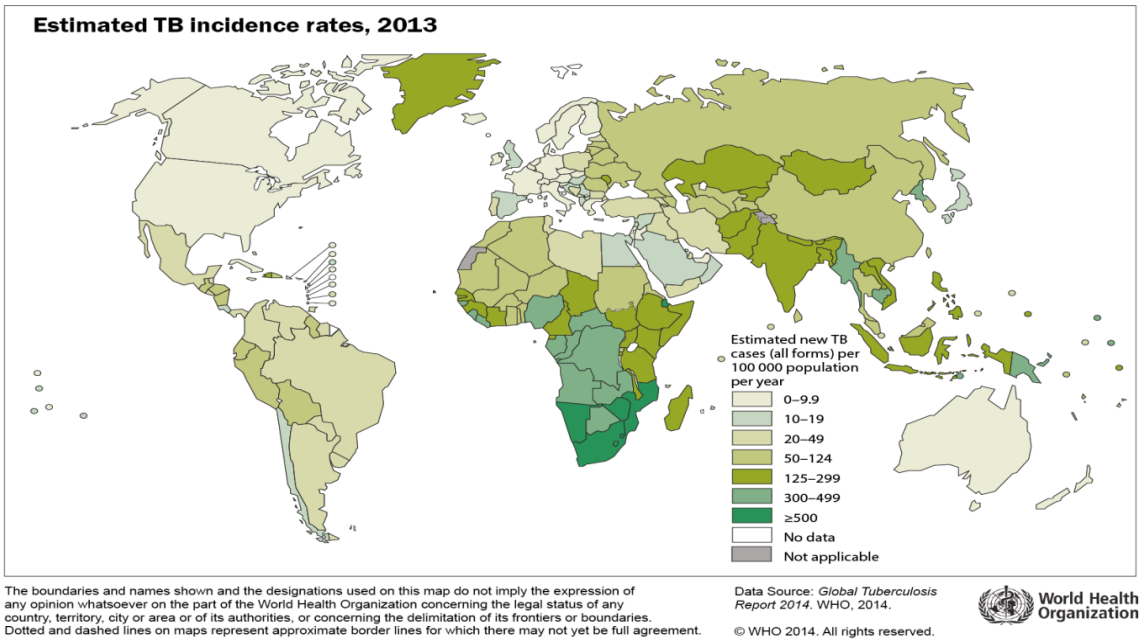


Figure 1-1: Global incidence rates of TB for 2013 as reported by the World Health Organization (WHO) in the Global Tuberculosis Report 2014 (WHO 2014).

Over the past two decades there has been a notable resurgence in research into TB. A study by Alavinia *et al.* (2013) analysed the publications in the field of paediatric tuberculosis in the period 1990–2010, examining 3 417 articles that have been cited 48 459 times (14.2 citations per article). From this study it was shown that South Africa had the second-largest output of articles (12%), as well as the highest total number of new TB cases in children — extracted from the WHO-based country report. Interestingly, this study also showed that the authors with the most articles related to paediatric tuberculosis were in fact South African. Within the current literature there exist numerous publications documenting and comparing the clinical (e.g., based on pathology, immunology and diagnosis), microbiological (e.g., Mtb genotyping and strain analysis, knockout studies and *in vitro* investigations), and treatment aspects of TB. However, there is still limited knowledge of Mtb metabolism within the host,

with gaps in knowledge regarding metabolic characteristics of the pathogen within the host and host response, particularly in the case of TBM.

In this PhD study, specific attention is given to the different arrays of metabolites from TBM patients. These metabolic profiles can serve not only as indicators of what has happened (providing a basis for diagnosis), but also as predictors of what will happen (a foundation for prognosis) in an Mtb-infected individual. Understanding these highly intricate networks of metabolic perturbations caused by TBM can ultimately lead to the identification of metabolic biomarkers (cross-validated single metabolites that provide definitive differential diagnosis) or a metabolic biosignature (comprising multiple markers for improving overall diagnostic accuracy and model stability). Metabolomics serves as an ideal platform (Denery *et al.* 2010; Kell 2006; Madsen *et al.* 2010), which is crucially needed for the more rapid and confident diagnosis of TBM, especially in culture-negative cases. Improved understanding of TBM at the metabolic level would, in practice, allow clinicians to better monitor disease/treatment progression.

1.2 Pathogenesis

Meningitis is a disease characterized by the inflammation of the meninges. The meninges are a three-membrane (termed the outer dura mater, inner pia mater and arachnoid mater) system that envelopes and protects the brain. An interval exists between the meninges, known as the subarachnoid space, through which the cerebrospinal fluid (CSF) flows. This well-regulated biofluid is produced in the ventricles of the brain and in the subarachnoid space, and fulfils a similar function to blood. CSF, typically collected via lumbar puncture, poses a unique analytical and ethical quandary. This biofluid is imperative in the definitive diagnosis of TBM, but it exhibits a diverse number of components in low concentrations. Furthermore, sample volumes collected are usually of limited quantity, especially in young children, as pain and health risks are involved. The aetiological agent responsible for meningitis can be bacterial, fungal or viral, with TBM being a special, chronic form of bacterial meningitis. The focus of this study is on the host response to Mtb in confirmed TBM cases; the reason for this stems from the fact that TB is endemic in the region under study — the Western Cape Province of South Africa — affording us an unusually high prevalence of TBM, which is considered a fairly rare disease in developed countries.

The pathogen is spread through inter-human airborne transmission in which inhaled Mtb bacilli are either eliminated immediately in the upper respiratory tract or reach the alveoli in

the lungs where they are either engulfed by alveolar macrophages or enter the circulatory system and disseminate to other parts of the body. There are four possible outcomes following initial Mtb infection, namely: 1) spontaneous healing, 2) active disease, 3) latent infection, 4) reactivation/reinfection (Godreuil *et al.* 2007; Young *et al.* 2008). The factors that determine which outcome of Mtb infection prevails are dependent upon the host responses, with every individual being unique with respect to the immune response.

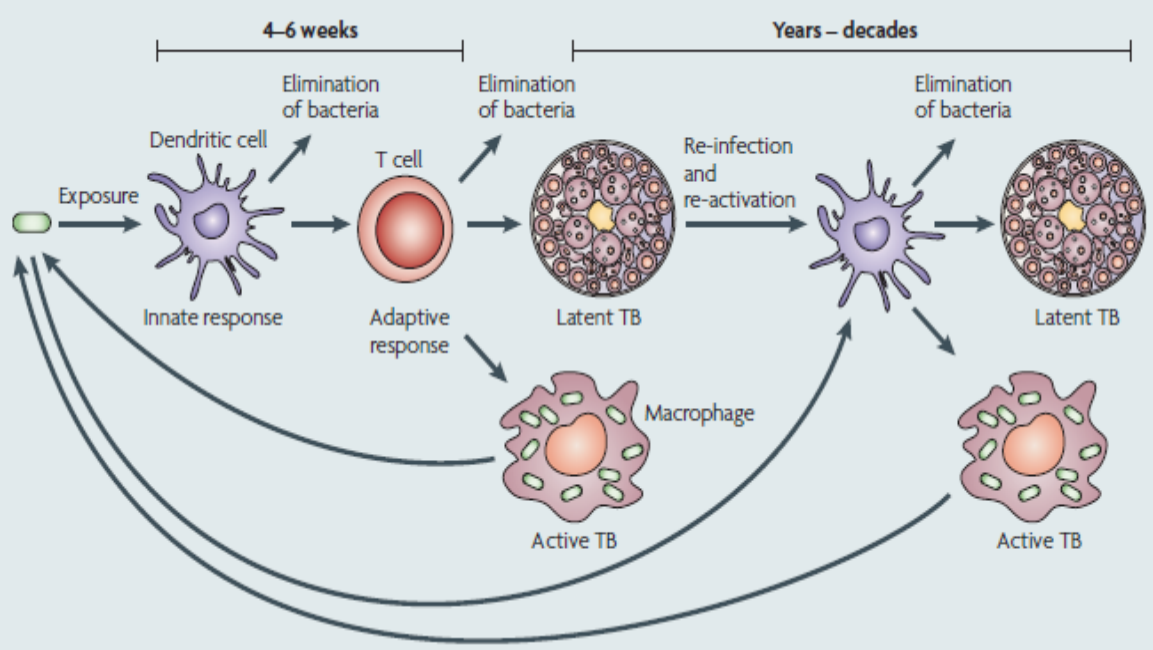


Figure 1-2: Possible outcomes of Mtb infection (with permission from Young *et al.* 2008).

Macrophages that phagocytose the infecting pathogen can act as effector cells, effectively eliminating the Mtb, or they can function as an isolated habitat, preventing antibodies from reaching the source of infection and allowing the bacteria to undergo intracellular replication. Cellular lysis of the infected macrophages results in infection of other macrophages and subsequent intracellular replication. The release of Mtb and cellular debris into the surrounding tissue results in a delayed-type hypersensitivity (DTH) reaction that attracts monocytes and granulocytes to the site of Mtb deposition, resulting in the accumulation of cells and the formation of loosely packed lesions. Activated antigen-specific T cells recognize the Mtb-infected macrophages in these lesions and induce infected macrophages to differentiate into epithelioid histiocytes that aggregate with lymphocytes to form granulomas (Jeong & Lee 2008; Kaufmann 2005; Kaufmann & Parida 2008).

The Mtb is either controlled within the granuloma or continues to multiply, resulting in the development of a caseous-like necrosis within the centre of the granuloma. Rupture/liquefaction of these granulomatous lesions causes the caseous material to be released, resulting in damage to the surrounding tissue and acting as an excellent growth medium for the Mtb to grow and multiply. At this stage of the infection the infected individual is classified as having active TB (also known as primary TB) and can be highly contagious, transmitting Mtb to other, uninfected individuals. In the case of latent infection, the host immune response is sufficient to contain and control the Mtb within the granulomata but unable to eradicate it completely. During this stage of equilibrium no clinical symptoms specific for TBM manifest. The Mtb microbes within these latent granulomatous lesions are surrounded by a fibrotic wall and develop a hypoxic environment. The pathogen enters a state of dormancy, switching from aerobic to anaerobic metabolism and metabolic activity is reduced, which results in highly reduced growth of the bacterium. Latent infection is thus characterized by the persistence of viable Mtb that carries a risk of secondary disease (Godreuil *et al.* 2007; Jeong & Lee 2008; Young *et al.* 2008; Kaufmann 2005; Kaufmann & Parida 2008). If the immune system of an individual with latent Mtb infection becomes compromised (e.g., by HIV infection) or a subsequent reinfection occurs, even with a different strain of pathogen, then the dormant Mtb is resuscitated, resulting in a shift from low to high metabolic activity. Disruption of latent lesions will cause an increase in oxygen content within the granulomas and a subsequent shift of Mtb metabolism back to aerobic respiration. Granulomas liquefy, caseate, cellular debris is released, Mtb flourishes and infection progresses to active disease. This secondary form of active TB disease is also known as reactivation disease. Figure 1-3 illustrates the life cycle of the granuloma from formulation through to lysis during infection/disease.

TBM is the most severe extra-pulmonary manifestation of TB, occurring in approximately 7–12% of pulmonary TB (PTB) cases (Youssef *et al.* 2006). TBM can occur in isolation but typically it is secondary to PTB as a result of the disseminaton of Mtb bacilli into the lymphohematogenous circulatory system. The brain is a very attractive site for the establishment of metastatic foci owing to its rich vascular and oxygen supply, with the bacteria traversing the blood–brain barrier (BBB) by means of a postulated Trojan horse mechanism (Faksri *et al.* 2012), whereby the *Mycobacterium* bacillus imitates a friendly cell. The initial establishment of foci results in the development of small, loculated, tuberculous lesions/plaques known as the Rich focus (Rich & McCordock 1933). These small lesions may remain dormant for years. Pathogenesis of TBM typically occurs as active TB disease in infants and reactivation disease in adults and adolescents. The rupture of caseous, tuberculous lesions discharges cellular debris and Mtb into the subarachnoid space and/or

the CNS (Tung *et al.* 2002; Doerr *et al.* 1995; Evans 2008; Sharma & Mohan 2004; Thwaites & Hien 2005). The pathological features associated with the subsequent DTH reaction are a result of cellular immune response to the Mtb antigens and cellular debris (Sharma & Mohan 2004).

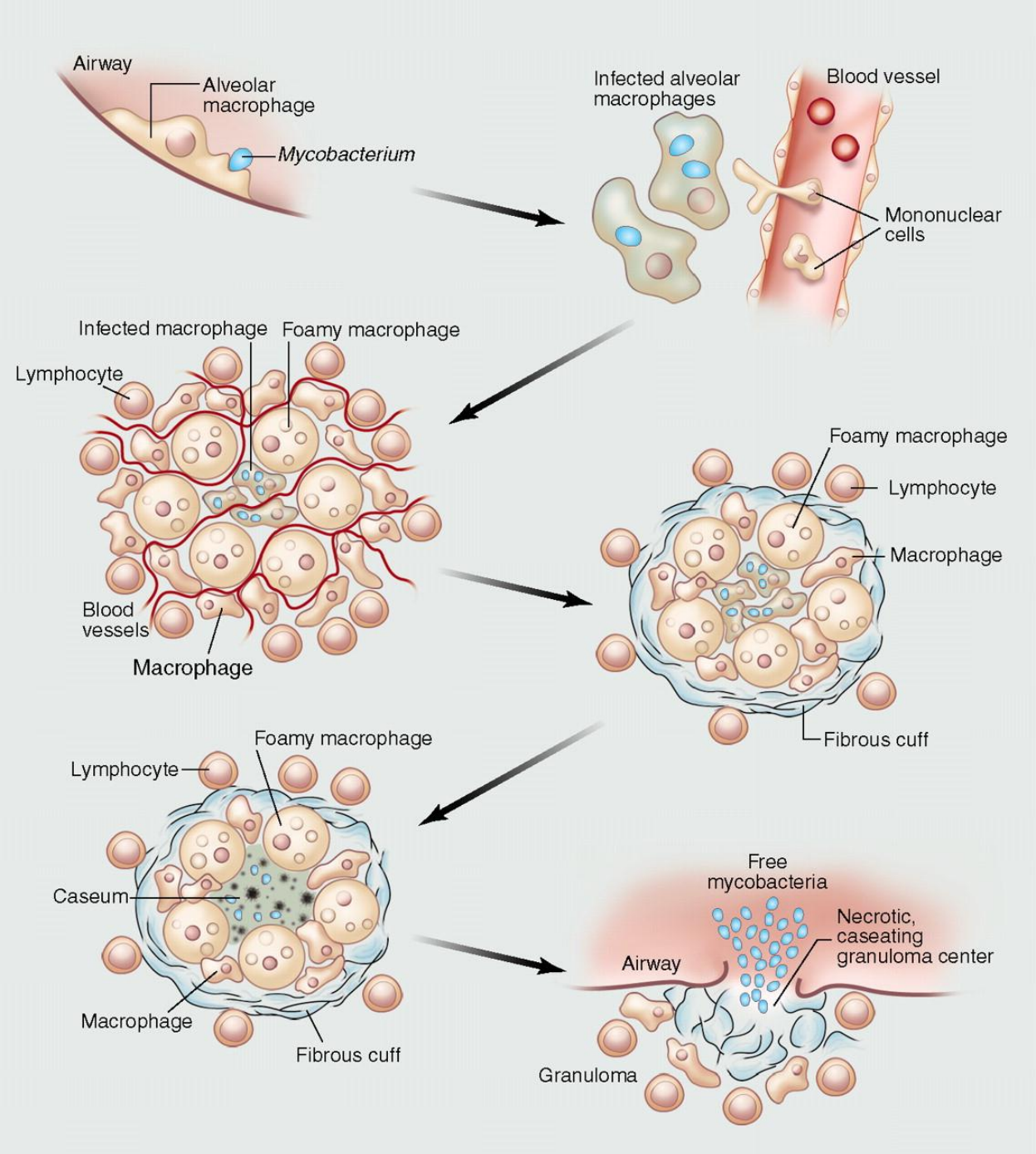


Figure 1-3: Life cycle of the granuloma — from formation to lysis (with permission from Russell *et al.* 2010).

During the early stages of disease, the first cells to respond are polymorphonuclear cells (neutrophils), which recruit monocytes to the site of infection. Approximately three days after infection there is a shift from neutrophilic predominance to mononuclear cell (lymphocyte) predominance at the site of infection. Within the CNS, microglial cells are the resident macrophages of the brain parenchyma. A study by Rock *et al.* (2005) showed that microglia are preferentially infected by Mtb, rather than astrocytes, and that infected microglia elicited production of cytokines and chemokines. Cytokines modulate the immune response and increase the permeability of the BBB, allowing plasma and increased migration of leukocytes into the CNS, typically compartmentalized at the site(s) of infection. Anti-inflammatory cytokines, such as tumour necrosis factor-α (TNF-α) (Thwaites & Hien 2005; Asano *et al.* 2010; Nagesh Babu *et al.* 2008; Shendurnikar & Shastri 1994; Du *et al.* 2006) and interferon-γ (IFN-γ) (Mastroianni *et al.* 1997), are produced in response to pro-inflammatory cytokines in order to mitigate the inflammatory response. Cytokines thus form a network of stimulatory and inhibitory influences on the inflammatory processes with microbial components promoting inflammation by limiting the host's production of anti-inflammatory cytokines.

Thus, it is evident that TBM is a serious complication of Mtb infection, often with some degree of neurological damage. The most efficient manner of minimizing fatalities and long-term neurological deficits is by early diagnosis, and subsequently adequate early treatment. The next section describes the sensitivities, or their lack, of the current diagnostic capacity available to clinicians. It will be made clear that no one diagnostic test is adequate to make a confident diagnosis, and symptoms can be so vague that false negatives are an alarmingly high possibility. Hence, as will be outlined at the end of this chapter, the use of metabolomics in defining the metabolic profiles of TBM cases can be considered invaluable and a technological aspect that needs to be implemented into TBM diagnostics.

1.3 Diagnostic capacity

In order to gain an idea of the range and sensitivity of existing diagnostic methods available to address the TBM disease, an in-depth retrospective overview of the current literature was conducted. A total of 66 studies on TBM were examined, which consisted of 13 on infants and children only, 31 on adults only, and 12 on both children and adults. Since this PhD study focuses on the pediatric population specifically, a detailed comparative analysis was performed on the 13 studies from the literature that focused on infants and children with TBM (Andronikou *et al.* 2004, Delage & Dusseault 1979, Doerr *et al.* 1995, Farinha *et al.* 2000, Kumar *et al.* 1999, Lee 2000, Singh *et al.* 1994, Tinsa *et al.* 2010, Tung *et al.* 2002,

van den Bos *et al.* 2004, van Well *et al.* 2009, Yaramis *et al.* 1998, Youssef *et al.* 2004), summarized in Table 1-1.

The small number of studies in the literature focused specifically on the paediatric population gives a clear indication that there exists limited information regarding TBM in infants and children and highlights how important this particular PhD study is at furthering our understanding of this disease in order to aid and improve upon the rapid diagnosis needed and subsequent treatment.

Table 1-1 highlights the most common clinical information, CSF characteristics and diagnostic markers pertaining to TBM in infants and children. The current tests used to diagnose TBM differentially hold little value on their own; the assessment of results from several tests, along with clinical symptoms, is vital. The simple skin test has low and varied sensitivities (42.4 ± 19.1%); microbiological detection by means of CSF culture and acid-fast bacilli (AFB) smear of CSF have similar low sensitivities of 19 ± 21.1% and 5.7 ± 11%, respectively. Even in three of the adult studies of 100% CSF culture-proven TBM cases, the AFB smear of CSF exhibited positive results in only 2–5% of patients (Girgis *et al.* 1998; Karstaedt *et al.* 1998; Kilpatrick *et al.* 1986). Numerous Mtb-specific antigen tests have been developed that function by using certain antibodies for the detection of particular Mtb antigens. They are simple, rapid, inexpensive and highly specific but exhibit varying sensitivities and are recommended as an adjunct to conventional diagnostic methods.

In diagnosing TB using urinary antigens, lipoarabinomannan (LAM), a mycobacterium-specific lipopolysaccharide component of the bacillus's cell wall, became an important focus. A commercial ELISA test to detect LAM in urine (MTB ELISA Test®, Chemogen, Portland, OR, USA) is now available as Clearview® TB ELISA (Inverness Medical Innovations, Inc., Waltham, MA, USA). Initial assessments were encouraging but more recent evaluations showed that the current commercial urine LAM ELISA test is not useful for independent diagnosis of PTB (Reither *et al.* 2009). Evaluation of the diagnostic value of the Clearview® TB ELISA test, using urine samples from a cohort of paediatric patients suspected of TBM, likewise proved to be of little value (Blok *et al.* 2014).

Therefore, multiple diagnostic tests exist, yet none stands alone as the true ‘gold standard’ method. It is up to the clinician to assess the results of multiple tests to reach a diagnosis, while it is the responsibility of the scientist to continue to research, test and improve upon existing methods to aid the clinician in making the best decision for the patient.

Table 1-1: Summary of sensitivities of various clinical measures, CSF characteristics and diagnostic markers pertaining to TBM cases as reported by 13 published studies consisting of children only.

		13 children studies
Age – mean (range)		23–49 months (< 1–13 years)
Mortality[a]		24.6 ± 14.8 (10–56)
Survivor morbidity[a]		53.4 ± 18.5 (30–80)
HIV [a,b]		3.4 ± 2.2 (0–7)
TBM stage[a]	**I**	6.2 ± 5.1 (8–22)
	II	49 ± 9.7 (30–57)
	III	44.4 ± 9.2 (32–64)
Clinical presentation and neurological involvement[a]	**Fever**	75.4 ± 13.4 (50–100)
	Headache	32.3 ± 12.2 (10–58)
	Neck stiffness	60.3 ± 38.6 (11–98)
	Meningeal signs	77 ± 16.5 (47–86)
	Cough	31.8 ± 5.6 (29–100)
	Lethargy	20 ± 10.4 (16–100)
	Vomiting	60.5 ± 12.3 (25–87)
	Bulging fontanel	18.7 ± 24.3 (5–71)
	Seizures	53.5 ± 13.1 (16–71)
	Significant weight loss	42.6 ± 6.5 (16–46)
	Altered mental state	56 ± 18.1 (16–80)
	Cranial nerve deficits	26.4 ± 1.6 (23–32)
	Paresis/Plegia	61.9 ± 7.1 (14–63)
	Hydrocephalus	74.3 ± 18.7 (10–100)
	Meningeal enhancement	72.4 ± 29.4 (14–100)
	Tuberculoma	10.5 ± 4.7 (2–27)
	Cerebral infarction	24.1 ± 12.9 (10–62)
	Brain edema	ND
Extrameningeal TB[a]	**Chest X-ray**	66.7 ± 12.5 (40–87)
CSF characteristics	**WBC count**	137–222 cells/mm^3 (20–620)
	Protein	100–342 mg/dL (10–1090)
	Glucose	16.7–40 mg/dL (0–49)
	CSF:blood glucose	0.26–0.29 (0–0.92)
	Adenosine deaminase	ND
Serum	**Sodium**	126–131 mmol/L (116–143)
Diagnostic measures[a]	**Tuberculin skin test**	42.4 ± 19.1 (16–86)
	CSF culture[c]	19 ± 21.1 (0–82)
	CSF AFB smear	5.7 ± 11 (0–51)
	PCR	ND

[a] Values, where applicable, given as percentages: weighted mean (%) ± SD (range).

[b] Studies involving 100% HIV as inclusion criteria are excluded from calculations.

[c] Studies involving culture-proven TBM cases only are excluded from calculations.

(ND = not determined).

Often clinical symptoms are few, or even non-existent, particularly in children. The early stage of the disease is accompanied by vague, constitutional symptoms, such as: general ill-health, behavioural changes, irritability, lethargy, apathy, weight loss/anorexia, limb weakness and disturbed sleep patterns (Doerr *et al.* 1995; Sharma & Mohan 2004; Pardasani *et al.* 2008; Tan *et al.* 1999; Anderson *et al.* 2010). As the disease progresses, signs of meningeal irritation begin to become noticeable. The most common clinical symptoms typically observed in TBM patients upon presentation include: fever, vomiting/nausea, seizures, neck stiffness, meningeal irritation and altered mental state. Reports of headaches are markedly fewer in children, probably because infants cannot indicate to a clinician the presence of a headache, whereas symptoms such as bulging fontanel are only detectable in infants, and are a clear indication of raised intracranial pressure.

As shown in Table 1-1, 44.4% of child cases are admitted at stage 3 and as a result often present more commonly with neurological complications such as hydrocephalus and paresis. Indeed, neurological abnormalities detectable by cranial computed tomography (CT) scan occur in approximately 62–94% of TBM cases, of which the severity is dependent on the intensity of the inflammatory response and intracranial pressure. The typical predictors of poor outcome in TBM cases include (Doerr *et al.* 1995; Misra *et al.* 1996; Tan *et al.* 1999; Anderson *et al.* 2010; Sinha *et al.* 2010; Croda *et al.* 2010):

- Extremes of age and advanced clinical stage of disease at presentation.
- Delayed diagnosis and treatment.
- Associated chronic systematic disease (including extraneural TB and HIV co-infection).
- No BCG vaccination[1].
- Presence of focal neurological deficits such as cranial nerve palsies and especially hydrocephalus, leading to raised intracranial pressure.

Most TBM patients present at an advanced stage, often with various combinations of these poor prognosis predictors; as such TBM is associated with a high mortality rate (particularly in young children), with an estimated 30% of patients with TBM dying despite treatment (Youssef *et al.* 2006; Thwaites *et al.* 2002). A literature review of 16 studies, between 1960 & 1977, by Delage & Dusseault noted that the overall mortality rate of TBM was 34.8%, of which the great majority involved children. The highest mortality rates of TBM patients are of

[1] Bacilli Calmette-Guerin (BCG) vaccine – an attenuated strain of mycobacterium related to the pathogen causing human TB that provides protection against severe forms of TB. It is only effective in young children.

those co-infected with HIV (Silber *et al.* 1999; Cecchini *et al.* 2007; Thwaites *et al.* 2005; Hakim *et al.* 2000). However, one study (van Toorn *et al.* 2014) showed that the mortality of childhood TBM recorded at Tygerberg Hospital, Cape Town, was only 3.8%. This result might be specific to this setting, and probably related to the inpatient care and intensive, high-dose anti-TB drugs afforded to the patients.

Because massive and chronic inflammatory responses in the brain are the cornerstone of meningitis (a feature that is not only evident in the CSF but also detectable in the urinary profile, as it will be shown later), many survivors (ranging from as few as 9% to as many as 80% of cases) are often left with some form of permanent neurological damage. Individuals with TBM are typically categorized as one of three clinical stages based upon disease severity:

- Stage I – isolated meningeal disease with normal consciousness/alertness and no neurological abnormalities. Glasgow coma score (GCS)[2] = 15.
- Stage II – signs of meningeal irritation, behavioural changes, lethargy, no/slight change in consciousness/sensorium and mild/moderate focal neurological deficits. GCS = 11–14.
- Stage III – meningeal disease with severely altered consciousness/sensorium (delirium, coma, stupor) and major focal neurological deficits (seizures, abnormal movements, paresis). GCS ≤ 10.

Noteworthy are the studies with a greater percentage of TBM-infected individuals presenting at a relatively late stage of the disease, typically having a higher than average rate of mortality and morbidity (Girgis *et al.* 1998, Hosoglu *et al.* 1998, Morgado *et al.* 2013, Torok *et al.* 2008, Verdon *et al.* 1996), supporting the assertion that an advanced clinical stage of TBM disease at presentation is a predictor of poor prognosis.

Biochemical examination of CSF, collected by lumbar puncture, in TBM cases typically reveals the following characteristics of this biofluid (Bhigjee *et al.* 2007; Youssef *et al.* 2006; Tung *et al.* 2002; Doerr *et al.* 1995; Anderson *et al.* 2010; Thwaites *et al.* 2002):

[2] Glasgow coma score = practical assessment of coma and impaired consciousness. The score ranges between 3 and 15, where 3 is the worst and 15 is the best. Three factors are assessed: (A) eye response (1 = no eye opening, 2 = eye opening to pain, 3 = eye opening to verbal command, 4 = eyes open spontaneously), (B) verbal response (1 = no verbal response, 2 = incomprehensible sounds, 3 = inappropriate words, 4 = confused, 5 = orientated) and (C) motor response (1 = no motor response, 2 = extension to pain, 3 = flexion to pain, 4 = withdrawal from pain, 5 = localizing pain, 6 = obeys command).

- Clear appearance.
- High white blood cell (WBC) count (pleocytosis), greater than 10 cells per microlitre CSF, exhibiting polymorphonuclear cell predominance (neutrophilic, >50% neutrophils or >5 neutrophilic cells per microlitre) during early stages of disease progressing to mononuclear cell predominance (lymphocytotic, >50% lymphocytes).
- Protein levels defined as being either elevated – greater than 40 mg/dL (lower cut-off), or significantly elevated – greater than 100 mg/dL (higher cut-off).
- Lowered glucose defined as less than 2.2 mmol/L, or less than 40 mg/dL, as an absolute value or relatively as less than 0.5 CSF:blood glucose ratio.

These CSF measures are important for aiding rapid differential diagnosis, which is helped by other predictive factors of TBM: 1) length of illness as >5 days, 2) weight loss, 3) cough or night sweats, 4) close TB contact, 5) focal motor deficits, 6) cranial nerve palsies, and 7) altered level of consciousness, all of which are useful in distinguishing TBM from other types of meningitis, particularly BM (Marais *et al.* 2010). Accuracy of using predictive values for the differential diagnosis of TBM varies according to prevalence of TB and if treatment has already started (Anderson *et al.* 2010; Thwaites *et al.* 2002; Kumar *et al.* 1999; Youssef *et al.* 2006). Thus, TBM is typically classified as being either definite, probable or possible, based upon certain criteria (van Well *et al.* 2009; Katrak *et al.* 2000; Sinha *et al.* 2010; Thwaites *et al.* 2005; Kashyap *et al.* 2006; Solari *et al.* 2013; Torok *et al.* 2008; Marais *et al.* 2010).

1.4 Biochemistry of TBM

Based on the limited information in the literature, a CSF chloride level of less than 125 mmol/L (<100 mg/dL) and a CSF lactate level greater than 2 mmol/L can also be diagnostic measures indicative of TBM. Chloride levels in CSF reported in four TBM studies (Patel *et al.* 2004; Silber *et al.* 1999; Thwaites *et al.* 2002; Thwaites *et al.* 2004) ranged between 108 and 113 mmol/L, whereas CSF lactate levels in TBM cases have been reported to range between 4.8 and 5.8 mmol/L in three studies (Thwaites *et al.* 2002; Thwaites *et al.* 2004; Torok *et al.* 2008), far exceeding the normal reference range of 0.45–2.1 mmol/L (Hoffmann *et al.* 1993). The reason why lactate is not currently used as a biomarker of TBM is possibly because it is often, wrongly, considered simply a by-product of normal metabolism. Another important biochemical feature of TBM patients is decreased serum sodium levels — hyponatraemia, the underlying cause of which is largely unknown but possible explanations include: cerebral salt wasting syndrome, inappropriate antidiuretic hormone or hyponatraemic natriuretic syndrome (Torok *et al.* 2008; Thwaites *et al.* 2005). It is postulated

that chronic hypo-osmolarity of serum eventually leads to hypo-osmolarity of CSF (Singh *et al.* 1994), contributing to neurological complications attributed to TBM cases. Based on current data, a person is hyponatraemic if they exhibit a serum sodium level less than 130–135 mmol/L, with TBM-infected individuals experiencing mean serum sodium levels of approximately 127 ± 9 mmol/L (range: 84–149 mmol/L).

The Mtb bacillus is unable to produce certain key factors which it acquires from the host during infection. Thus an imbalance in host metabolic homeostasis is expected. The acquisition of iron and vitamin B12 from host cells by Mtb results in TB patients exhibiting anaemia and vitamin B12 deficiency. The pathogen induces the production of reactive oxygen species (ROS) and reactive nitrogen species (RNS), with malnutrition associated with TB causing impairment of the host's antioxidant capacity, resulting in severe oxidative stress. The toxic ROS not only directly cause damage (e.g., mitochondrial and DNA damage, and membrane instability) but also result in the excessive oxidation of metabolites, leading to the production of harmful lipid peroxidation products (e.g., malondialdehyde) (Reddy *et al.* 2009), as well as increased nitrites/nitrates[3] and myeloperoxidase activity (pro-inflammatory and pro-oxidant mediators) (Çetin *et al.* 2002; Christen *et al.* 2001; Miric *et al.* 2010). Oxidative stress in the neural cells is also evident from an initial increase followed by a depletion of antioxidants, such as: manganese superoxide dismutase (MnSOD — an enzyme that protects against oxygen toxicity (Hirose *et al.* 1995)), ascorbate (which regulates glutamate-induced excitotoxicity) and reduced glutathione (Christen *et al.* 2001; Ghielmetti *et al.* 2003). All of these responses result in neuronal injury related to oxidative stress.

TNF-α, produced in response to neuroinflammation, induces the production of matrix metalloproteinases (MMPs) (Harris *et al.* 2007; Kolb *et al.* 1998; Lee *et al.* 2004; Shapiro *et al.* 2003). Two MMPs in particular (MMP-2 and MMP-9) are implicated in the lysis of the BBB via dissolution of the basement membrane underlying the endothelial cells, thereby contributing to pathogenesis. Elevated levels of both MMP-2 and MMP-9 occur in TBM, according to the severity of neuroinflammation. Dexamethasone decreases CSF MMP-9 concentrations early in treatment and this may represent one mechanism by which corticosteroids improve outcome in TBM (Green *et al.* 2010). Hence, MMP-2 and MMP-9 can be tentatively considered as biochemical biomarkers, as well as prognosis indicators of the development of neurological sequelae.

[3] CSF nitrite levels are used as an indicator of endogenous nitric oxide (NO) production in the biofluid. NO is an inflammation mediator and free radical that is involved in pleocytosis and neuronal injury. It is found to be significantly elevated in CSF in TBM and BM (Çetin *et al.* 2002; Nagesh Babu *et al.* 2008).

It has been noted that TB patients exhibit various forms of vitamin deficiencies (Blumenthal *et al.* 2009). Vitamins not only act as cofactors/coenzymes in enzyme reactions but also exhibit antibacterial properties. The active form of vitamin D (1,25-dihydroxy vitamin D) offers protection against Mtb: it suppresses growth of the pathogen, induces superoxide bursts and enhances phagolysosome fusion (Martineau *et al.* 2007; Möller & Hoal 2010), whereas retinoic acid (vitamin A) suppresses the transcription of an Mtb coat protein that inhibits phagolysosomal fusion (Blumenthal *et al.* 2009). Thus the reduced levels of vitamins A and D in TB patients may simply be a consequence of malnutrition although it is also possible that Mtb may actively decrease these anti-microbial vitamins to ensure its own survival.

1.4.1 Metabolic aspects of Mtb–host interaction

Mtb is a facultative human pathogen that replicates and persists within human macrophages. This chemoheterotroph is highly adaptable, able to survive in hostile and nutrient-poor conditions and utilizes a wide range of sources of carbon and nitrogen — often from the human host. The bacterium is able to shift from a metabolically active growing state in oxygen-rich environments to a metabolically inactive state (dormancy) in hypoxic conditions, minimalizing growth and ensuring persistence. The complexity of Mtb–host metabolic relationships, from the view of the microbe, extends beyond this biochemistry dissertation; it deserves a specialized review, possibly even a thesis, in its own right, in microbiology. Key points, however, need to be addressed here in order to facilitate our understanding of the biochemistry of TBM.

There are two universal, intermediate metabolites that are important for *in vivo* Mtb energy metabolism, namely, acetyl-CoA (a 2-carbon molecule) and propionyl-CoA (a 3-carbon molecule). The first, acetyl-CoA, is produced by the catabolic β-oxidation of even-chained fatty acids and is also biosynthesized by the anaplerotic glyoxylate cycle of Mtb. Acetyl-CoA is utilized in the citric acid cycle and is an important precursor in energy metabolism. The second metabolite, propionyl-CoA, is produced by the β-oxidation of odd- and branched-chain fatty acids and the catabolism of branched-chain amino acids (BCAA). Propionyl-CoA is a high-energy metabolite used in the growth and persistence of Mtb *in vivo*; however, it is highly toxic to host cells when it accumulates, and has also been linked to the production of virulence factors (Savvi *et al.* 2008). The predicted pathway model of propionate metabolism in Mtb by Savvi *et al.* (2008), selected for the purpose of this review and shown in Figure 1-4, provides a holistic graphical view of the precursors of propionyl-CoA — in particular the catabolism of cholesterol and BCAA, and the β-oxidation of fatty acids — and summarizes the subsequent energy-related pathways of the pathogen.

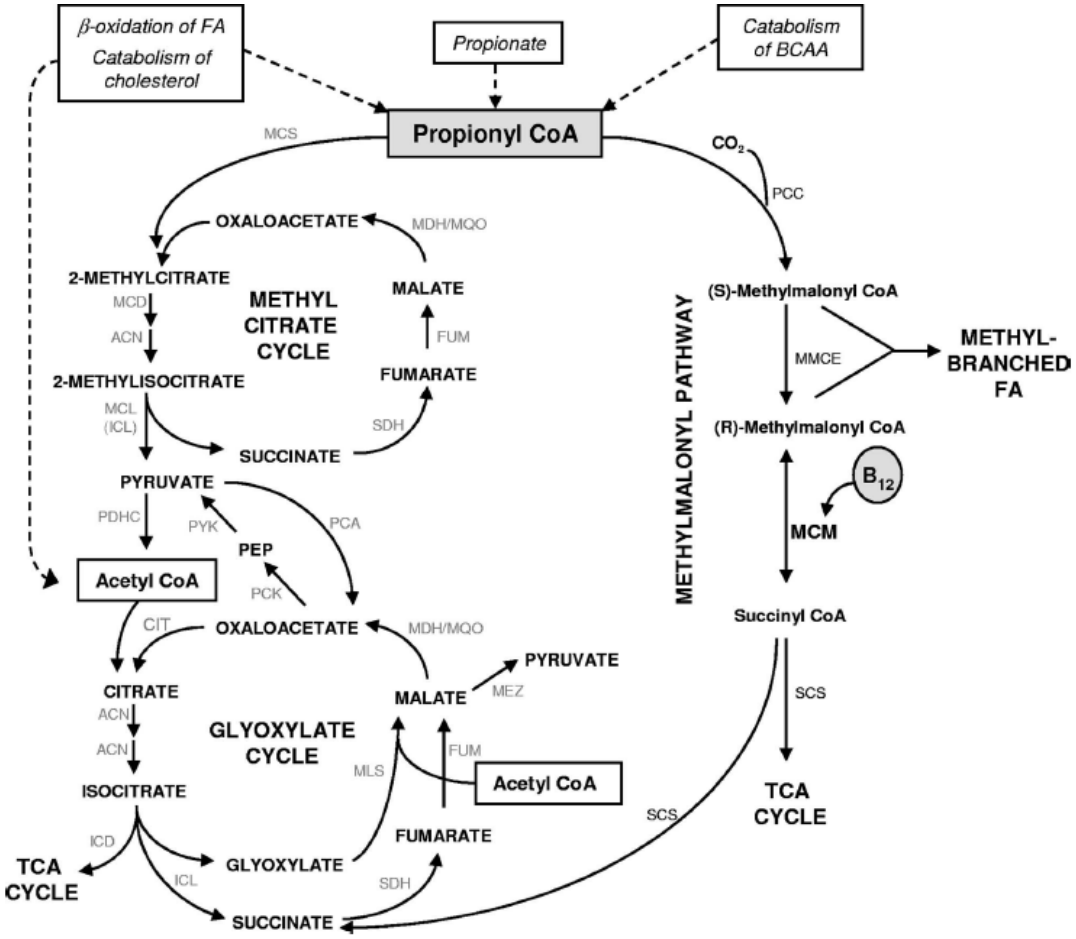


Figure 1-4: Predicted pathways of propionate metabolism in Mtb as proposed by Savvi *et al.* (2008).

The metabolism of host cholesterol by Mtb is an important factor for both its virulence and pathogenesis. TB patients typically exhibit hypocholesterolemia (Deniz *et al.* 2007; Pérez-Guzmán & Vargas 2006), which results in impaired activity of enzymes, cell membrane fluidity and, in particular, the differentiation of lymphocytes. A large cholesterol degradation locus has been observed in the Mtb genome (Cole *et al.* 1998; Griffin *et al.* 2011; Nesbitt *et al.* 2010; van der Geize *et al.* 2007). A study by Thomas *et al.* (2011) elucidated the catabolic pathways of the steroid ring degradation (not presented here) and the side chain β-oxidation of cholesterol as performed by Mtb (Figure 1-5). These intricate chemical reactions of cholesterol catabolism have recently been described in even greater detail (Yang *et al.* 2015), highlighting their complexity. The end products of cholesterol catabolism are the important energy metabolites acetyl-CoA and propionyl-CoA, which are used in Mtb energy metabolism and in the citric acid, methylcitric acid and glyoxylate cycles (Figure 1-4 & Figure 1-6).

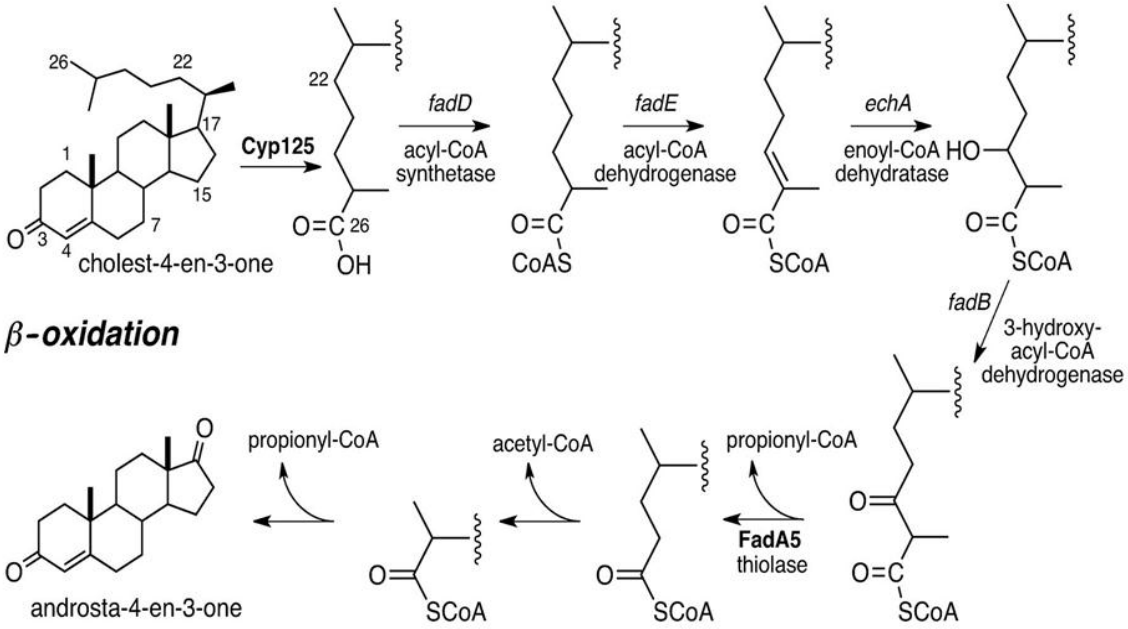


Figure 1-5: Side chain β-oxidation by Mtb of cholesterol (adapted from Thomas *et al.* 2011).

The glyoxylate cycle, unique to plants and bacteria but not present in humans, is important for the intracellular growth of Mtb. It is involved in the biosynthesis of vital bacterial components and replenishes intermediates in the citric acid cycle (see Figure 1-4), thereby ensuring constant energy production. The glyoxylate cycle thus allows Mtb to subsist primarily on fatty acids rather than carbohydrates (Muñoz-Elías *et al.* 2006; Savvi *et al.* 2008). Two enzymes specific to the glyoxylate cycle (not found in the citric acid cycle) are malate synthase (MLS), which mediates the condensation of glyoxylate and acetyl-CoA into malate, and isocitrate lyase (ICL), which catalyses the cleavage of isocitrate to form glyoxylate and succinate). The latter exists as two isoforms, ICL1 and ICL2. Mutagenesis studies have revealed that Mtb lacking ICL1 (but not ICL2) shows impaired persistence in chronic infection. Moreover, Mtb lacking both ICL isoforms is incapable of growth on fatty acids or in macrophages, thereby highlighting the importance of the glyoxylate cycle for the growth and survival of the pathogen, especially for long-term persistence in the host (Beste *et al.* 2007; McKinney *et al.* 2000; Muñoz-Elías *et al.* 2006). In dormancy models, transcriptomic profiling of Mtb has indicated down-regulation of genes involved in energy metabolism, in particular induction and repression of specific genes implicated in the glyoxylate cycle (Deb *et al.* 2009).

Propionyl-CoA is converted to pyruvate in the methylcitrate cycle (see Figure 1-4), which is an essential pathway in propionate metabolism in Mtb. Savvi *et al.* (2008) showed that Mtb was able to use propionate as a sole carbon source in the absence of a functional methylcitrate cycle and with inhibition of the glyoxylate cycle. This observation suggests that Mtb is also able to metabolize propionate through an active methylmalonyl pathway when exogenously supplied with vitamin B12. The final reaction in the methylmalonyl pathway is the reversible intramolecular rearrangement of methylmalonyl-CoA to succinyl-CoA. This reaction is catalysed by vitamin B12-dependent methylmalonyl-CoA mutase (MCM). Mtb, however, is unable to produce vitamin B12 and must therefore acquire this important cofactor from host cells.

The predicted propionate metabolic pathways in Mtb are important in energy metabolism, as proposed by Savvi *et al.* (2008). Figure 1-4 illustrates the sources of propionyl-CoA and the recruitment of acetyl-CoA into the anaplerotic glyoxylate cycle and the processing of propionyl-CoA in the methylcitrate cycle, as well as in the alternative methylmalonyl pathway. However, the notion that fatty acids constitute the predominant carbon substrates used by Mtb in the host has largely neglected the possibility that pathways associated with fatty acid metabolism may serve both endogenously generated and exogenous substrates. This possibility is demonstrated by the constitutive activity of Mtb's methylcitrate cycle functioning on both fatty acid and carbohydrate substrates and its shared necessity for metabolism of both acetate and propionate (Eoh & Rhee 2014).

Mtb possesses its own citric acid cycle; however, α-ketoglutarate dehydrogenase (KGD), which is used in the host's citric acid cycle to convert α-ketoglutaric acid to succinyl-CoA, is lacking in the pathogen. Instead, the bacterium operates separate oxidative and reductive citric acid half-cycles to convert α-ketoglutaric acid to succinic acid using succinic semialdehyde (SSA) as the intermediate (Tian *et al.* 2005). When the glyoxylate pathway is inoperable, the pathway proposed by Tian *et al.* (2005) should support Mtb growth on carbohydrates but this pathway still requires succinyl-CoA to be produced by alternative means. Baughn *et al.* (2009) demonstrated that Mtb in fact is responsible for anaerobic-type α-ketoglutarate ferredoxin oxidoreductase (KOR) activity that results in the formation of succinyl-CoA, thereby completing the pathogen's citric acid cycle. Given the concurrent function of the KOR pathway with the glyoxylate shunt, Baughn *et al.* (2009) predict that the flux of α-ketoglutaric acid runs largely through KOR, rather than through KGD, during the pathogen's growth. This function drives the Mtb citric acid cycle in an alternative, unconventional way.

Thus, it is evident that the intricacies of the host–Mtb metabolic energy pathways are both complex and difficult to distinguish. Figure 1–6 illustrates the integrated model of routes and regulations within just the Mtb citric acid cycle (Baughn *et al.* 2009). Mtb utilizes two distinct tricarboxylic acid cycle (TCA) pathways involving α-ketoglutaric acid, one that functions concurrently with β-oxidation (KOR dependent), and one that operates in the absence of β-oxidation (KGD dependent). When facing starvation, the pathogen may metabolize branched-chain keto and amino acids as sources of energy (Venugopal *et al.* 2011). Thus, Mtb has traditionally been described as a persistent bacterium as can be seen by its ability to maintain metabolic pathways of energy production under various conditions, many of which are still being elucidated.

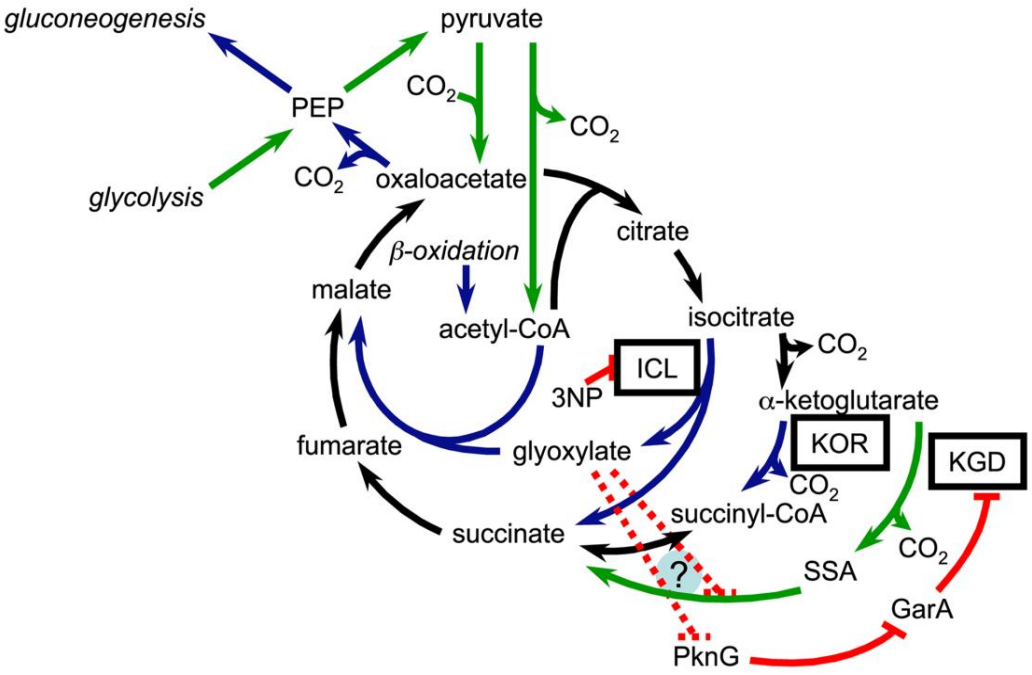


Figure 1-6: **Integrated model of routes and regulation in the Mtb citric acid cycle as proposed by Baughn *et al.* (2009).**

The glyoxylate cycle (inner cycle), canonical citric acid cycle (medial cycle), and variant citric acid cycle (outer cycle) are depicted. Blue lines indicate pathways that are utilized concurrently with β-oxidation, green lines indicate pathways that are utilized during growth on carbohydrates as the sole carbon source, and black lines indicate pathways that are common to both modes of growth. Red lines indicate blocks imposed by 3-nitropropionate (3NP) on isocitrate lyase (ICL), serine threonine kinase (PknG) on the protein (GarA), and GarA on KGD. The dotted red lines represent the putative blocks imposed by glyoxylate on SSA dehydrogenase and PknG.

Other important metabolic pathways necessary for Mtb survival and virulence include arginine catabolism — the bacterium utilizes arginine as a nitrogen source during anaerobic metabolism to ensure persistence (Sürken *et al.* 2008). A further pathway involves sulphur metabolism; mycothiol, a sulphur-containing metabolite, has been shown to participate in a redox defence mechanism, making it important for the bacterium's survival (Schnell & Schneider 2010).

A study by Qureshi *et al.* (1998) identified an altered amino acid profile in TBM patients. These perturbations included: elevated levels of aspartic acid, glutamic acid, gamma-aminobutyric acid, glycine, tryptophan, phenylalanine, arginine and homocysteine as well as decreased levels of taurine and vitamin B12. Some of these amino acids act as precursors to catecholamines. The perturbed levels of these metabolites in the CSF, and subsequently the brain, were attributed to the inflammatory response to Mtb infection and the resulting altered permeability of the BBB. A study by Spranger *et al.* (1996) highlighted the role of excitatory amino acids (EAAs), namely glutamate, in meningitis. EAAs have the unique ability to open receptor-gated ion channels, altering the permeability of membranes and resulting in an excessive increase in intracellular ions and water. This response subsequently leads to cell death, and is associated with glial, neuronal and endothelial impairment. EAA-mediated cell death results in further release of otherwise compartmentalized excitatory neurotoxins. Thus a self-sustaining vicious circle is generated, leading to widespread neuronal and glial damage. The swelling of cells, induced by overstimulation of neurons and astrocytes by EAAs, and ion imbalances lead to increased intracranial pressure and subsequent neuronal injury (Stover *et al.* 1997). The source of EEAs during meningeal inflammation is not fully understood. One postulate is the passive diffusion from compromised glutamatergic neurons; however, Spranger *et al.* (1996) were able to implicate the role of monocytic (but not lymphocytic) inflammatory response in the release of glutamate in BM. Thus EAAs are considered neurotoxic in excessive amounts and are implicated in neuronal cell injury, making them a prime target during the management and treatment of meningitis, which is not yet addressed in current treatment regimes.

1.4.2 Perspective

There is clear evidence that there are various biochemical markers, hitherto unrecognized, which contain valuable information potentially useful in the diagnosis of TBM. This PhD study highlights the importance of small molecules such as organic acids (e.g. lactic acid) and amino acids (e.g. tryptophan), their dynamic nature and the roles they play in TBM. Attention is also given to the views and expectations of paediatricians dealing with TBM. Both these aspects were considered in an attempt to formulate a biological question for this

study. Hence, the investigation reported here strives to bring about the beginning of a shift of view on TBM and to encourage researchers and clinicians to look more closely at the metabolic profiles for diagnostic and prognostic purposes.

1.5 Identifying the biological question for this study

Although the pathogenesis and pathophysiology of TBM are often well described, there exist numerous challenges and gaps in knowledge within the field of TB research, particularly within TBM. The early diagnosis of TBM is essential to ensure prompt treatment in order to reduce the risk of mortality and morbidity. However, the early diagnosis of children is difficult as symptoms are often nonspecific; or even non-existent (atypical). In some cases diagnostic methods produce negative results (e.g. smear and culture evidence can be negative or ambiguous) even though the patient exhibits clinical symptoms of TBM. From the review presented here, it is cleat that currently no definitive, unique diagnostic method, nor biomarker, of TBM exists with sufficiently high sensitivity and specificity. Yet there exist large amounts of, as yet, untapped sources of metabolic information. Metabolomics grants us new opportunities in TBM research, with the overall outcome of identifying novel metabolic biomarkers.

The term biomarker is often used loosely in the literature. In the strictest sense a metabolic biomarker is defined as a quantifiable metabolite that represents: 1) a reference point corresponding to the normal physiological state, 2) a perturbed value for a specific pathophysiological condition (e.g. a particular infectious disease or inborn error of metabolism), or 3) a response indictor to a controlled and defined intervention (e.g., treatment regime). The most important criterion for a biomarker is that it needs to be validated; this typically requires lengthy studies involving a multidisciplinary approach. As stated by Denoroy *et al.* (2013), such a validation may tackle two questions: 1) is this metabolite, which differs in concentration between samples from the test population, indicative of a specific pathophysiology?; and 2) does an abnormal level of this specific metabolite clearly indicate a specific pathophysiology when the other symptoms are considered? Denoroy *et al.* (2013) also state that biomarker validation may comprise three different aspects: 1) validation of the scientific hypotheses derived from the analysis of the metabolite perturbation, 2) technical validation by means of measuring the same perturbation using a different technology platform or a similar platform employing different settings, and 3) validation of the device used to measure the perturbation. As stated by Wallis *et al.* (2013), biomarkers can be the basis for surrogate endpoints that replace typical

clinical endpoints that describe how a patient feels, functions or survives. The validation of a putative surrogate endpoint in TB, however, remains extremely challenging, especially because of the lack of well-characterized biobanks containing biospecimens from patients who have had adequate follow-up to establish long-term treatment outcomes. With this in mind, it is important to understand that this PhD study is geared toward the first step in identifying metabolic biomarkers of TBM. We therefore refer to them as putative biomarkers. The reason for this is that: 1) they are based upon a small test population and there is the danger of false discoveries, especially when sample sizes are small (Kell 2006); and 2) they require rigorous testing and cross-validation, which demands substantial time, patience and good scientific work beyond the aims of this thesis. A successful model publication succinctly outlining the proof-of-concept, scientific approach, frame of mind, terminology and, ultimately, statement of a putative biosignature (the format of which this thesis aspires) is presented at the end of chapter 2.

This thesis attempts to address questions arising from the literature review and sometimes articulated by paediatric clinicians. 'What are the metabolic implications of the composition of the CSF of an infant or child diagnosed with TBM?' 'In what way is the TBM disease reflected in the urinary metabolic profile?' 'Can these metabolic perturbations provide biological information?' And, ultimately, 'Can metabolic biomarkers be identified through a non-invasive procedure?' An amalgamation of these questions leads to the proposed biological question addressed by this study, namely:

> **Can biologically relevant metabolic perturbations be identified in the CSF of infants and children with TBM and are these perturbations reflected in the urine (noninvasive sample collection) as putative biomarkers?**

The experimental approach to this question is a metabolomics investigation; the outline of a typical metabolomics workflow is given in the following chapter.

CHAPTER 2 METABOLOMICS, AIMS OF THESIS AND EXPERIMENTAL DESIGN

2.1 What is metabolomics?

The terms “metabolome” (Oliver *et al.* 1998) and “metabolomics” (Fiehn 2002) have been well defined since their introduction and evolved into an established field of research. Its endpoint can be succinctly articulated as follows: metabolomics is a holistic scientific approach that examines all small molecules defined as metabolites in a domain (e.g. organelle, organism or biofluid), using multivariate statistical methods. The quantified concentrations of metabolites relate to low molecular weight (<1500 Da) molecules and are products of cellular, metabolic and regulatory or pathological processes. Their levels can be regarded as the response of biological systems to influences that are endogenous (e.g., genetic, physiological or age-related) or exogenous (e.g., related to nutrition, infection or medication). A metabolomics analysis can thus be described as a data-driven process to characterize the quantitative composition of these low molecular weight chemicals with the objective of identifying those metabolites of interest in relation to the endogenous or exogenous perturbation under investigation. Metabolomics is viewed as a hypothesis-generating method as opposed to hypothesis-testing of the traditional scientific method; thus, a hypothesis (or hypotheses) is formulated and subsequently evaluated in terms of experimental observations of the metabolomics data generated. The most frequently used analytical techniques in metabolomics — nuclear magnetic resonance (NMR) spectroscopy, liquid chromatography–mass spectrometry (LC-MS) and gas chromatography–mass spectrometry (GC-MS) — require a high level of sophistication from the analysts to generate the metabolomics data. In addition, a biological sample to be analysed may contain a few hundred metabolites, having markedly heterogeneous chemical properties and may be present in concentrations that differ more than 5000-fold.

In a typical systematic approach to a metabolomics study, one should begin with the most holistic, and untargeted approach, which involves NMR spectroscopy. The term untargeted here refers to the simultaneous analysis of multiple classes of metabolites, originating from (apparently) unrelated metabolic pathways. ^{1}H NMR spectroscopy is ideal for this as it involves the analysis of protons in various compounds above the detection limit. These compounds include carbohydrates, lipids, organic acids, amino acids, purines, pyrimidines, and various other classes of metabolites. ^{1}H NMR spectroscopy is a highly specific method,

whereby different compounds display unique spectral patterns based on specific surrounding chemical environments, and has the distinct advantage that it does not require the sample being analysed to be altered chemically (e.g., unlike during the compulsory chemical derivatization in GC-MS), does not come into contact with any part of the analytical apparatus (e.g., the GC or LC column), nor does it destroy the sample as happens in mass spectrometry, making ^{1}H NMR the ideal starting platform. However, established NMR does not match up to the sensitivity of MS-based technology.

Upon identifying the class or classes of metabolites that are of interest in the perturbation(s) under investigation, the next step is a more sensitive semi-targeted approach, which is GC-MS. Here, semi-targeted refers to the analysis of a (apparently related) sub-group of the overall metabolome, such as organic acids, by means of an extraction technique during sample preparation. The higher sensitivity of this technique yields a far greater volume of detectable metabolites than NMR, but often interspersed with artefacts, as discussed in more detail in chapter 7. Finally, after careful analysis of the GC-MS data, one often is left with a few ideal biomarker candidates, based on the formulated hypotheses, for additional analysis, which is best done by means of a highly sensitive targeted approach, such as LC-MS. Here, targeted refers to the analysis of a limited number of known metabolites at specified retention times using an LC column with particular specifications, in line with the compounds under examination. LC-MS yields a more accurate and definite quantifiable result if a stable isotope is used. Thus, ideally, the compound of interest identified by NMR, that is also identifiable using GC-MS, should be clearly distinguishable and quantifiable in LC-MS. A perfect example of this in the case of this study is the identification of highly elevated lactic acid in the CSF of TBM cases using NMR (chapter 5), followed by confirmation by means of GC-MS, and then determination of the exact concentration, as well as the ratio of the physiological L-form or microbiological D-form lactic acid, through UPLC-ESI-MS/MS (Mason *et al.* 2016b – Chapter 10).

2.2 Metabolomics workflow

Metabolomics experiments are not driven by algorithms — finite sets of unambiguous instructions performed in a prescribed sequence — but instead are approached with an open mind and require a rigorous experimental design (Goodacre *et al.* 2007). This design strategy typically encompasses five successive stages. This approach should, however, not be regarded as a mechanistic process to be followed routinely. In practice it may include frequent iterative steps between the analytical and bioinformatics aspects of the stages and

also include some overlap between successive stages, as well as inputs from the multidisciplinary team typically participating in metabolomics investigations.

The typical metabolomics workflow follows five basic steps. Firstly, the biological question to be investigated requires a multidisciplinary reflection, clear definitions and meticulous selection of cases and controls, including attention to aspects like case and clinical descriptions, sample collection and storage (Dunn & Ellis 2005) as well as the bioinformatics approaches intended for data analysis. The planning and design of the overall experiment needs to be addressed by all involved in the study to ensure coherent understanding and logical flow of thought and action. In NMR metabolomics the second stage is basic sample preparation, in which the sample is not chemically changed, does not make contact with the NMR spectrometer nor gets destroyed during analysis.

In GC-MS metabolomics, in contrast, the second stage involves extraction of the section of the metabolome under consideration and derivatization of the extracted components followed by the fractionation of the derivatized substances on the GC column. Subsequently, each component included in the GC chromatogram, which may be one of hundreds of components ranging from just above the detection limit of the equipment to high concentrations of several major components, have to be identified and quantified, using information from the MS detector. The complex composition of the derivatized extracts leads to skewed peak profiles, drift in retention times and variation in response for the various components. This requires pre-processing and deconvolution to convert the raw instrumental data to usable information. For these aspects an array of semi-automated and automated software programs are commercially available. Similarly, NMR metabolomics employs software programs that convert the raw NMR signal into the frequency domain through Fourier transformation, adjust the baseline and phase of the spectrum, and calibrate the spectrum according to an internal standard peak. Calibration and modelling issues of the raw data for both GC-MS and NMR data require consultation from biostatisticians in order for the analyst to produce the n x v matrix of the original data, where n represents the rows which refer to the cases and v indicate the columns which refer to the variables. Typically in a metabolomics data matrix $v > n$ (i.e., there are usually far more variables than cases).

Stage three includes data pre-treatment, processing and post-processing, again dependent on bioinformatics inputs to develop the original matrix further. Selection of the appropriate data pre-treatment is an essential step in the analysis of these data as it may fundamentally affect the identification of the metabolites present (van den Berg *et al.* 2006). The quality of data processing is an essential requirement to enable interpretation of the data, as the primary interest in metabolomics experiments is the differentiation of *relevant* biological

variations from *obscuring* sources of variability (Katajamaa & Orešič 2007). Results that are too optimistic may occur due to methods that over-fit the data, making rigorous validation a compulsory component of the design of a metabolomics experiment (Westerhuis *et al.* 2008).

The fourth stage closes the loop of the metabolomics investigation by biological reflection and interpretation of the research question in view of whatever valid information is generated from the data matrix. However, interpreting the data from global untargeted metabolomics experiments often requires a substantial paradigm change, from the conventional reductionist mode of thinking of the traditional hypothesis-driven scientific method towards the hypothesis-generating context of knowledge generation (Kell 2004). Moreover, metabolomics investigations of human samples infected by pathogens, such as Mtb, disclose a complex profile of host–pathogen interactions. These profiles are superimposed on the complex inter- and intra-individual biological variations of the cases under investigation as a result of dietary differences, co-infections, age and gender that can influence metabolite profiles. Interpreting metabolomics information thus often requires transdisciplinary reflection and participation, which turns out to be one of the most rewarding aspects of this field of study.

Finally, the outcome of an investigation must become public. The primary motivation of the members of the Metabolomics Standards Initiative (MSI) working group (Fiehn *et al.* 2007) was to establish acceptable practices and involvement of those from the metabolomics and related scientific and professional communities. It should be noted, however, that three overriding factors influence the eventual outcome of metabolomics investigations: 1) Participation of a multidisciplinary team in the project, especially when clinical aspects and sampling are involved. 2) It is essential to include the bioinformatics experts from the design to the final stages of the process, which allows for the identification of potential biases and improves the validity of the outcomes. 3) Give equal emphasis to the quality of each stage of the process, as each stage pre-supposes the validity of the preceding stage.

Each metabolomics study engineers its study design based upon unique parameters at hand (e.g., sample type, biological question at hand and depth of analysis). Examples of such metabolomics workflows are illustrated thoughout this thesis — Figures 2-2, 4-1, 7-1 and 8-1.

2.3 Metabolomics of TB and meningitis

The biological question and aims defined for this PhD study are not new. Several investigations on TB and meningitis using a metabolomics approach have been reported during the present investigation (2011 to 2015).

The potential of metabolomics in the research into TB has been suggested by several publications (Kaufmann & Parida 2008; Parida & Kaufmann 2010), with the value of metabolomics being that it is able to: 1) look simultaneously at a wide array of metabolites from different biochemical pathways, 2) detect and isolate patterns of disturbance for additional, targeted investigations, and 3) generate hypotheses to be tested. Godreuil *et al.* (2007) believe that new quantitative and bioinformatics approaches to study the interaction between Mtb and the infected host and how this influences the infection process is of particular importance. This is because it is fully accepted in the scientific area of infectious diseases that the outcome of transmission, infection and disease is dependent on both the intrinsic characteristics of the microbes and of the host.

There have been various metabolomics studies on TB, as well as on the agent responsible for the disease. These studies have revealed information on the following: from a purely analytical viewpoint the methodology behind the metabolomics analysis of TB, the microbial metabolism of Mtb in cultures, the metabolic profiling of Mtb infection in animal models, and, most importantly for this study, in-depth metabolomics analyses of TB in humans.

From a metabolomics methodology point of view, Schoeman & du Preez (2012) investigated four different pre-extraction methods for sputum from PTB patients using two-dimensional gas chromatography time-of-flight mass spectrometry (GCxGC–TOFMS), demonstrating the power and sensitivity of untargeted metabolomics in TB research. Also using GCxGC–TOFMS metabolomics, Meissner-Roloff *et al.* (2012) examined the Beijing lineage of cultured Mtb and were the first to identify the metabolite markers associated with an increased state of virulence, indicating elevated metabolic activity, raised growth/replication rates, increased cell wall synthesis and an altered antioxidant mechanism, all of which would contribute to this organism's increased pathogenicity and survival. Eoh & Rhee (2013) applied ^{13}C-based metabolomic profiling together with LC-MS metabolomics to cultured Mtb to characterize the activity of the pathogen's tricarboxylic acid cycle during adaptation to and recovery from hypoxia, a physiologically relevant condition associated with non-replication. Eoh & Rhee (2013) show that, as Mtb adapts to hypoxia, it slows and remodels its tricarboxylic acid cycle to increase production of succinate, which is used to flexibly sustain membrane potential, ATP synthesis, and anaplerosis in response to varying degrees of

oxygen limitation and the presence or absence of nitrate as an alternative electron acceptor. In a similar type of study, de Carvalho *et al.* (2010) used untargeted LC-MS metabolomics for the metabolite profiling of Mtb growing on ^{13}C-labelled carbon substrates, which revealed that Mtb could catabolize multiple carbon sources simultaneously to achieve enhanced monophasic growth. Moreover, de Carvalho *et al.* (2010) identified for the first time that Mtb differentially catabolized each one of multiple carbon sources through the glycolytic, pentose phosphate, and/or tricarboxylic acid pathways to distinct metabolic fates.

Using animal models, Shin's group reported ^{1}H NMR-based metabolomics results on mice infected with virulent Mtb, using global metabolite profiling of organ tissues and serum. They showed a distinction between control and TB-infected mice. These researchers found that precursors of membrane phospholipids, phosphocholine and phosphoethanolamine, as well as glycolysis, amino acid metabolism, nucleotide metabolism, and the antioxidative stress response, were altered due to the presence of Mtb infection (Shin *et al.* 2011). Through their study, Shin *et al.* suggest that NMR-based global metabolite profiling could potentially provide insight into the metabolic changes in the host's response to Mtb infection, and lead to the elucidation of diagnostic markers for TB progression and response to therapeutics. Also in 2011, Somashekar *et al.* examined TB infection, by a strain of Mtb of modest virulence and drug susceptibility, in tissue (lung, spleen and liver) of guinea pigs through solid-state NMR (high-resolution magic angle spinning (HRMAS)-NMR) metabolomics, as well as serum analysis via ^{1}H NMR-based metabolomics. They identified several metabolites that changed in concert with the progression of infection and discovered that Mtb within the lung granulomatous environment was metabolically active at all stages of infection. Clear group separation between infected and uninfected tissues revealed metabolic fingerprints (in terms of glycolysis, gluconeogenesis and gluataminolysis) that reflected the cycle of active replication and persistence. Interestingly, Somashekar *et al.* (2011) state that their most relevant observations are very similar to metabolic changes seen in cancer during tumour development. They conclude that their study supports the potential of NMR-based metabolomics in proposing biosignatures with potential as surrogate markers for vaccination or therapeutic interventions.

More in-depth metabolomics studies of TB in humans reveal interesting new knowledge. Weiner *et al.* used ultra-high-performance liquid chromatography–tandem mass spectrometry (UPLC-MS/MS) and GC-MS metabolomics in a study in 2012 with the aim of better understanding the biological processes operative during infection and disease of TB. Their study led to the definition of a serum signature (comprising 20 metabolites) for reliable discrimination between active TB and latently infected or uninfected healthy individuals;

these results ultimately led to biomarkers for inflammation, immunosuppression and stress. Several of their findings are congruent with the NMR-based animal studies of Shin *et al.* (2011) and Somashekar *et al.* (2011). Of particular interest, Weiner *et al.* (2012) identified, in active TB cases, lower relative abundances of amino acids, of medium-chain fatty acids and lysophosphatidylcholines, higher relative abundances of fibrinopeptides and of the adenosine degradation products inosine, hypoxanthine and ribose, as well as of other compounds, such as bile acids and uremic toxins. They concluded that their study provided a first step towards the development of metabolomics-based diagnosis of TB and, moreover, towards understanding host metabolic processes during TB, which could ultimately inform individualized supplementary therapy in combination with chemotherapy.

Zhou *et al.*, in 2013, using ^{1}H NMR-based metabolomics, examined a total of 77 serum samples from patients with TB (n = 38) and healthy controls (n = 39). They unambiguously identified 35 metabolites in the sera of the TB patients, of which 17 were different from the controls. The principal group of altered endogenous metabolites in the sera of TB patients were energy related — citric acid cycle intermediates, products of glycolysis, amino acids and indicators of enhanced lipid degradation. The metabolic processes found to be most significantly changed in TB patients were protein biosynthesis, followed by alanine metabolism, phenylalanine and tyrosine metabolism, and ammonia recycling. Zhou *et al.* (2013) concluded also that increased glycolytic products could be an index of tissue hypoxia and the extent of necrosis as the infection progresses, that enhanced lipid degradation can be correlated with caseation of host TB granulomas, and that elevated formate reflects an increased requirement for nucleotide biosynthesis that corresponds to active inflammatory cell division in the host. Zhou *et al.* (2013) also state that their results partially validate previous reports on metabolomics profiling of Mtb-infected murine models (Shin *et al.* 2011; Somashekar *et al.* 2011) and active TB patients (Weiner *et al.* 2012). A follow-up ^{1}H NMR-based metabolomics study in 2015 by Zhou *et al.* highlighted the specificity of the metabolite profile in the sera of TB patients, compared with similar diseases, namely: representative metabolism-related diseases (such as diabetes mellitus), wasting diseases (malignancy), and lung inflammation. Consistent with their earlier study in 2013, in the plasma they found increased levels of lactate, pyruvate, lipids and ketone bodies, and decreased levels of glucose, glutamate, glutamine, glycerophosphocholine, very-low-density lipoproteins and branched-chain amino acids. Zhou *et al.* (2015) could distinguish TB patients statistically, but the greatest overlap occurred with the plasma profiles of patients with malignancy, indicating that, to some extent, the metabolism of TB mimics that of tumour cells — a similar conclusion was reached by Somashekar *et al.* (2011). The 2015 follow-up study by Zhou *et al.* concluded that these results strongly support the notion that NMR-based metabolomics

could contribute to an improved understanding of disease mechanisms and may offer clues to new TB clinical diagnosis and therapies.

While fewer in number, some metabolomics studies on meningitis have also been reported. A small NMR-based metabolomics investigation by Bundy *et al.* in 2005 examined two cultured ecotypes of *Bacillus cereus*, three avirulent and three responsible for BM. Based on their study, the ecotypes were clearly separated in terms of their metabolite profiles, showing that it is possible to use metabolomics methods to classify pathogens according to their expressed physiology, even when it is not possible to infer a direct mechanistic link to specific virulence factors.

A metabolomics study more relevant to this PhD study, that included cases on TBM, was reported in 2005 by Coen *et al.*, who identified two out of 11 of their adult BM patients as having TBM. Although the patient group were heterogeneous based on the infectious agents, the ^{1}H NMR spectra of patients were clearly different from those of control subjects, indicating the distinct presence of glucose, creatine and creatinine organic acids (lactic, pyruvic, citric, glutamic, 3-hydroxybutyric, acetoacetic and acetic acids) and amino acids (glutamine, alanine, valine, isoleucine and leucine). These results suggested a clustering by pathogen (e.g. *Streptococcus pneumoniae* and *Neisseria meningitidis*), which primarily reflects the extent of derangement of brain or CSF metabolism and the host's metabolic response. On the basis of a mean-centred loadings plot the authors mention, but do not present, results that the metabolites with the highest statistical weight were lactic acid and glucose, which were greatly increased (lactic acid) or significantly reduced (glucose) in cases of BM. The authors interpreted the low glucose concentrations in patients with BM to be primarily caused by suppression of membrane carrier-facilitated glucose transport and a switch to anaerobic metabolism (glycolysis) in the brain due to ischemia and cytokine release. They concluded that the high lactate levels were generated mainly by cerebral glycolysis, with species-dependent, minor contributions from microbial metabolites. Coen *et al.* in 2005 did not substantiate the potential source of lactate, originating from the host or microbe, and no mention was made of how increased lactate corresponded to TBM specifically, as they dealt with all patients as a single group, irrespective of the bacterial origin of their infection. The highly elevated lactate concentrations have been reported for patients with BM. Lactate levels have been proposed as a more sensitive and specific marker than the CSF to blood glucose ratio in cases of postsurgical meningitis, although it was noted in a monograph on major problems in neurology that lactate has not been a useful discriminator in several studies on neurological disorders (Wood & Anderson 1998). The study by Coen *et al.* (2005), albeit on a limited number of samples, highlights the

potential of metabolomics in the study of metabolic/neurological perturbations in infectious diseases, particularly meningitis. These authors conclude that ^{1}H NMR-based metabolomics is a potentially valuable diagnostic tool but is complementary to current diagnostic methods, because it cannot yet replace experiments with microbial cultures to determine antimicrobial susceptibility.

A particular limitation of the studies by both Bundy *et al.* (2005) and Coen *et al.* (2005) was the limited number of samples available. The complex issue of having sufficient and well-characterized patients and controls with meningitis was subsequently addressed by Himmelreich *et al.* in 2009 through an experimental animal study. They induced BM in rats through injections of *Cryptococcus neoformans* or *Streptococcus pneumonia* and could thereby analyse a total number of 175 CSF samples collected from 76 animals. These authors reported increased lactate, attributed to anaerobic cerebral glycolysis in response to infection, and low levels of glucose, due to reduced membrane carrier-facilitated glucose transport, microbial metabolism and increased glycolysis (Himmelreich *et al.* 2009). Other discriminatory metabolites included elevated glutamine, acetate and branched-chain amino acids (isoleucine, leucine and valine). They went as far as to suggest that a rapid aetiological diagnosis of meningitis is possible without prior culture.

The only study in the literature that pertains to meningitis in children with a statistically significant number of TBM cases and uses metabolomics-based technology, namely NMR, is by Subramanian *et al.* (2005). They describe an in-house NMR software system, called MENEXSYS, used as a diagnostic tool through the quantification of 12 well-separated and commonly observed CSF metabolites recorded by NMR — hydroxybutyrate, lactate, alanine, acetate, acetone, acetoacetate, pyruvate, glutamine, citrate, creatine/creatinine, glucose (total) and urea. These criteria were applied to CSF from 191 cases of children suffering from bacterial meningitis (n = 85), tuberculous meningitis (n = 47), viral meningitis (n = 35); the controls numbered 24. Subramanian *et al.* (2005) found that the controls could be differentiated from the disease group with a success rate of 96%, followed by the differential diagnosis of tuberculous meningitis with a corresponding value of 77%. The main limitation of this investigation was that it was not an untargeted metabolomics study but instead one built around the testing of a diagnostic software program on pre-defined metabolites developed in-house and thus did not yield any new diagnostic markers nor advance any novel hypotheses related to the pathogenesis of meningitis. Moreover, the study by Subramanian *et al.* (2005) identified, for the first time in any such work, cyclopropane in the CSF of TBM patients. Cyclopropane is a very simple cyclic organic compound (C_3H_6) that exists in gaseous form and, by itself in the biofluid tested, cannot be explained from a

biochemical or biological point of view. The presence of cyclopropane reported by Subramanian *et al.* is most likely an artefact from their unique sampling protocol, which involved snap-freezing samples using liquid nitrogen. Hence, the experimental group studied by Subramanian *et al.* (2005) most closely matches that which was examined in my study by using a similar technology platform (^{1}H NMR). However, most importantly, it was not a metabolomics study by design as they pre-selected 12 CSF metabolites to analyse with the pre-defined, hypothesis-testing aim of determining whether their in-house software system was applicable as a diagnostic tool for the differential diagnosis of meningitis in children.

The value of metabolomics in TBM research is evident from the preceding studies on TB, meningitis and, especially, for my study of TBM. Some of the clear and defining limitations of previous studies were the lack of samples of TBM cases and their clinical descriptions, and the fact that most were pilot studies — no cross-validation nor generation of hypotheses to be subsequently tested were included. Although there is close similarity between those studies and the present investigation, some unique aspects distinguish my study from these previously reported studies. For example, my study involves:

(1) A sufficient[4], homogeneous[5], well-described[6] patient group of TBM cases in a defined age group — infants and children, from the same population, as well as two matched control groups.
(2) Application of both untargeted NMR and semi-targeted GC-MS-based metabolomics techniques on two types of biofluids, CSF and urine, which encompass a relatively broad spectrum of the metabolome.
(3) More in-depth analysis of the metabolic perturbations leading to novel putative biomarkers, validated at various levels.
(4) Completion of the loop of new knowledge by proposing a novel hypothesis, and subsequent conceptual model, based on experimental data and suggesting directives for testing this hypothesis.

[4] The term sufficient as used here is based on the initial meetings with all members of the team involved in the experimental design and the general consensus that a minimum of 20 cases would be necessary in order to obtain statistically relevant results.

[5] Homogeneous here refers to the patient group coming from the same geographical community, thereby sharing similar environmental backgrounds and conditions so as to avoid any confounding outside influences that might influence the results, and consisting solely of pediatric participants – a unique and well-defined age group.

[6] All patients are well described in terms of the comprehensive clinical information collected upon their admission to hospital and following treatment, both aspects of which are highly relevant and necessary for the biological interpretation of metabolomics data – emphasizing the multidisciplinary roles needed by the researchers involved.

(5) Suggested directives for testing the proposed hypothesis and recommended directives to be followed in TBM research subsequent to that presented in this thesis, including one published paper on a possible future study (Mason *et al.* 2016b).

2.4 Aims of thesis

This thesis, up to this point, can be encapsulated by five focal elements: 1) Establishment of a conceptual understanding of the neuropathophysiology and current diagnostic capacity of TBM, particularly in infants and children. 2) Identification of a gap in the current literature and hence formulation of a biological question to be addressed. 3) Description of the workflow of the scientific method of choice — metabolomics. 4) Critical analysis of existing studies that encompass metabolomics investigations in the field of TB and meningitis — identifying their contributions and limitations. 5) Specifically outlining potential new contributions of this thesis to TBM research. From these features, the aims of the thesis can thus be succinctly articulated as three points of intent:

(1) Gain new insight(s) into the global metabolite profile close to the site of infection of TBM by analyzing CSF using untargeted NMR metabolomics with the intention to obtain a holistic/broad overview of host metabolic response towards the Mtb infection.

(2) Investigate correlating urine samples by using the more sensitive semi-targeted GC-MS metabolomics approach — unravelling new information on the vast range of excreted metabolites as a result of the perturbations caused by TBM.

(3) Apply the novel urinary metabolomics information towards proposing a putative biosignature for the noninvasive diagnosis and monitoring of TBM in infants and children from the population group being studied.

2.5 Experimental design

To accomplish the above aims, the objectives are given below in the form of the experimental design of this study, described chronologically, in order to produce high quality metabolomics data.

First, certain major requirements need to be considered:

- Well-described clinical information on participants.
- High quality samples, involving controlled sample collection and analytical protocols, a sufficient number of well-defined samples to ensure statistical validation of data, and appropriate, corresponding control groups that are as homogeneous as possible).
- Quality assurance of analytical methods to ensure reliable metabolomics data.
- An initial holistic approach, involving NMR, to obtain a global metabolite profile of CSF, followed by a more targeted and sensitive approach, using GC-MS, to identify biomarkers in urine.
- Sophisticated bioinformatics for the interpretation of data.
- Assignment of biological relevance to the statistical information.

With the above in mind, the experimental design was given careful thought and attention. The design of this metabolomics study was aimed at applying untargeted as well as semi-targeted metabolomics techniques that can provide as much metabolic information as achievable. In addition, meticulous care was taken to ensure that sampling was well supervised, and that the data output of the experiments could be considered reliable, on the basis of analytical repeatability studies, before considering the biological relevance and context of the results. The progress of this study can be subdivided into several chronological stages:

1) Sampling protocol — first stage: collection of well-defined, high quality samples of sufficient quantity and clinical classification.

- Informed consent of the participants, including their parents.
- Selection of sample group of infants and children (ages: 6 months to 13 years).
- Participants: confirmed tuberculous meningitis (clinical and CSF ─ molecular and microbiological ─ diagnosis) and meningitis-negative cases (controls ─ suspected meningitis later confirmed as negative for any form of meningitis). Also included: a second, independent, age-matched control group of sample subjects unrelated to TB.
- Collection of biofluid samples: CSF from lumbar puncture and urine upon patients' admission to hospital, and additional defined periodic urine samples following specific treatment (up to 2 weeks). A second homogeneous control group consisted of one sample per individual to affirm differentiation from the patient group.
- Collection of all relevant clinical information about patients.

2) Analytical protocol — second stage: orientation of data generation (ensuring high quality data) and untargeted holistic validation for more targeted analysis in the third stage.

- NMR study:

- Analytical orientation and repeatability exercises.
- Broad, nontargeted overview of metabolome of CSF from TBM cases.
- Identification of various metabolites and associated pathways that present as being important with respect to TBM-infected individuals.
- Validation of a more-targeted MS study of specific metabolite group(s) (e.g organic acids).

• GC-MS analytical (repeatability) study:
  - Repeat analysis of independent CSF and urine samples by GC-MS to determine the reproducibility and repeatability of metabolomics data, aimed at ascertaining the quality and accuracy of data produced.

3) Comprehensive data analysis — third stage: in-depth, targeted analysis of TBM and control groups.

• Global comparative TBM study:
  - Large-scale comparison of control and TBM cases, involving both CSF and urine.
  - Untargeted NMR analysis.
  - Utilization of a more-sensitive metabolomics platform (GC-MS).
  - Multivariate statistical methods for processing and interpretation of data.
  - Assigning biochemical importance to variables that have been statistically isolated and identified.

• Time-dependent treatment study:
  - Specialized investigation aimed at analysing any time-dependent metabolic changes of the urinary profile of TBM patients undergoing treatment.

4) Interpretation of newly generated metabolomics information — the final stage: prediction of biomarker(s)/biosignatures and discussion linking identified biomarker(s)/biosignatures to specific metabolic pathways and discussion of causes/sources, implications and possible treatments/preventions of perturbation(s).

5) Reflection — reviewing the new knowledge gained from the overall study and discussion of the translational context of data in terms of real-world application in a clinical setting, as well as addressing questions raised during the course of the investigation and future prospects.

2.6 Template paper outlining the first step in biomarker identification

DNA is designated as a template in biochemical terminology. As a nucleic acid, it serves as a pattern (template) for the synthesis of mRNA, another, but related, nucleic acid. The paper by Smuts *et al.* (2013) acted as a starting point for me and template for the present investigation on TBM — providing proof-of-concept on how a combination of metabolomics technologies could disclose information to define a putative biosignature for a defined disorder. The motivation for that investigation was: "*The identification of mitochondrial disorders (MDs) in patients* [especially in infants and small children] *is still a major clinical and diagnostic challenge in mitochondrial medicine, especially in view of the extensive genetic, phenotypic and clinical heterogeneity of these disorders.*" This motivation clearly relates closely to the background for the biochemical question formulated in my study. Moreover, the key metabolomics technology reported in Smuts *et al.* (2013) was GC-MS, but my contribution to the investigation was a parallel untargeted NMR analysis of samples from the same patients and controls. I co-authored the paper by Smuts *et al.* (2013), which served as a foundation for my first steps towards revealing a biosignature for a serious disease suffered by very many humans. That paper is as follows:

2.6.1 Disclosure of a putative biosignature for respiratory chain disorders through a metabolomics approach

Izelle Smuts[1], Francois H. van der Westhuizen[2], Roan Louw[2], Lodewyk J. Mienie[2], Udo F.H. Engelke[3], Ron A. Wevers[3], Shayne Mason[2], Gerhard Koekemoer[4] and Carolus J. Reinecke[2]

[1] Department of Paediatrics and Child Health, Steve Biko Academic Hospital, University of Pretoria, Pretoria, South Africa,

[2] Centre for Human Metabonomics, North-West University (Potchefstroom Campus), South Africa,

[3] Laboratory for Genetic, Endocrine and Metabolic Diseases, Department of Laboratory Medicine, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands,

[4] Statistical Consultation Services, North-West University (Potchefstroom Campus), South Africa.

Published: *Metabolomics*, 2013, 9(2):379–391.

Abbreviated title:

A biosignature for respiratory chain disorders

Acknowledgements:

We would like to thank Dr M. Duran from the Laboratory for Genetic and Metabolic Diseases, Academic Medical Centre (AMC), Amsterdam, The Netherlands, for his comments on the original manuscript. This study formed part of BioPAD Project BPP007, funded through the South African Department of Science and Technology. Additional financial support from North-West University is likewise acknowledged. S.W. Mason is a recipient of a Vrije Universiteit (VU) Amsterdam-National Research Foundation (NRF)-Desmond Tutu PhD Fellowship.

Keywords:

Metabolomics, respiratory chain disorders, urinary organic acids, urinary amino acids, data reduction, biosignature

2.6.1.1 Abstract

The diagnosis of respiratory chain deficiencies (RCDs) is complicated and the need for a diagnostic biomarker or biosignature has been widely expressed. In this study, the metabolic profile of a selected group of 29 RCD patients, with a predominantly muscle disease phenotype, and 22 controls were investigated using targeted and untargeted analyses of three sub-sections of the human metabolome, including urinary organic acids and amino acids [measured by gas chromatography–mass spectrometry (GC-MS)], as well as acylcarnitines (measured by electrospray ionization tandem MS). Although MS technologies are highly sensitive and selective, they are restrictive by being applied only to sub-sections of the metabolome; an untargeted nuclear magnetic resonance (NMR) spectroscopy approach was therefore also included. After data reduction and pre-treatment, a biosignature comprising six organic acids (lactic, succinic, 2-hydroxyglutaric, 3-hydroxyisobutyric, 3-hydroxyisovaleric and 3-hydroxy-3-methylglutaric acids), six amino acids (alanine, glycine, glutamic acid, serine, tyrosine and α-aminoadipic acid) and creatine, was constructed from uni- and multivariate statistical analyses and verified by cross-validation. The results presented here provide the first proof-of-concept that the metabolomics approach is capable of defining a biosignature for RCDs. We postulate that the composite of organic acids ≈ amino acids > creatine > betaine > carnitines represents the basic biosignature for RCDs. Validated through a prospective study, this could offer an improved ability to assign individual patients to a group with defined RCD characteristics and improve case selection for biopsy procedures, especially in infants and children.

2.6.1.2 Introduction

The identification of mitochondrial disorders (MDs) in patients is still a major clinical and diagnostic challenge in mitochondrial medicine, especially in view of the extensive genetic, phenotypic and clinical heterogeneity of these disorders (Koene & Smeitink 2011). Accordingly, a wide range of medical specialists, including paediatricians, cardiologists, gastroenterologists, neurologists and ophthalmologists, may first encounter these patients (Wong *et al.* 2010). As a consequence, various criteria have been developed from clinical, genetic and biochemical points of view to direct the diagnosis of the RCDs.

Skeletal muscle provides the key material for histological and biochemical analysis of mitochondrial function and RCD diagnosis. Given the invasive procedure of a muscle biopsy under general or local anaesthesia in children or adults, respectively, distinct clinical and

biochemical information is desirable as a directive for a biopsy. Elevated transaminases and creatine phosphokinase are generally accepted as non-specific enzymatic indicators of MDs (Wong *et al.* 2010). Recently, Suomalainen *et al.* (2011) proposed fibroblast growth factor (FGF-21) as a biomarker for muscle-manifesting mitochondrial respiratory chain deficiencies, which needs to be confirmed by a prospective study, including appropriate patient groups (Turnbull 2011). An analysis of urinary metabolites including lactate, alanine, other amino acids, Krebs cycle intermediates and other organic acids provides the least invasive indicators of RCDs, but still lacks specificity as well as selectivity, as pointed out by Koene & Smeitink (2011). By using metabolic profiling of data generated by mass spectrometry (MS), plasma creatine was recently proposed as a specific and sensitive indicator of RCDs (Shaham *et al.* 2010). A subsequent preliminary study confirmed that plasma creatine is elevated in RCDs, suggesting that it could be used in combination with other biomarkers for the diagnosis of MDs (Boenzi *et al.* 2011). In this regard, it has been suggested that “omics” approaches, such as metabolite profiling, might expand the global view of metabolism due to RCD pathology directly or indirectly (Suomalainen 2011), and so support the more efficient identification of improved biomarkers for RCDs. This concurs with the findings of a metabolomics investigation which disclosed the presence of 24 organic acid metabolites that were practically and statistically highly significant for a well-defined group of RCD patients (Reinecke *et al.* 2012).

A biomarker is defined as a feature that is objectively measured and evaluated as an indicator of normal biological processes, pathological conditions or pharmacological responses to a therapeutic intervention (Atkinson *et al.* 2001). A profile of combined biomarkers is called a biosignature. Measuring single markers seems insufficient in dealing with complex diseases, such as RCDs, as outlined above. It has been argued that for complex infectious diseases, such as tuberculosis, a combination of molecular profiles is likely to have more value than single biomarkers (Jacobsen *et al.* 2008), and that a global approach is the analytical route to reveal such markers. Metabolic profiling using a global approach thus proved valuable in the search for biomarkers of complex conditions employing an experimental model for an infectious condition (Wikoff *et al.* 2008) as well as for the inherited RCDs (Shaham *et al.* 2010). We recently proposed that a global approach might disclose a metabolite profile with the potential to define an extended and characteristic biosignature that can be used as a non-invasive screening instrument for RCDs (Reinecke *et al.* 2012).

In the study reported here we have further investigated the metabolite profile in RCDs by analysis of three sub-sections of the human metabolome, included in the evaluation by the

Mitochondrial Medicine Society's Committee on Disease as different laboratory modalities that can contribute to the establishment of RCDs (Haas *et al.* 2008). The three sub-sections are the organic acids and amino acids [measured by gas chromatography–mass spectrometry (GC-MS)], and acylcarnitines [measured by electrospray ionization tandem mass spectrometry (TMS)]. MS technologies are highly sensitive and selective, but also restrictive by applying only to sub-sections of the metabolome. We therefore also included untargeted NMR spectroscopy in this investigation. Although less sensitive than MS analysis, NMR spectroscopy proved to be highly successful as a complementary technique in studies of inherited metabolic diseases (Engelke *et al.* 2004).

2.6.1.3 Materials and methods

2.6.1.3.1 Reagents

Reagents and standards for the extraction of the organic acids were purchased from Merck Chemical Co. (Darmstadt, Germany) and ethylacetate, diethylether and sodium sulphate and 3-phenylbutyric acid from Sigma-Aldrich (St Louis, MO, USA). All the reagents for the amino acid analysis, including the standards (200 µM each), GC column (10 m × 0.25 mm ZB-AAA) and liner were provided in the EZ:faast^{TM} amino acid analysis sample testing kit by Phenomenex, Inc. (Torrance, CA, USA). For the carnitine analysis, acetonitrile, formic acid, and methanol were purchased from Merck Chemical Co. Butanolic HCl (3N) was purchased from Sigma-Aldrich Co. The following standards were obtained from Dr H J ten Brink, Free University Medical Center (VUMC), Amsterdam, The Netherlands: L-carnitine·HCl, acetyl-L-carnitine·HCl, propionyl-L-carnitine·HCl, isovaleryl-L-carnitine·HCl, octanoyl-L-carnitine·HCl, hexadecanoyl-L-carnitine·HCl, [methyl-d_3]-L-carnitine·HCl, [d_3]-acetyl-L-carnitine·HCl, [3,3,3-d_3]-propionyl-L-carnitine·HCl, [d_9]-isovaleryl-L-carnitine·HCl, [8,8,8-d_3]-octanoyl-L-carnitine·HCl, and [16,16,16-d_3]-hexadecanoyl-L-carnitine·HCl.

2.6.1.3.2 Subjects and the selection of samples for the metabolomics analysis

Ethical approval for the study was obtained from the relevant Ethics Committees of the University of Pretoria (No. 91/98 and amendments) and North-West University (No. 02M02). Informed consent was obtained from the parents of patients and controls for the use of the urine samples and biopsy material (where applicable) of their children for research purposes. The original RCD experimental group consisted of 101 clinically selected patients, including the cohort of South African patients described by Smuts *et al.* (2010). Urine samples were obtained at the Paediatric Neurology Unit of the Steve Biko Academic Hospital, Pretoria,

South Africa, at the time when the muscle biopsy was performed; the patients did not receive any specific treatment or supplements often given to patients with MDs. The use of anti-convulsants, such as valproate, is known to cause metabolic derangements and mitochondrial toxicity (Sztajnkrycer 2002), which might lead to biased mitochondrial-related markers; its use was not stopped in the patients, however, because of health risks and attendant ethical consequences. The controls were selected from among children referred to the clinic, but for whom no prevailing disorder was detected. Aliquots of all samples were stored at –80^{o}C prior to metabolomics analyses. This cohort provided the basis for the selection of samples from patients and controls for these analyses.

Metabolomic investigations are most successfully conducted with control and patient groups which are clearly distinguished from one another, because sample selection is one of the most important aspects of any metabolomics analysis. As mutational analyses of mtDNA and nDNA were not part of the routine procedures used to diagnose the present RCD patient group, other available clinical as well as biochemical parameters were selected to ensure the clear distinction between controls and patients, while retaining the intrinsic heterogeneity of the selected RCD group. Three inclusion criteria were thus formulated to define the patient and control groups: (1) clinical criteria characteristic of RCDs [including the intrinsic property of having a predominantly myopathic phenotype as described in Smuts *et al.* (2010)]; (2) a proven deficiency in one or more complexes of the RC as measured by biochemical enzyme analyses; and (3) elevated excretion of the total urinary organic acids in the patient group, to the extent of there being no overlap of these values for the controls and patients. Although elevated urinary excretion of organic acids is not a recognized principle for diagnosis of RCDs, it was included to ensure separation between the groups used in unsupervised multivariate analyses for the comparative metabolomics (Reinecke *et al.* 2012). Samples from 51 cases (29 patients and 22 controls, designated as Group 1) satisfied these criteria and were available for the MS analyses of the organic acids, amino acids and acylcarnitines; sufficient urine from only 34 of these cases (20 patients and 14 controls, designated as Group 2), however, was available for the NMR analyses. The characteristics of Groups 1 and 2 are shown in Tables 2-1 and 2-2.

2.6.1.3.3 Biopsy material and enzyme analyses from the patient group

Enzyme analyses were performed on muscle biopsies from the *vastus lateralis* muscle of all patients complying with the Mitochondrial Disease Criteria as defined by Wolf & Smeitink 2002. The analyses were conducted according to the procedures fully described previously; we also recognized the two criteria used to identify an enzyme deficiency in this patient group (Reinecke *et al.* 2012). As summarized in Table 2-1, the 29 patients selected thus had

a muscle deficiency of either complex I (CI; five cases), complex III (CIII; four cases) or several different deficiencies of more than one RC enzyme (CM; 20 cases).

2.6.1.3.4 Acquisition of metabolite data

2.6.1.3.4.1 Untargeted metabolic analysis using nuclear magnetic resonance spectroscopy

Proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy was included in this investigation for its high selectivity, provision of unambiguous information about a metabolite and the direct analysis of samples that did not require any prior sub-fractionation for metabolite selection. This work was conducted at the Laboratory for Genetic, Endocrine and Metabolic Diseases, Department of Laboratory Medicine, Radboud University Nijmegen Medical Centre according to standard procedures used there (Engelke *et al.* 2007). Urine samples from 34 cases of Group 2 were used in the ^{1}H-NMR study. These urine samples were analysed using one-dimensional (1D) ^{1}H-NMR spectroscopy. One millilitre of urine was centrifuged at 3000 rpm for 10 min and 700 µL supernatant was transferred into a clean test tube. To this, 70 µL of 20.2 mM standard trimethylsilys-2,2,3,3-tetradeuteropropionic acid sodium salt (TSP) in ^1H_2O was added, the pH adjusted to 2.5 ± 0.05 with concentrated HCl, and 650 µL was transferred to a 5-mm NMR tube. Each sample was analysed in a 500 MHz Bruker DRX spectrometer at 256 scans with a pulse of 7 µsec and a delay of 4 sec. The resulting free induction decay (FID) was converted into frequency domain by Fourier transformation, thereby yielding a ^{1}H-NMR spectrum. The instrument was equipped with a sample changer and each urine sample's ^{1}H-NMR spectrum was analysed individually. The dominant metabolites typically present in urine were detected in all samples. Six notable metabolites were identified based upon their chemical shift resonances at pH 2.5, namely, alanine (1.51 ppm (doublet)), betaine (3.26 ppm (singlet)), creatinine (3.13 ppm (singlet)), creatine (3.05 ppm (singlet)), lactic acid (1.41 ppm (doublet)) and succinic acid (2.66 ppm (singlet)). Each of the above peaks was manually selected and the area under the peak was calculated using a software program (Bruker Amix). Each selected metabolite was quantified relative to creatinine, using the integral and the number of protons with respect to each peak. Interference from medication made the selection, and thus the quantification, of certain metabolites (particularly alanine) impossible for some urine samples. The creatine and betaine values obtained for 14 controls and 20 patients are included in Table 2-2, where applicable.

Table 2-1: Summary of the three inclusion criteria and selected metabolomics data of patients used in this study.

Patients	Criterion 1	Criterion 1	Criterion 1	Criterion 1	Criterion 1	Criterion 2	Criterion 3	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6	Data 7
Number	Clinical profile[1]	MDC score[2]	Lactate (mmol/l)	Pyruvate (mmol/l)	L/P	RC enzyme defect: Percentage of the lowest control value	Total OA (mmol/mol Cr)	Total AA (mmol/mol Cr)	Acetyl-Car (mmol/mol Cr)	Carnitine (mmol/mol Cr)	DC372/370	OC344/342	Creatine (mmol/mol Cr)	Betaine (mmol/mol Cr)
P10_1	M, CNS, DD, DYS	6	2.50	0.10	25.00	CI: 95	1159	1485	2.10	3.39	0.12	0.86	0.84	0.05
P27_1	M, CNS, Eye, ENT, DD, DYS	7	6.00	0.47	12.80	CI: 94	2695	4765	15.27	35.80	0.10	0.15	nd	nd
P30_1	CNS	6	3.53	0.11	32.18	CI: 76	619	1874	8.45	13.30	0.24	1.81	nd	nd
P47_5	M, Eye, R, DD, PNS	7	1.60	0.07	22.85	CI: 66	3049	9719	17.96	44.24	0.42	0.08	3.3	0.38
P73.1_A	M, CNS, Eye, L, DD, PNS	8	0.80	0.08	10.00	CI: 72	1514	901	25.48	22.88	0.41	2.83	0.46	0.04
P36_1	M, CNS, DD	3	1.60	0.13	12.31	CIII: 90	1397	1034	12.85	7.43	0.10	0.11	0.77	0.04
P59_1	M, CNS, Eye, L, DR	7	2.40	0.12	20.70	CIII: 71	1137	3284	14.71	35.02	0.10	0.08	0.66	0.19
P60_1	M, AID	4	1.07	0.13	8.20	CIII: 95	1223	1399	4.57	3.82	0.41	0.24	nd	nd
P84_1_A	M, CNS, Eye, G, R, DD	8	nd	nd	nd	CIII: 97	1297	1119	5.06	6.56	0.14	0.07	0.56	0.03
P01_2	M, Eye, PNS, DD	7	2.80	0.20	14.00	CI, CII+III: 0, 0	873	2877	4.12	2.29	0.41	0.67	nd	nd
P07B_1	M, Eye, DD	3	1.90	0.14	13.57	CII, CII+III: 80, 64	2041	2068	6.18	12.35	0.21	0.25	3.23	0.06
P11_1	M, CNS, Eye, R, DD	8	2.50	0.35	7.10	CI, CIII: 79, 78	845	1058	2.12	1.32	0.58	2.76	0.26	0.01
P14_1_1	M, CNS, Eye, DD, DR	6	1.10	0.10	11.00	CCI, CII+III: 77, 89	1985	2492	4.94	12.86	0.39	12.51	2.03	0.1
P21_1	M, Eye, PNS, End	8	2.80	0.32	8.80	CIII, CII+III: 98, 92	3218	4781	10.07	15.36	0.17	0.09	0.52	1.44
P39_3	M, R, DD	4	1.30	0.20	6.50	CI, CII, CIII, CII+III: 91, 93, 81, 83	1841	6544	29.42	74.93	0.40	0.13	nd	nd
P43_1	M, Eye, DD	7	2.00	nd	nd	CII, CIII, CII+III, CIV: 98, 94, 70, 91	1208	3961	9.71	27.13	0.28	0.35	2.56	0.07
P55_1	M, End, ENT, G, DD	10	nd	nd	nd	CI, CIII: 86, 71	849	1784	5.33	7.55	0.25	0.13	2.24	0.13
P62_1	M, CNS, End, S, DD	8	3.30	0.17	19.41	CIII, CIV: 87, 80	864	904	1.80	0.54	0.07	0.19	0.92	0.04
P67_1	M, CNS, L, DD	6	1.01	0.08	12.60	CIII, CIV: 65, 92	941	850	3.94	7.83	0.29	0.55	0.88	0.12

P69.1_A	M, End, DD, DYS	5	1.10	0.10	11.00	CCII+III, CIV: 90, 97	779	1901	9.44	2.07	0.06	1.21	nd	nd
P70.1_A	BE, CNS & PNS	4	1.90	0.16	11.90	CCI, CIII: 84, 82	586	1862	6.44	12.06	0.08	0.17	0.85	0.06
P71_1_A	M, G, Car, DD	4	2.20	0.12	18.30	CIII, CIV: 99, 71	1077	1745	5.56	4.78	0.38	0.18	0.93	0.03
P75.1_A	CNS, Eye, S, DD, DR, BE	6	3.00	0.14	22.40	CI, CIII, CIV: 90, 97, 37	820	695	9.85	6.54	0.20	0.26	0.13	0.01
P76.1_A	M, CNS, Eye, ENT, S, DD	5	nd	nd	nd	CI, CIII: 37, 33	814	1622	6.47	1.82	0.31	0.19	nd	nd
P78.1_A	M, CNS, L, S, DD	5	1.30	0.08	16.25	CI, CIII, CIV: 28, 38, 54	746	1054	6.39	1.25	1.25	11.29	nd	nd
P82.1_A	M, CNS, Eye, G, ENT, DD	8	3.60	0.15	24.00	CI, CIII, CIV: 72,70, 65	2297	1739	31.15	8.41	0.14	0.33	nd	nd
P83.1_A	M, CNS, End, G, DD, DYS	6	2.30	0.14	16.40	CIII, CIV: 65, 93	1762	4222	3.29	5.94	10.22	0.67	0.14	0.12
P86_1_A	M, CNS, G, End, BE, DD, DR	8	3.70	0.20	18.50	CI, CIII, CII+III, CIV: 11, 65, 76, 39	1416	980	12.30	9.05	0.12	0.16	0.21	0.01
P87_1_A	M, CNS, Eye, Skin, DD, DR	8	1.90	0.17	11.20	CII+III: 86	580	3186	10.01	30.93	0.11	0.18	0.02	0.15

[1]Clinical profile includes: M, muscle involvement; CNS, central nervous system involvement; Eye, vision involvement; DD, developmental delay; DR, developmental regression; Dys, dysmorphism (minor and major); BE, behaviour and emotional abnormalities; ENT, sensori-neural deafness; PNS, peripheral neuropathy; G, gastro-intestinal tract involvement; R, renal involvement; Car, cardiac involvement; End, endocrine abnormalities; AID, auto-immune disorder; L, liver involvement; S, skeletal involvement.

[2]MDC score: Mitochondrial Disease Score (Wolf & Smeitink 2002). AA, amino acids; AcCar, acylcarnitines; CI-IV; complexes I-IV; CAR, carnitines; Crea, creatine; Cr: creatinine; L/P, lactate:pyruvate ratio; nd, not done; OA, organic acids; RC, respiratory chain; SD, standard deviation; nd: not detetmined.

2.6.1.3.4.2 Analyses of organic acids using gas chromatography–mass spectrometry

The organic acids were isolated from the urine, derivatized and separated by gas chromatography according to a procedure described previously (Reinecke *et al.* 2012). The volume of urine used for organic acid analysis was based on urinary creatinine values, transferred to silanized glass tubes (Kimax) and the internal standard (3-phenylbutyric acid) was added to a final concentration of 180 mmol/mol creatinine. The samples were acidified with 5N HCl to a pH less than 2, followed by the addition of 6 ml of ethylacetate to each sample which were then shaken on a rotary wheel for 20 minutes. After centrifugation of each mixture for 2 minutes at 1300 x g, the upper ethylacetate phase was transferred to a clean glass tube; 3 ml of diethylether was added to the water phase, shaken for a further 10 minutes and centrifuged at 1300 x g for 10 min. The upper phase was removed and added to the ethylacetate. A small amount of sodium sulphate (BDH) was added to the ethylacetate/diethylether mixture to remove any residual water. After a subsequent centrifugation step, the organic phase was transferred to a clean glass tube. The organic solvents were evaporated to dryness under nitrogen at 37 °C.
O-bis(trimethylsilyl)trifluoroacetamide (BSTFA): trimethylchlorosilane (TMCS): piridine (5:1:1, and volume added based on the creatinine values)) was used for derivatization. The volume of urine used gave a creatinine concentration equivalent to 21 µmol/ml derivatization reagent. The samples were derivatized at 85 °C for 45 minutes in a sand bath. The derivatized mixture was transferred to a 1.5 ml vial for GC-MS analysis. The Agilent GC-MS system used in this study consisted of a model 7890A gas chromatograph, a model 5975C mass selective detector, an HP 5970C MS and Agilent Chemstation (Revision E.02.00), and the GC-MS analysis was done as previously described (Reinecke *et al.* 2012). Peak identification and feature annotation was done by using AMDIS software (Version 2.66) linked to NIST Mass Spectral Search Program for the NIST/EPA/NIH Mass Spectral Library (Version 2.0F, built Oct. 8, 2008). The semi-quantitative identification of the organic acids was conducted according to Chen *et al.* (2009). All organic acids identified above the detection limit of the equipment used were expressed as mmol per mol creatinine.

Table 2-2: Summary of the urinary parameters for the respective controls (22/12) and patients (29/22).

Patients	Organic acids (mmol/mol Cr)	Amino acids (mmol/mol Cr)	Acylcarnitines (mmol/mol Cr)	Carnitines (mmol/mol Cr)	DC372/370	OC344/342	Creatine (mmol/mol Cr)	Betaine (mmol/mol Cr)
Minimum	579	694	1.80	0.5	0.06	0.07	0.02	0.01
Mean	1366	2479	9.83	24.2	0.62	1.33	1.08	0.15
Maximum	3217	9719	31.15	74.9	10.22	12.51	3.3	1.44
SD	724	2006	7.73	22.4	1.86	3.02	1.01	0.31
Controls								
Minimum	164	246	0.59	0.26	0.05	0.08	0.02	0.01
Mean	348	494	3.96	6.5	0.19	0.26	0.12	0.02
Maximum	565	882	20.54	10.6	0.41	0.72	0.5	0.03
SD	123	164	4.05	5.8	0.09	0.17	0.15	0.01
***p*-value**	**> 0.0001**	**> 0.0001**	**> 0.001**	**> 0.001**	0.232	0.070	**> 0.0001**	0.047

SD: standard deviation; DC372/370: decanoyl-carnitine:decenoyl-carnitine; OC344/342: octanoyl-carnitine:octenoyl-carnitine

2.6.1.3.4.3 Analyses of amino acids using gas chromatography–mass spectrometry

GC-MS analysis of the amino acids was conducted on an Agilent Technologies (Chemetrix, Midrand, South Africa) 6890 series GC system with an Agilent Technologies 5973 Mass Selective Detector and a 7683 series dual tower and autosampler, all controlled by the MSD ChemStation E.02.00 (Palo Alto, CA, USA). The amino acid standards and urine were prepared as prescribed by the suppliers of the EZ:faast^{TM} amino acid analysis sample testing kit. One hundred microlitres of internal standard (norvaline at 200 µM) and amino acid standards (10, 25, 50, and 100 µL of each standard at 200 µM) or 100 µL urine were combined in a glass vial and further procedures were conducted according to the method supplied with the testing kit. Two microlitres of the extracts prepared according to the prescribed method was injected into the GC-MS for analysis and also analysed according to the prescribed method. The standard range analysis was used to calibrate the identification and quantification of the amino acids, using the MSD ChemStation E.02.00 software with a linear regression curve fit.

2.6.1.3.5 Analyses of carnitines using tandem mass spectrometry

The electrospray ionization TMS method was used to quantify urinary acylcarnitines. Ten µL of urine was added to a 1.5-ml centrifuge tube before 400 µL of the deuterated acylcarnitines (internal standard solution) with the following concentrations was added: [methyl-d_3]-L-carnitine·HCl (30.45 µmol/l), [d_3]acetyl-L-carnitine·HCl (20.83 µmol/l), [3,3,3-

d_3]propionyl-L-carnitine·HCl (19.69 µmol/l), [d_9]isovaleryl-L-carnitine·HCl (17.73 µmol/l), [8,8,8-d_3]octanoyl-L-carnitine·HCl (15.43 µmol/l) and [16,16,16-d_3]hexadecanoyl-L-carnitine·HCl· (11.47 µmol/l). After the samples were evaporated to dryness under a gentle stream of nitrogen (55°C), the remaining procedures were followed as described by Mels *et al.* (2011). Acylcarnitines were quantified by comparing the signal intensities of carnitine and acylcarnitines against those of the corresponding deuterated analogues. The concentrations of carnitine and acylcarnitines analysed were expressed as mmol per mol creatinine.

2.6.1.3.6 Statistical analysis

Variables with no variation (e.g. the internal standards) were removed from the original data sets for the organic acids, amino acids and carnitines and each of these data sets was initially analysed separately to identify their role as potential biomarkers. In addition, a data filter, based on the approach of Bijlsma *et al.* (2006), was applied to each variable to eliminate those that contained more than 40% zero values ("60% rule") for the control and patient groups. Standard univariate analyses, including t-tests and the Mann-Whitney U-test, were applied to all the remaining variables after application of the 60% rule to assess the statistical significance of those variables that would eventually be considered as components of a putative biosignature. The subsequent data pre-treatment, in the first instance, consisted of zero replacement, where the zero values represented the detection limit of the analytical equipment. The zeros were replaced by a random sample of values from a Beta (0.1;1) distribution bounded between zero and the detection limit. Thereafter, a shifted logarithmic transformation with a shift parameter set at one was performed, ensuring that the scales of the various metabolite concentrations were more comparable, after which the transformed data were centred prior to further statistical analyses.

The effect size of each individual variable was measured to ascertain the importance of the single variables (Ellis & Steyn 2003). An effect size of $d > 0.5$ can be considered as being of medium practical importance, whereas an effect size of $d > 0.8$ can be considered as highly practically significant. Descriptive statistics, such as minimum and maximum values, means and standard deviations, were included as applicable.

Multivariate analyses used for the identification of important variables were principal component analysis (PCA), as an unsupervised pattern recognition method (Johnson & Wichern 1998), and a partial least squares discriminant analysis (PLS-DA) as a supervised method (Barker & Rayens 2003). Variables listed by the PCA with a modelling power greater than 0.5 were regarded as potential biomarkers (Brereton 2003); and for variables important in projection (VIPs) from PLS-DA, the 'greater than one rule' was used as the criterion for

variable selection (Chong & Jun 2005). The primary criterion for selection of important metabolites was that identified by PLS-DA, based on a VIP > 1.0 for each variable in the three data sets. The specificity and sensitivity estimates of the outcomes of the PLS-DA approach were evaluated by cross-validation as described below. Fit statistics of the PCA and PLS-DA models were reported as the percentage variance explained for the metabolites (R^2X), the percentage variance for the group membership of the patients and controls (R^2Y), and the predictive R^2Y values (Q^2).

A putative biosignature was derived from a consolidated data set, consisting of the important metabolites identified by the PLS-DA of the three MS-based analyses and the two important variables from the ^{1}H-NMR analysis. Because the scales of these four data sets were quite different, we compared various approaches of scaling for normalization, from which we selected the scale function provided in the R-statistical program, expressed as $Z = \log[X/\sqrt{\{1/(n-1)\Sigma X^2\}}]$. The scaled variables were then further transformed by using a shifted logarithmic transformation ($Y = \log[Z+1]$). The transformed data were subsequently centred prior to PLS-DA analysis, and the important variables identified and validated as described below.

A generic description of the cross-validation, which was constructed on the outcomes of the applicable PLS-DA models, includes the following aspects. A data set was constructed which included only the important metabolites that were identified. Next, a PLS-DA model was built for this data set and an appropriate cut-off point was determined by calculating the Youdin index (Fluss *et al.* 2005). Then, we let $\mathbf{P_{CON}}$ and $\mathbf{P_{PAT}}$ be the observed occurrence probabilities of a control and a patient, respectively, and let α be the fraction of cases to be removed in the cross-validation. Next, 10 000 unique stratified samples of size $n_{CV} = [\alpha \cdot n]$, with n = total number of controls and patients, were selected from the data, stratified according to the observed occurrence probabilities, that is, $n_{CV} = n_C + n_P$, where n_C and n_P are the sample sizes from the controls and patients, respectively. For each of the 10 000 samples, the n_{CV} cases were withheld, a PLS-DA model was built using the remaining cases and the group membership of the withheld cases was predicted. For this, the sensitivity and specificity as well as the percentage of misclassified cases were recorded. Lastly, the standard deviation and the average of the recorded information were calculated over the 10 000 samples and reported as cross-validated estimates of sensitivity, specificity and percentage of misclassifications, as well as the respective values for α and the cut-point.

2.6.1.4 Results and discussion

2.6.1.4.1 Profile of the control and patient groups

Table 2-1 summarizes the inclusion criteria and selected metabolomics data of the 29 RCD patients investigated. With regard to criterion 1 (5 aspects), assessments were based on a detailed history and clinical examination of all patients (Smuts *et al.* 2010), indicating that an intrinsic property of the selected patients was their predominant myopathic phenotype. Original baseline investigations included lactate (L), pyruvate (P), creatine kinase (CK), and ammonia (NH_3) determinations. Lactic acidosis was present in five (17%) and a raised pyruvate and lactate:pyruvate ratio (>18) in nine (31%) cases of the selected group. All patients had a deficiency in one or more complexes of the RC (criterion 2), established by enzyme assays of biopsy material; and no patients with a deficiency in the pyruvate dehydrogenase complex (PDH) were included in the group. The total excretion of organic acids (criterion 3) of the patients was statistically significantly increased relative to the controls [mean value of the 22 controls was 348.9 (SD=123.5) and 1336.6 (SD=724.5) mmol/mol creatine for the 29 patients with $p < 0.0001$ for the t- as well as the Mann Whitney U-tests]. No patients (minimum = 579 mmol/mol creatinine) or controls (maximum = 565 mmol/mol creatinine) were included in the group with an overlap in the total organic acid content, as shown in Table 2-2. With regard to the 7 sets of metabolomics data shown, statistically significant differences between the patients and controls were also found for the total amino acid excretion ($p < 0.0001$), total acylcarnitines ($p < 0.001$), free carnitine ($p < 0.001$), creatine ($p < 0.0001$) and betaine ($p < 0.047$). Although the mean values for the ratios of octanoyl-carnitine:octenoyl-carnitine and decanoyl-carnitine:decenoyl-carnitine were respectively three times and five times higher than the controls, these differences were not statistically significant (the *p*-values from the robust Mann Whiney U-test were 0.424 and 0.246 for these two ratios respectively), which coincides with the view that high values for the ratio of certain acylcarnitine esters may in certain cases be useful in supporting specific diagnosis (Haas *et al.* 2008). It has also been suggested that the severity of symptoms observed in mitochondrial fatty acid β-oxidation defects may correlate with the concentrations of accumulating acyl carnitines, with possible application in newborn screening programmes (Giak Sim *et al.* 2002). We therefore, finally, made a cluster and conducted other univariate analyses to investigate a possible correlation between some distinct clinical and other phenotypes of the patient group and the metabolite profiles described above. Comparison between phenotypes and metabolite profiles did not reveal any meaningful relationships. A way of improving genotype, phenotype and metabolite interrelationships might be to classify phenotypes in greater depth by also including

transcriptional information as reported for genetic networks in liver metabolism (Ferrara *et al.* 2008).

A distinct difference between the biochemical profile of the patients and the controls, as shown in Table 2-2, is an important point of departure for metabolomics investigations. This was substantiated by a PCA conducted on the 29 patients and 22 controls for all the original 291 variables measured in the organic acid, amino acid and acylcarnitine (including free carnitine) analysis (Figure 2-1). The outcome of the PCA, shown as a two-dimensional (PC1 and PC2) score plot for all the cases, indicates that the patient group was distinguished from the controls. This reveals that the metabolic profiles of the two groups were distinctly different due to the perturbation induced by the respective CI, CIII or CM deficiencies. Moreover, the heterogeneity, which is characteristic of RCDs, was retained in the patient group, as shown by the spread of these cases in the PCA. These observations already had the potential to identify biomarkers that could distinguish RCDs, but required data reduction for this identification.

2.6.1.4.2 Identification of important metabolites

The work-flow followed to identify important metabolites is shown schematically in Figure 2-2 NMR-based metabolic profiling enables the simultaneous examination of a complex mixture of metabolites in a biological sample and requires only a limited knowledge of sample composition prior to analysis. NMR metabolomics may thus be regarded as an untargeted mode of analysis. By contrast, MS-based analyses are mostly semi-targeted as they distinguish a specific sub-section of the metabolome, extracted from a biofluid by an appropriate analytical procedure. Thus, as shown in Figure 2-2, we included both these analytical approaches in our metabolomics investigation to optimize the detection of possible biomarkers for RCDs.

As indicated, urine samples from only 20 patients and 14 controls, which included information on the urinary organic acids, amino acids and acylcarnitines, were available for the ^{1}H-NMR analysis. Although these cases were fewer than the 51 cases of the total group, the clinical profile of the 34 cases strongly resembled that of the group of 51 patients. Alanine, lactic acid, succinic acid, creatine and betaine were found to be the important variables that distinguished the patient and control groups. As the first three of these are included in the MS-based analysis, only the values obtained for creatine and betaine were considered for the final consolidation of all important variables identified by the different analytical approaches.

Identification of important metabolites from GC-MS and TMS data required preprocessing to generate a data matrix of variables and cases of an operational size, to be followed by multivariate analyses to identify only the relevant analytical information. The total number of features in the original data set of the 51 cases, generated by an untargeted analysis in each of the three metabolite groups, yielded 291 substances that could be annotated as metabolites, namely, 189 organic acids, 51 amino acids, 50 acylcarnitines and free carnitine. Using the data filter, these compounds were reduced to 120 metabolites: 39 organic acids, 36 amino acids, 44 acylcarnitines and free carnitine. With regard to the long-chain acylcarnitines, it should be noted that they are strongly protein-bound in the plasma and thus escape excretion into the urine, as do the free fatty acids. Their presence in the urine in appreciable amounts may thus be due to renal malfunctioning or damage resulting in proteinuria and the relatively elevated amino acid excretion found in the RCD patients.

The subsequent data pretreatment first included zero replacement and logarithmic scaling. Mean values, as well as the standard deviation of all variables for the controls and patients, were determined on the unscaled data, followed by t-test and Mann-Whitney analyses. The two traditional methods of multivariate analysis chosen (PCA and PLS-DA) proved to be valuable for selection of variables and were subsequently applied to all three data sets, followed by effect size analyses. All variables with a VIP > 1.0 and/or a power value > 0.5 and an effect size > 0.8 were designated as important metabolites due to the RCDs in the patient group. A total of 26 metabolites was identified by this selection method, and included 11 organic acids (adipic, fumaric, homovanillic, lactic, suberic, succinic, vanilmandelic, 2-hydroxyglutaric, 3-hydroxyisobutyric, 3-hydroxyisovaleric, and 3-hydroxy-3-methylglutaric acids), 13 amino acids (alanine, asparagine, aspartic acid, glutamic acid, glutamine, glycine, lysine, proline, serine, threonine, tyrosine, α-amino adipic acid and β-alanine), acetyl carnitine and free carnitine. Thus, from the ^{1}H-NMR-based and the MS-based analyses a total of 2 + 26 = 28 metabolites were identified as important indicators of RCDs, from which a final list of biomarkers was selected and validated according to the cross-validation procedure described in the statistical methods section.

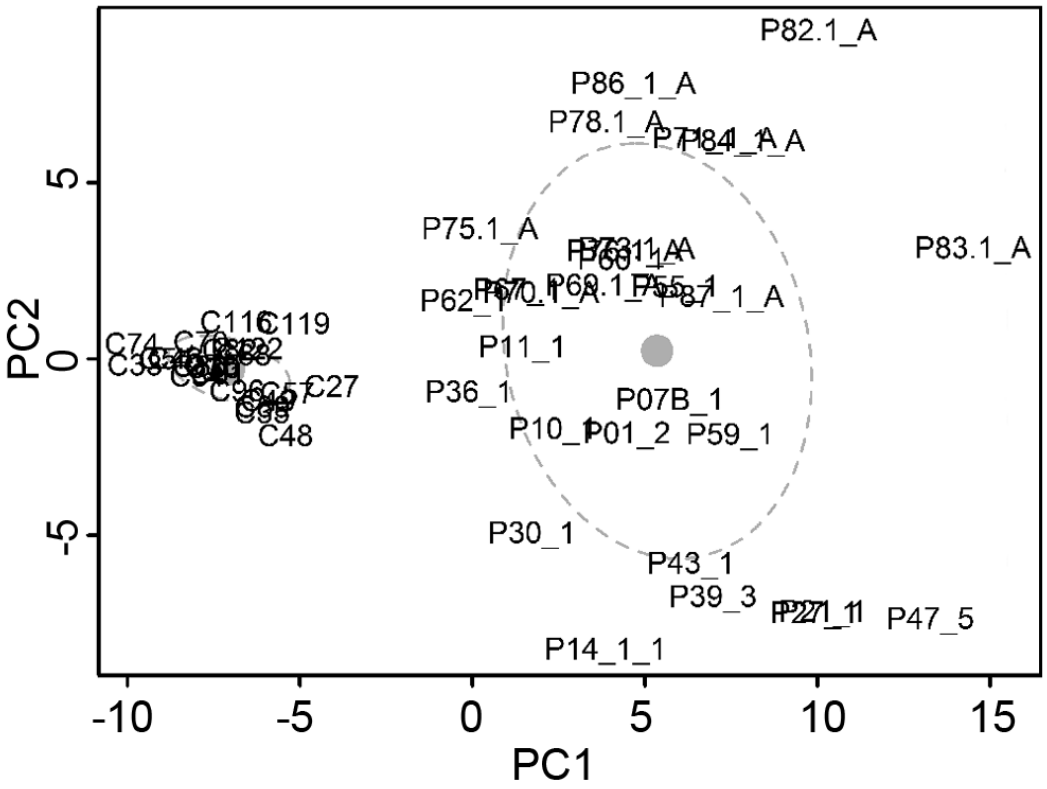


Figure 2-1: **Two-dimensional principal component analysis of the controls (indicated by a C and the case number) and patients (indicated by a P and the case number).**

This analysis was based on all 291 variables present before data reduction. The circles were drawn to indicate a 50% probability level and the averages of the group scores are indicated by the solid dots. Owing to the density of the data references of the controls, the corresponding dot and most of the circle are obscured (see grouping in left-hand side of figure). Principal component 1 explained 35.7% of the total variance whereas principal component 2 explained 10.58% of the variance.

The outcome of the cross-validation for the two experimental groups (51 or 34 cases, respectively) is shown in Table 2-3. The cut-off points for the metabolite groups were determined for each group separately; the differences relate to the numerical characteristics of the data sets for these variables. The sensitivity refers to the percentage of patients in the experimental group who were correctly classified as such by using the important metabolites identified from the organic acids (11), amino acids (13), carnitine (2) and creatine plus betaine, respectively. The specificity relates to the ability of the selected metabolites to

identify the controls. The percentage misclassification includes the results obtained for the patients and controls taken together. The value of 100 obtained for the selectivity and the specificity for the organic acids clearly relates to the selection of the control and patient groups on the basis of a complete separation of the total urinary organic acids excreted by the groups (criterion 3). From the misclassification outcome it is clear that the ranking of importance of the metabolite groups is organic acids ≈ amino acids > creatine and betaine > carnitines, with the respective percentage of misclassifications being 0%, 3.08%, 16.58% and 26.64%, respectively. A comparable ranking was obtained for the outcome of the sensitivity and specificity measures. The final conclusion from these cross-validations is that all 26 important metabolites from the three MS-based analyses should be included in a consolidated matrix, with the two metabolites identified by ^{1}H-NMR analysis. From this matrix a biosignature for the group of RCD patients could then be constructed.

2.6.1.4.3 Identification of a biosignature for the RCD patient group

First, the consolidated data set of 28 metabolites was formed, followed by data pretreatment as described above. Subsequently, a PLS-DA model was constructed for this data set to identify the metabolites that could qualify for a biosignature for the group of RCD patients. Sixteen metabolites with a VIP > 1.0 were identified as possible components of a biosignature, of which 13 are eventually summarized in Table 2-4, following exclusion of three of the original 16 as indicated below.

The RC is essentially involved in cellular reduction/oxidation (redox) status and energy (ATP) production; deficiencies in any component of this supramolecular complex inevitably affect a wide array of metabolic and other processes (Reinecke *et al.* 2009; Elstner & Turnbull 2011). Eight of the organic acids could accordingly directly be related to a consequence of RCDs and were included in the biosignature. Vanilmandelic acid (VMA) and homovanillic acid (HVA) were excluded from the biosignature because of their properties as indicators of neurological stress conditions (Frankenhaeuser *et al.* 1986; Rauste-von Wright & Frankenhaeuser 1989), rather than being specifically related to RCDs. Betaine, the final component, was not included into the biosignature, because of a *p*-value ($p = 0.047$) on the borderline between significant and not significant. Thus, 13 components can be related to RCDs, and are shown in Table 2-4 as part of the putative biosignature for RCD.

An important consequence of RCDs is a relative increase in levels of NADH (NADH/NAD^+ ratio) and $FADH_2$, as well as decreased ATP production, which may result from a defect at any site within the RC and lead to the well-established elevations in lactic acid and the lactate/pyruvate ratio. Increased succinic acid, 3-hydroxyisobutyric acid, 3-hydroxyisovaleric

acid, 3-hydroxy-3-methylglutaric acid and 2-hydroxyglutaric acid were all reported in a metabolomics investigation on global changes in organic acid metabolism (Reinecke *et al.* 2012).

Amino acids and solutes such as bicarbonate, phosphate and glucose are transported across the apical membrane of the proximal renal tubular cells. This transport is driven by the sodium gradient, which is established by the basolateral ATP-dependent sodium pump. A disruption of the ATP supply in the kidney is therefore likely to occur in a multisystem disease, such as RCD (Martín-Hernández *et al.* 2005), or might even occur owing to a single enzyme defect, as in methylmalonyl-CoA mutase (EC 5.4.99.2) deficiency (Morath *et al.* 2008). Disrupted ATP supply leads to renal dysfunction as part of the pathophysiological profile seen in these metabolic diseases. A related perturbation might exist in the present patient group, reflected by the presence of carnitine as well as the increased aminoaciduria, which resembles a Fanconi-Bickel excretion pattern of these metabolites (Odièvre *et al.* 2002). However, some amino acids may also increase as a response to other primary and secondary abnormalities due to RCDs. These amino acids include alanine, which follows from an increase in pyruvate and its consequent transamination, as well as glutamic acid that results from elevated amino acid catabolism (indicating also the possible hyperammonemia in RCD disorders) and tyrosine due to underlying liver damage (Levine & Conn 1967). Among the amino acids, α-aminoadipic acid has not been described for RCDs before, and clearly reflects a deficiency in lysine catabolism due to high FADH concentrations.

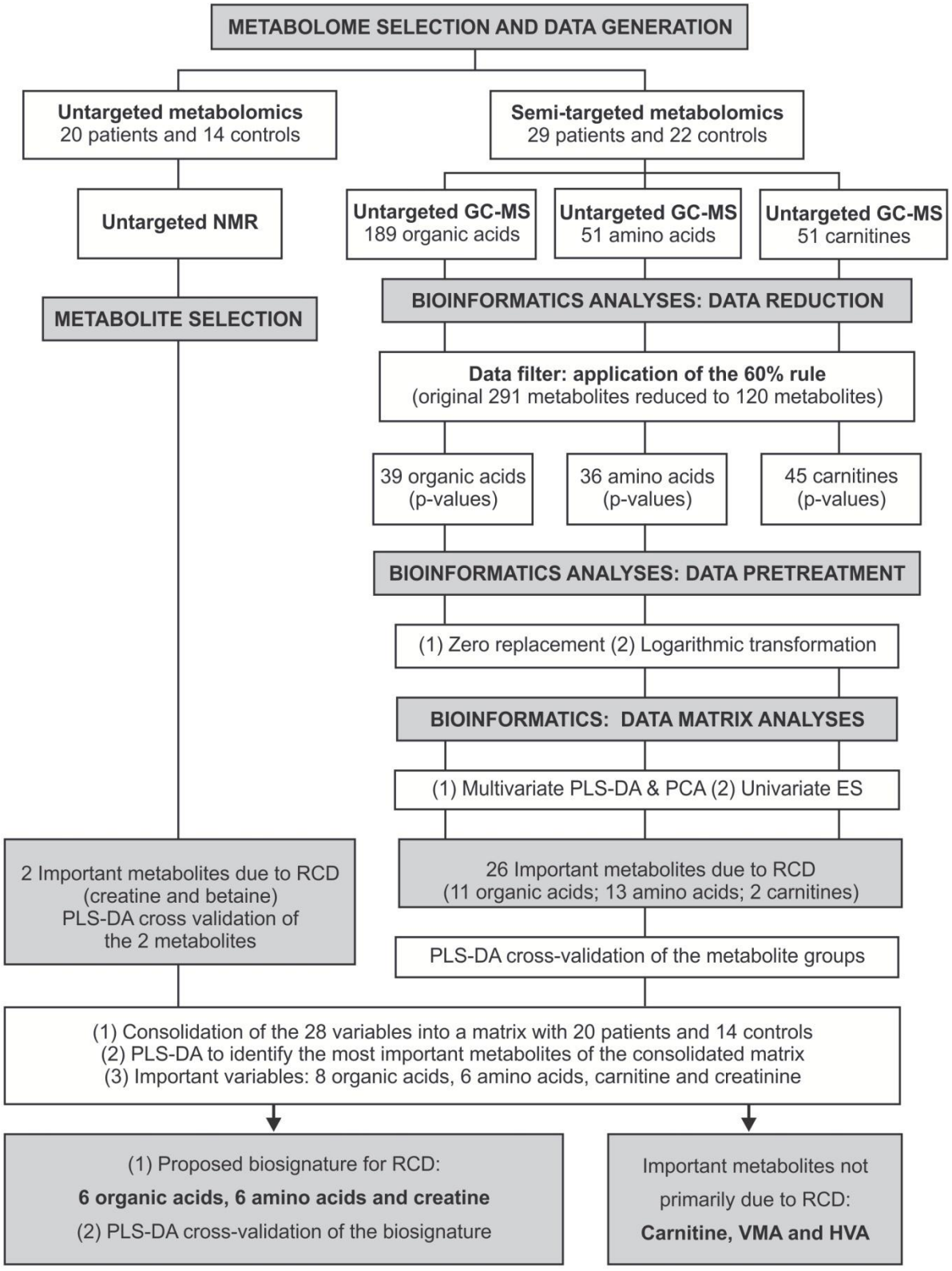


Figure 2-2: Metabolomics work-flow and cross-validation of metabolite groups and the biosignature.

Table 2-3: Cross-validation of individual metabolite groups and of the biosignature.

Number of cases (validation size)	Metabolite class (number)	Cut-off points	Sensitivity mean (SD)	Specificity mean (SD)	% Mis-classification (SD)
Cross-validation of individual groups of metabolites					
51 (n_C = 7, n_P = 9)	Organic acids (11)	0.11	100 (0)	100 (0)	0 (0)
51 (n_C = 7, n_P = 9)	Amino acids (13)	0.17	96.19 (6.67)	97.83 (5.67)	3.09 (4.02)
51 (n_C = 7, n_P = 9)	Carnitines (2)	0.06	73.78 (14.19)	72.79 (16.30)	26.64 (9.25)
34 (n_C = 6, n_P = 8)	Creatine and betaine	-0.01	72.95 (16.11)	97.39 (6.26)	16.58 (9.01)
Cross-validation of the biosignature					
34 (n_C = 6, n_P = 8)	Organic acids (6), amino acids (6) and creatine	0.23	98.12 (4.88)	97.96 (6.40)	1.95 (3.67)

n_C and n_P are respectively the sample sizes from the controls and patients used for validation.

Table 2-4: The proposed biosignature.

Metabolite	VIP	ES	C[Mean]	SD	P[Mean]	SD	P/C	t-value	*p*-value
lactic acid	1.15	2.2	3.2	2.7	65.1	95.2	20	+3.51	< 0.001
succinic acid	1.29	2.1	6.0	5.9	97.5	108.6	16	+4.53	< 0.001
2-OH-glutaric acid	1.26	2.6	1.5	1.3	15.3	13.1	10	+5.65	< 0.001
3-OH-isobutyric acid	1.39	2.3	2.8	2.9	26.8	17.5	10	+7.27	< 0.001
3-OH-valeric acid	1.35	2.5	3.7	2.8	42.1	51.1	11	+4.03	< 0.001
3-OH-3-me-glutaric acid	1.18	1.8	1.4	2.4	17.8	17.4	13	+5.02	< 0.001
alanine	1.05	1.9	22.9	11.3	196.9	213.4	9	+4.38	< 0.001
glycine	1.06	1.9	93.6	45.6	638.8	622.1	7	+4.70	< 0.001
glutamic acid	1.01	1.8	4.7	1.9	38.7	39	8	+4.69	< 0.001
serine	1.00	2	35.9	11	216	188.1	6	+5.14	< 0.001
tyrosine	1.04	2	14.5	5.6	64.8	45.6	4	+5.89	< 0.001
α-aminoadipic	1.04	1.7	2.5	1.5	34.6	42.7	14	+4.04	< 0.001
creatine	1.11	0.94	0.12	0.15	1.08	1.02	9	+4.11	< 0.001

P/C designates P[Mean] / C[Mean], that is, the mean values per metabolite for the patients and controls respectively; ES: effect sizes. VIP: variables important in projection, derived from the partial least-square discriminant analyses. The *p*-values of the t-test are shown in the table; all *p*-values of the Mann-Whitney analyses for the metabolites of the biosignature were below 0.0001 and are not included in the table.

We did not compare biomarkers that discriminate between mitochondrial and other myopathies. Urinary glycine ($p < 0.001$), creatine ($p < 0.0001$) and betaine ($p = 0.047$) were significantly elevated in our patient group relative to controls, and were reported also to be significantly increased [in the case of glycine ($p < 0.01$), creatine ($p < 0.001$) and betaine ($p = 0.001$)] in juvenile idiopathic inflammatory myopathy patients relative to control subjects (Chung *et al.* 2005). Elevated creatine in plasma was recently described for RCDs, by using the phosphocreatine shuttle, as a consequence of tissues in a low energy state (Shaham *et al.* 2010). Furthermore, several of the metabolites that can be attributed to increased catabolism of fatty acids and amino acids share a bioenergetics-sensing (hormone-modulated) induction pathway with FGF-21, which is also associated with a muscle disease phenotype response (Suomalainen *et al.* 2011).

From the metabolomics and statistical analyses, as well as from the biochemical considerations discussed here, the proposed biosignature for our experimental group consisted of 6 organic acids, 6 amino acids and creatine, as shown in Table 2-4.

2.6.1.4.4 Specifications for a biosignature

It has been proposed that the specification for a single metabolite (a biomarker), or a combination of metabolites (a biosignature), is the requirement to assign an individual patient to a unique group with defined characteristics (Jacobsen *et al.* 2008). The evaluation of a biosignature thus requires the use of a data set to validate the capacity of a putative biosignature to classify individual samples correctly. The data set from which a biosignature is defined may also be used for the validation, but an independent data set should preferably be used for this purpose. For an inherited metabolic disease, the latter can be generated only over a period of time or by the creation of a data set through information gathered from participants at several medical centres, as was recently reported for FGF-21 as a potential biomarker for an RCD (Suomalainen *et al.* 2011). In our investigation the original data sets had to be used to validate the RCD biosignature, as an independent data set was not available for this purpose.

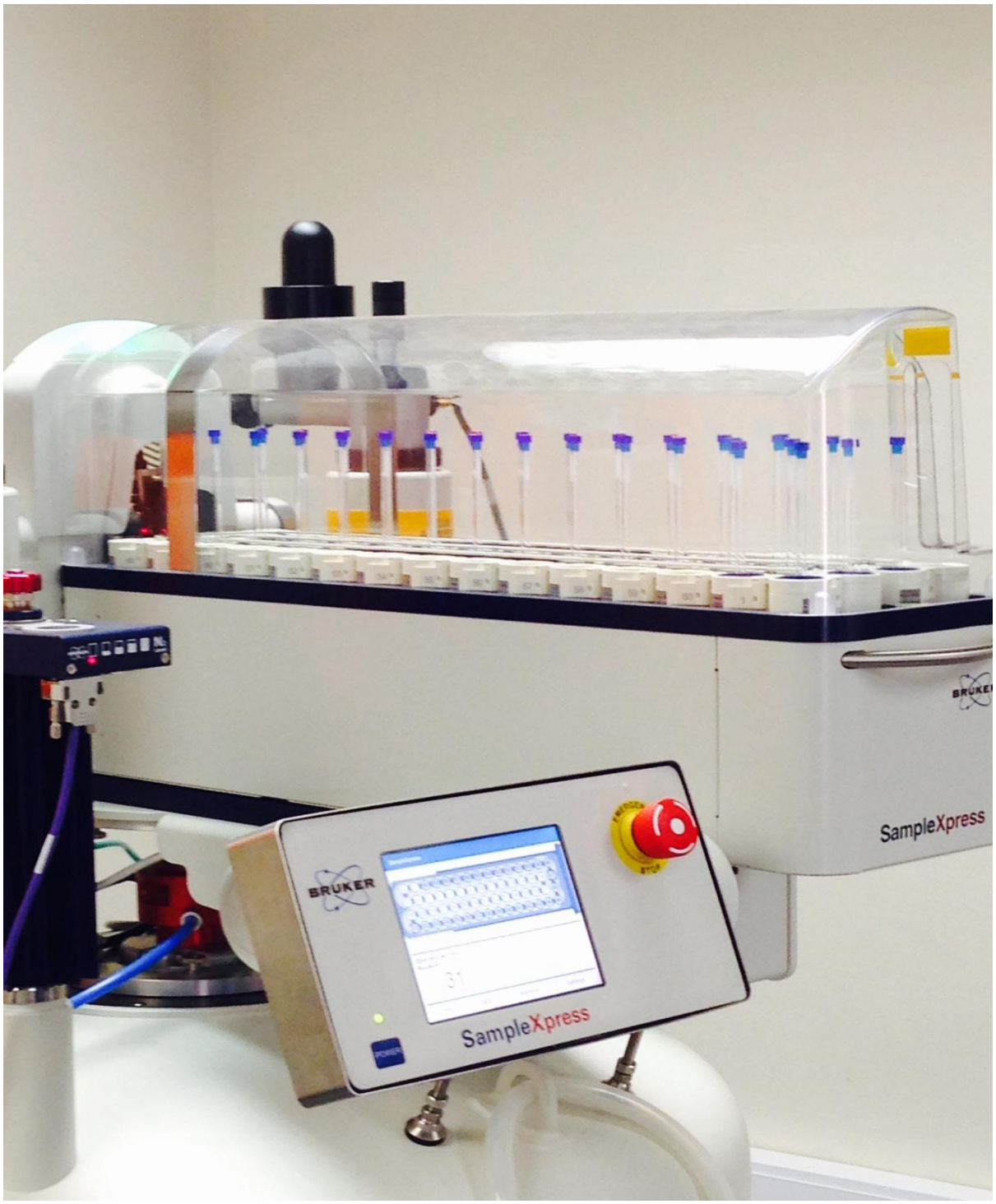
The cross-validation procedure described in the statistical section was used for the validation of the biosignature, and the outcome is summarized in Table 2-3. This validation was conducted for the data set consisting of the 13 metabolites (Table 2-4) and the 20 patients and 14 controls used in the ^{1}H-NMR analysis. The cut-off point for the cross-validation of the biosignature shown in Table 2-3 was determined for the consolidated set of variables. The cross-validation of the biosignature indicates its advantage as an indicator of an RCD compared with the use of a limited number of metabolite markers. Using the biosignature for

the larger group of patients, a separation between the controls and patients with a CI, CIII or CM-deficiency could be obtained by unsupervised PCA as well as comparison of supervised PLS-DA. Thus, the results presented here give proof-of-concept that metabolomics investigations can include inherited metabolic diseases in their field of investigation.

2.6.1.4.5 From metabolomics to the clinic

According to Mancuso *et al.* (2009), the requirements for an ideal biomarker for a metabolic disorder such as an RCD are that it should improve the timing and accuracy of diagnosis, minimize the invasive procedure needed for the final diagnosis, and be useful to monitor disease progression and efficacy of treatment. They concluded, however, “that to date, no one can bet on this, but we are all looking forward to find it”. The translation of research findings into clinical practice is not straightforward (Hu, 2011) but the criteria to be satisfied for a biomarker or biosignature of an inherited metabolic disease to become a practical and useful instrument in a clinical setting are clear, although complex (Turnbull, 2011). For RCDs, the inductive approach (Kell 2004) to define a biosignature for the present experimental group, and its successful validation through the method of cross-validation opens up the possibility of formulating a hypothesis for the further development of a biosignature for RCDs. We postulate that the composite of organic acids ≈ amino acids > creatine > betaine > carnitines represents the basic biosignature for RCDs. The experimental approach to test the hypothesis and define the suite of organic and amino acids of a validated and consistent/specific biosignature could be the subject of a future study that includes a cohort of more cases than those used in the present study, as well as additional controls of related but different mitochondrial disorders. The development of a sensitive and specific biosignature may well prove to be an essential step in selecting patients for more invasive and complex diagnostic procedures, and the availability of such a biosignature could influence or even eventually change clinical practice with regard to RCD diagnosis and monitoring of treatment.

PART 2: UNTARGETED NMR METABOLOMICS: TOWARDS HYPOTHESIS GENERATION – AN INDUCTIVE ENDEAVOUR INTERTWINED WITH DEDUCTIVE ASPIRATIONS



CHAPTER 3 AN ORIENTATION ON PROTON NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY OF BIOFLUIDS

3.1 Brief theoretical background

The term 'nuclear magnetic resonance' provides a starting point for the explanation of ^{1}H NMR – employed in this PhD study. In essence, it refers to the application of a constant magnetic field and manipulation of the nuclei of protons through the application of a radio frequency (RF) pulse that resonates with the protons in the research material of interest. The applied magnetic field orientates the nuceli in the experimental sample in a particular direction (i.e., all vertically, pointing either in a north–south direction or in a south–north direction). This allows the RF pulse to hit the nuclei at precisely 90 degrees. In this study, each 'scan' consists of a 4-microsecond pulse which causes the aligned nuclei to be 'knocked over'. Since charged particles typically possess a spin, the 'knocked over' nuclei then return to their original state by means of a precessional orbit (Figure 3-1), releasing the energy applied to the nuclei by the RF pulse.

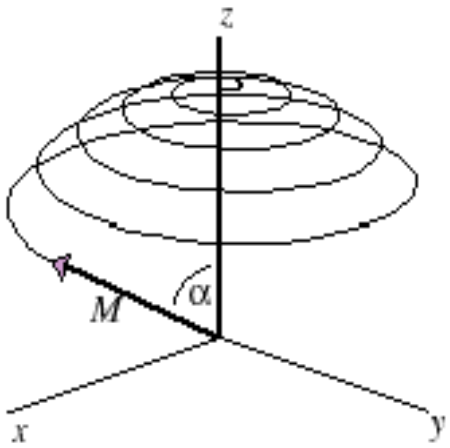


Figure 3-1: Schematic illustration of a precessional orbit of a nucleus whereby the direction of the magnetic field (M) returns to its original relaxed state.

The 4-microsecond pulse is typically followed by a delay, in which the release of energy is measured in the form of a free induction decay (FID), completing a single 'scan'. This FID is a measure of the intensity of protons of specific compounds in the sample and is converted from a time domain to a frequency domain by means of a Fourier transformation (Figure

3-2), which yields the familiar NMR spectrum, in which each peak represents an intensity of protons for a particular constituent of a compound. The greater the number of scans performed, the greater the signal to noise ratio of the NMR spectrum.

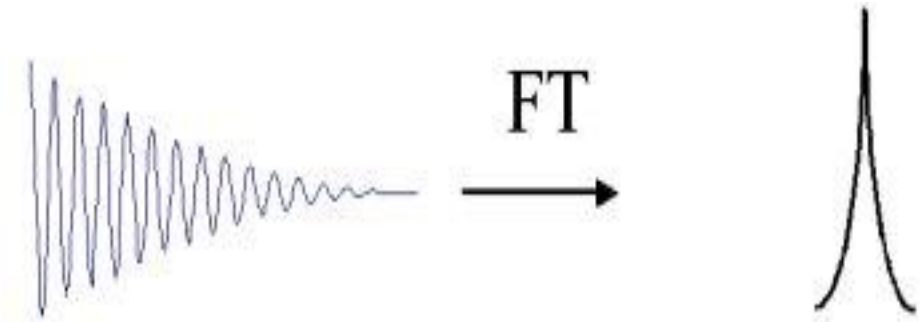


Figure 3-2: Conversion of free induction decay from time domain to frequency domain by means of Fourier transformation (FT).

3.2 Sample preparation of biofluids for NMR analysis

One of the great advantages of NMR metabolomics is the simplicity of sample preparation. First, for the purpose of analysing small biomolecules, the sample is centrifuged to remove macromolecules. In the case of biofluids with high protein content, such as blood and CSF, this centrifugation step is typically performed in a filtration unit, such as one with a 10 000-dalton filter membrane, as used in this study. This action removes any macromolecules, such as proteins in particular, which would otherwise cause peak broadening in the NMR spectrum.

Following centrifugation, a specific volume of supernatant is removed, to which a defined volume of an internal standard solution is added. This standard solution contains a single compound, such as trimethylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP) as used in this study, which acts as the source of a reference peak for the quantification and scaling of multiple NMR spectra for qualitative comparisons, together with a deuterated solvent, such as deuterium oxide (D_2O), which serves to lock the RF signal during analysis. An important point to pay attention to in ^1H NMR is the pH of the sample solution; an acidic solution results in chemical compounds being more protonated, resulting in a different chemical shift on the NMR spectrum when compared to a more basic solution, which tends to be less protonated. The matter of pH can be addressed in two ways: 1) after preparing the sample for analysis, the pH of each sample for the batch can be adjusted within a certain confidence

interval (e.g., pH 2.50 ± 0.05), or 2) the internal standard added to the sample can be prepared as a buffer solution, thereby stabilizing the pH of each sample at the time of preparation. Finally, the sample is placed in the NMR spectrometer for analysis. Figure 3-3 shows examples of NMR spectra for urine, serum and CSF taken from healthy individuals. Qualitative inspection indicates that there is a greater abundance of certain metabolites in serum and CSF; not clearly visible, unless the relevant part of the spectrum is expanded, is that urine contains far more peaks but corresponding to lower intensities than the other fluids. This is expected as urine is the outlet for all small molecules filtered by and excreted from the body.

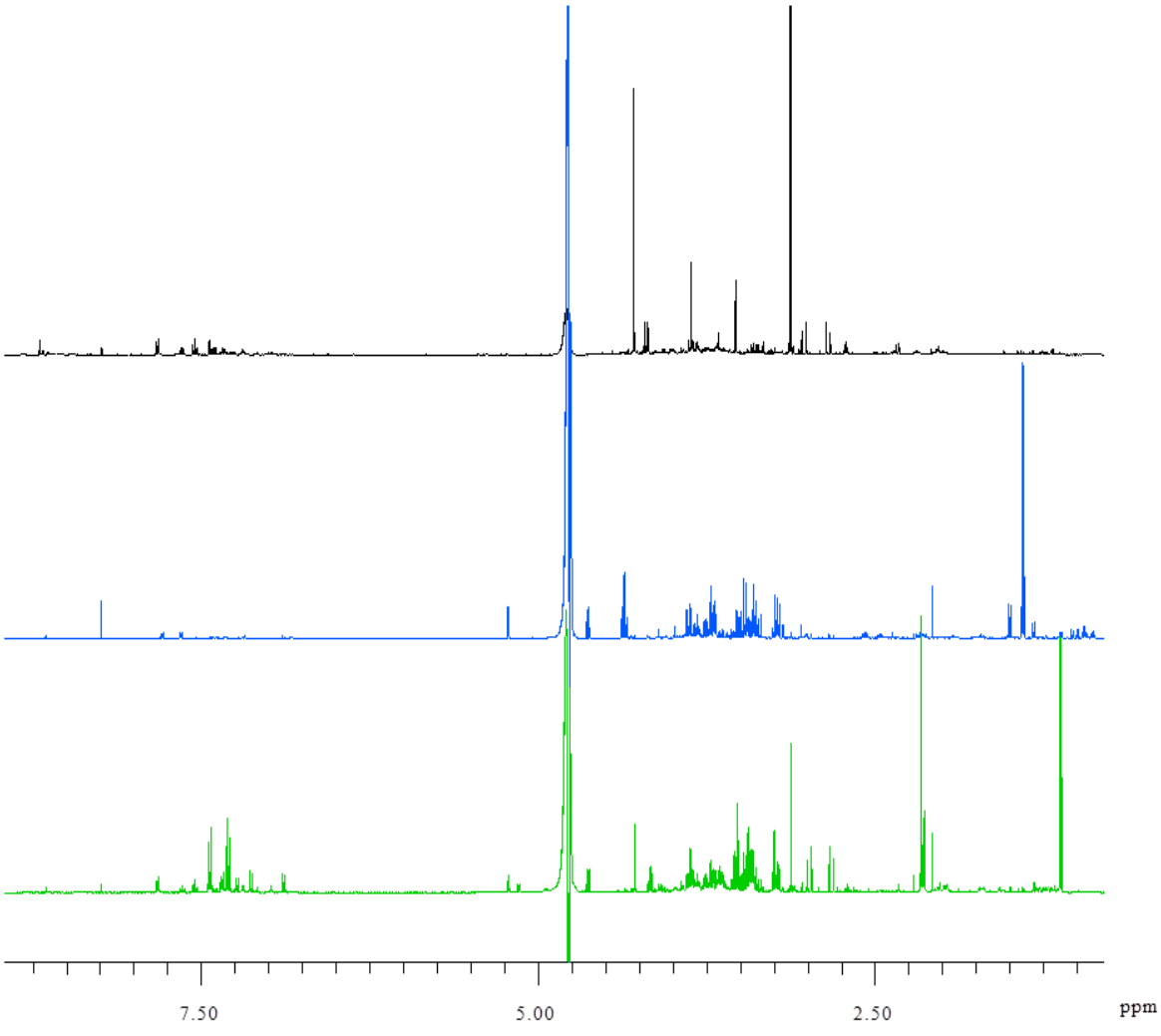


Figure 3-3: **NMR spectra for urine (top), serum (middle) and CSF (bottom) taken from healthy individuals and scaled according to the reference peak — TSP, at 0.00 ppm (not shown).**

3.3 Ensuring minimal technical variation

One of the unique and most important characteristics of NMR is that for each sample placed in the instrument's autosampler for batch analysis, the parameters for the NMR spectrometer are calibrated prior to analysis to ensure minimal technical variation caused by the characteristics of the machine. This calibration step involves 1) tuning the probe to resonate at the precise frequency of that particular spectrometer (e.g., 500 MHz), 2) locking onto the deuterated solvent in the sample, 3) ensuring that the magnetic field is accurately defined along all axes (a process known as shimming), 4) checking to ensure that the RF pulse is calibrated correctly, and 5) maintaining a constant temperature within the instrument. Since each sample has to undergo this calibration procedure, quality control samples (described in Chapter 7) that are necessarily used in GC-MS to detect possible trends over time, are not necessary for NMR. To demonstrate this, a pooled urine sample was created from an unrelated alcohol study and from this three quality control urine samples were prepared and analysed periodically over five weeks by an independent NMR analyst, resulting in a total of 72 samples. For each sample 28 pre-selected regions of the corresponding spectra representing clearly defined metabolites were integrated and their overall means, variances and, consequently, coefficients of variation (CV) were calculated. The overall CV value, presented as a percentage, for each metabolite is shown in Figure 3-4, which demonstrates that the CV did not exceed 8% for any of the 28 metabolites — a clear indication of the high repeatability of NMR.

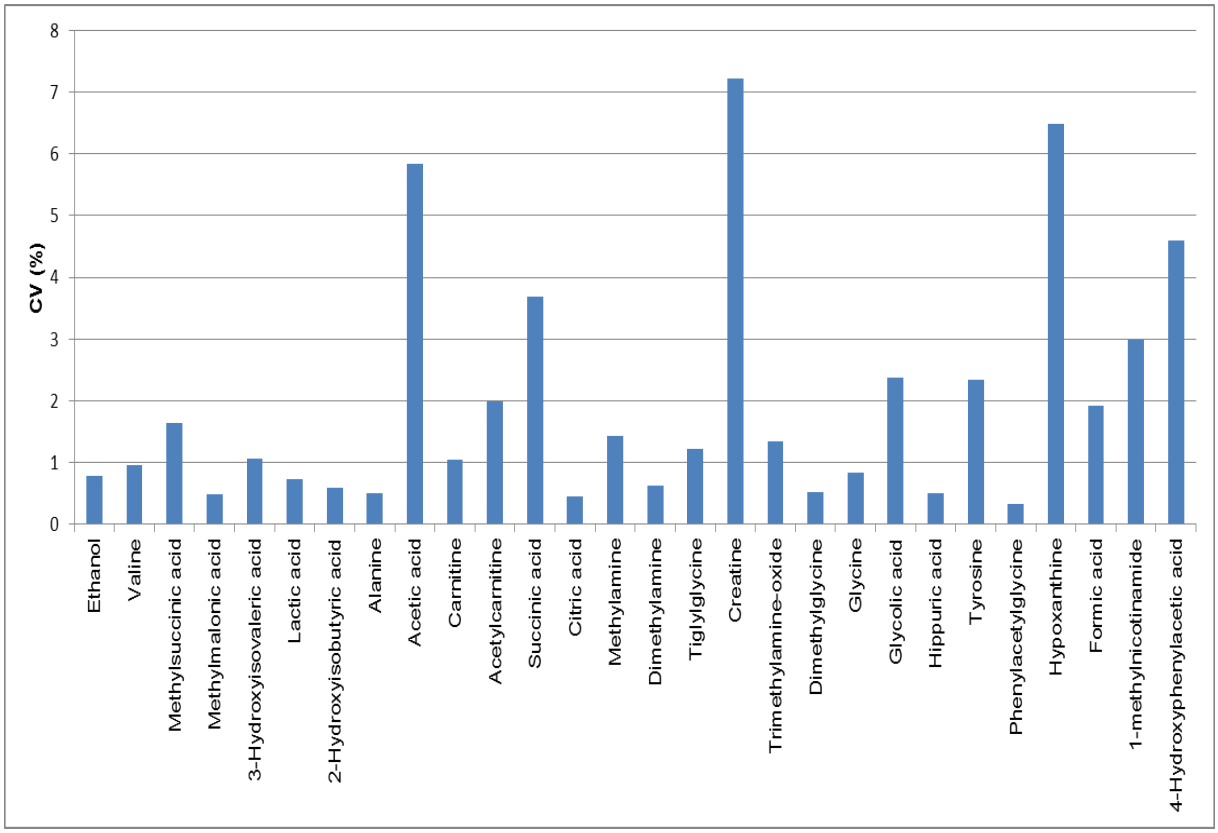


Figure 3-4: **Coefficients of variation, expressed as a percentage, of 28 pre-selected metabolites from 72 samples measured over a 5-week period.**

3.4 Quality of NMR spectra

Since the main constituent of all biofluids is water, the water in the NMR spectrum needs to be irradiated with a predetermined amount of energy during the analysis, thereby suppressing its otherwise large peak so that other, smaller peaks are clearly distinguishable. The disadvantage of this necessary procedure is that peaks neighbouring the water peak then become difficult to distinguish, and thus to quantify. Also, while the NMR spectra of each chemical compound are distinct, individual peaks often overlap. In such cases various other steps can be taken to identify the component peaks, such as by analysing at a different pH or performing additional two-dimensional analyses (e.g., 2D homonuclear J-resolved spectroscopy (JRES), 2D correlation spectroscopy (COSY) and 2D total correlation spectroscopy (TOCSY)).

After NMR analysis, each sample is pre-processed to ensure that the phase and baseline are corrected and the respective spectrum is calibrated by setting the TSP peak to 0.00

ppm. This ensures that the peaks of every NMR spectrum are aligned and allows identification of metabolites with a high degree of accuracy and specificity. These steps are necessary in order to facilitate the high quality of the spectra for further analysis.

3.5 Concluding remarks

^{1}H NMR is a highly repeatable method, useful as an initial, holistic means of assessing the perturbation under investigation in a metabolomics study. While less sensitive than MS-based metabolomics techniques, the unique nature of ^{1}H NMR allows for the determination of compounds in biofluids that are not detectable by means of either GC-MS or LC-MS; the converse is also true. Thus, ^{1}H NMR has its advantages and disadvantages (discussed in more detail in Chapter 5) and it should form a crucial part of any well-designed metabolomics study as it not only provides a 'first glance' but is complementary to other techniques, ultimately revealing a greater view of the total metabolome, as illustrated in the template paper (Chapter 2 – Smuts *et al.* 2013).

Since limited in volume, the valuable CSF samples collected for this PhD study were therefore first analysed via NMR as this method is nondestructive and involves minimal sample preparation — allowing the samples to be re-used for GC-MS analysis. The aim of the NMR approach was to determine: 1) what type of global host metabolic perturbations were discernible near the site of infection, 2) whether a unique neurological biosignature was evident and, based on experimental evidence, 3) whether biologically valid hypotheses could be formulated. This line of thinking was applied in Chapter 4 (Mason *et al.* 2015) and expanded upon in Chapter 5 (Mason *et al.* 2016a).

CHAPTER 4 A HYPOTHETICAL ASTROCYTE–MICROGLIA LACTATE SHUTTLE DERIVED FROM A ^{1}H NMR METABOLOMICS ANALYSIS OF CEREBROSPINAL FLUID FROM A COHORT OF SOUTH AFRICAN CHILDREN WITH TUBERCULOUS MENINGITIS

Shayne Mason[1], A. Marceline van Furth[2], Lodewyk J. Mienie[3], Udo F.H. Engelke[4], Ron A. Wevers[4], Regan Solomons[5], Carolus J. Reinecke[1]

[1]Centre for Human Metabonomics, Faculty of Natural Sciences, Private Bag X6001, North-West University (Potchefstroom Campus), South Africa,
[2]Department of Paediatric Infectious Diseases–Immunology and Rheumatology, Vrije Universiteit Medical Centre, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands,
[3]Potchefstroom Laboratory for Inborn Errors of Metabolism, Division for Biochemistry, Private Bag X6001, North-West University (Potchefstroom Campus), South Africa,
[4]Radboud University Nijmegen Medical Centre, Department of Laboratory Medicine, PO Box 9101, 6500 HB Nijmegen, The Netherlands,
[5]Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 19063, Tygerberg 7505, South Africa.

Accepted: *Metabolomics,* 2015, 11:822–837.

4.1 Abstract

Tuberculosis meningitis (TBM) is the most severe form of extra-pulmonary tuberculosis (EPTB) and is particularly intense in small children; there is no universally accepted algorithm for the diagnosis and substantiation of TB infection, which can lead to delayed intervention, a high risk factor for morbidity and mortality. In this study a proton magnetic resonance (^{1}H NMR)-based metabolomics analysis and several chemometric methods were applied to data generated from lumber cerebrospinal fluid (CSF) samples from three experimental groups: 1) South African infants and children with confirmed TBM, 2) non-meningitis South African infants and children as controls, and 3) neurological controls from the Netherlands. A total of 16 NMR-derived CSF metabolites were identified, which clearly differentiated between the controls and TBM cases under investigation. The defining metabolites were the combination of perturbed glucose and highly elevated lactate, common to some other neurological disorders. The remaining 14 metabolites of the host's response to TBM were likewise mainly energy-associated indicators. We subsequently generated a hypothesis expressed as an "astrocyte–microglia lactate shuttle" (AMLS) based on the host's response, which emerged from the NMR-metabolomics information. Activation of microglia, as implied by the AMLS hypothesis, does not, however, present a uniform process and involves intricate interactions and feedback loops between the microglia, astrocytes and neurons that hamper attempts to construct basic and linear cascades of cause and effect; TBM involves a complex integration of the responses from the various cell types present within the CNS, with microglia and the astrocytes as main players.

4.2 Introduction

Tuberculosis (TB), one of the major contemporary pandemics, is caused by *Mycobacterium tuberculosis* (Mtb); its worldwide impact remains a serious concern. The disease in children is particularly severe as diagnosis is difficult since its identification in children usually results from a combination of, often asymptomatic, clinical criteria, non-specific TB tests and diagnostic markers, all of which have a low sensitivity and specificity (van Well *et al.* 2009), leading to delayed intervention, and a high risk of morbidity and mortality. According to the WHO Global TB Report for 2012 (WHO 2013), there were approximately 325 000 clinically defined new cases of TB in South Africa alone, of which 42 000 (14%) were of the extra-pulmonary forms (EPTB). Central nervous system (CNS) TB accounts for an estimated 1–10% of all EPTB (Cherian & Thomas 2011; Bhigjee *et al.* 2007; Rock *et al.* 2005), with tuberculous meningitis – also known as TB meningitis (TBM) – being not only the most

severe complication of the disease but also the most common form of bacterial meningitis (BM) in South African children (Wolzak *et al.* 2012). Infants and young children are particularly prone to dissemination of TB, especially TBM, due to their young age and immature immune systems. Most TBM patients upon admission present with an advanced stage of disease due to difficulty of diagnosis in the early stages, resulting in an estimated 30% of patients dying despite treatment (Youssef *et al.* 2006).

Cerebrospinal fluid (CSF) is the biofluid that is in direct contact with the site of TBM infection and contains a wealth of information potentially useful for TBM diagnosis and treatment. Evaluation of the characteristics of CSF is already a key factor in the diagnosis of several CNS diseases (Seehusen *et al.* 2003; Watson & Scott 1995), particularly of TBM (van Well *et al.* 2009). CSF from TBM cases typically shows a clear appearance, moderate pleocytosis with a predominance of lymphocytes, increased protein content and a very low glucose concentration (Principi & Esposito 2012). To improve on these general clinical observations, more sophisticated clinical-chemistry tests on CSF have been proposed but none with elevated sensitivity and specificity (Ho *et al.* 2013), although potentially useful to support the diagnosis in some cases (Principi & Esposito 2012). Improved clinical outcomes are highly dependent on timely diagnosis and initiation of appropriate treatment for survival benefit (Ruslami *et al.* 2013), leaving monitoring of treatment likewise a key developmental need for this complex infectious disease, to which an untargeted metabolomics investigation could make a distinct contribution, as shown in this paper.

Metabolomics information has the advantage of being context dependent by disclosing the complement of metabolites associated with a physiological, developmental, or pathological state prevailing in a cell, tissue, organ, or organism. Infection of the Mtb bacilli occurs through inhalation of aerosolized droplets and harbouring of the bacilli in the pulmonary system. Should dissemination into the lympho-hematogenous circulatory system occur (Krishnan *et al.* 2010; El-Kebir *et al.* 2013), the brain becomes a very attractive site for the establishment of metastatic foci due its rich vascular and oxygen supply. Mtb appears to enter the subarachnoid space via rupture of an adjacent parenchymal tubercle (Leonard & Des Prez 1990) and CD14 receptors of the microglia may render them as the principal cell target in the CNS (Peterson *et al.* 1995). Microglial cells are the resident macrophages of the brain parenchyma; they share many, if not all, of the properties of macrophages in other tissues (Nareika *et al.* 2005), such as production of a variety of cytokines and chemokines on activation. Despite this hostile bactericidal environment, it provides protection and shelter to Mtb from other immune cells and promotes development of small tuberculous foci,

commonly known as granulomas (Ernst 1998; Davis & Ramakrishnan 2009), presenting one of the ultimate paradoxes in the study of TBM host–pathogen interactions (Thi *et al.* 2012).

A unique aspect of TB in children is the unnoticed infection by Mtb in the often rapid development of the disease (Swaminathan & Rekha 2010), usually in children younger than 5 years and/or in HIV-infected children (Starke 2003). The outcome of TBM correlates with the stage at which disease treatment starts, underlining the need for early detection and a prompt clinical response. Metabolomics proved to be suited for a powerful global study of diseases of the human CNS, with the potential of mapping early biochemical changes in these diseases, providing opportunities for timely diagnosis and development of signatures and hence for assisting in effective interventions (Dunne *et al.* 2005; Kaddurah-Daouk & Krishnan 2009; Madsen *et al.* 2010; Sinclair *et al.* 2009). The systematic characterization of the CSF metabolome is well established with contributions coming from studies using various platforms (Guo *et al.* 2011; Mandal *et al.* 2012; Stoop *et al.* 2010; Wevers *et al.* 1995; Wishart *et al.* 2008). To date, the presence of 476 detectable metabolites with an extensive dynamic range has been confirmed in the human CSF metabolome, of which 170 are readily measurable, 75 have concentrations above 1 µM and 47 metabolites have been quantified by NMR by Wishart *et al.* (2008). Due to its hydrophilic nature, NMR can be considered the best platform for characterizing CSF. Major challenges for investigations of the CSF metabolome are (1) the diverse nature and low concentrations of metabolites, (2) limited availability and (3) small sample volumes (Guo *et al.* 2011; Stoop *et al.* 2010), as a result of ethical issues, such as pain and health risks, that restrict the general availability of CSF for research purposes (Geiszler *et al.* 2013). To date, NMR-based metabolomics of TB has been conducted on infected mice (Shin *et al.* 2011), guinea pigs (Somashekar *et al.* 2011) and humans (Zhou *et al.* 2013), on other forms of meningitis (Coen *et al.* 2005; Himmelreich *et al.* 2009) and a study by Subramanian *et al.* (2005), which selected 12 CSF metabolites for differential diagnosis of meningitis in children through in-house software diagnostics.

The focus of the NMR metabolomics study of CSF reported here is directed towards the host's metabolic response to the TBM infection. The unusually high incidence of TB in childhood in isolated regions in the Western Cape province of South Africa provided a unique opportunity to obtain CSF samples for a metabolomics study based on a cohort of infants and small children definitively diagnosed with TBM. Two control groups were likewise used in this investigation: the first consisted of South African infants and children initially suspected to be suffering from meningitis (but subsequently shown to be negative), and secondly, control samples from an unrelated neurological study in the Netherlands. The

outcome of this NMR investigation revealed metabolic perturbations due to TBM, and has considerable potential for improving the early clinical assessment of the disease state in TBM and for monitoring responses to treatment regimes. The metabolic profile also contributed to hypothesis formulation, proposed as an “astrocyte–microglia lactate shuttle” (AMLS), encapsulating the unique metabolic plasticity in the CNS metabolism towards the energy requirements for the microglia-driven neuroinflammatory responses, of which TBM may be one such example.

4.3 Material and methods

4.3.1 Experimental design

Metabolic fingerprinting of NMR data requires pre-processing to generate a data matrix of variables and cases of an operational size, to be followed by multivariate analysis to identify the relevant analytical information. The work flow that we followed is shown schematically in Figure 4-1. Our selected experimental group consisted of 33 children with confirmed TBM and two control groups, as described below. Untargeted ^{1}H NMR analysis of CSF samples yielded spectra that were resolved into 110 buckets of variable size. Following removal of outliers from the patient and two control groups, the important, known metabolites observed in the TBM-infected cases were identified and cross-validated by statistical modelling. The 16 NMR-derived metabolites differentiating TBM from non-TBM, based on this South African cohort of patients, were then quantified, compared to normal reference ranges and the corresponding biological interpretation given, from which our hypothesis was generated.

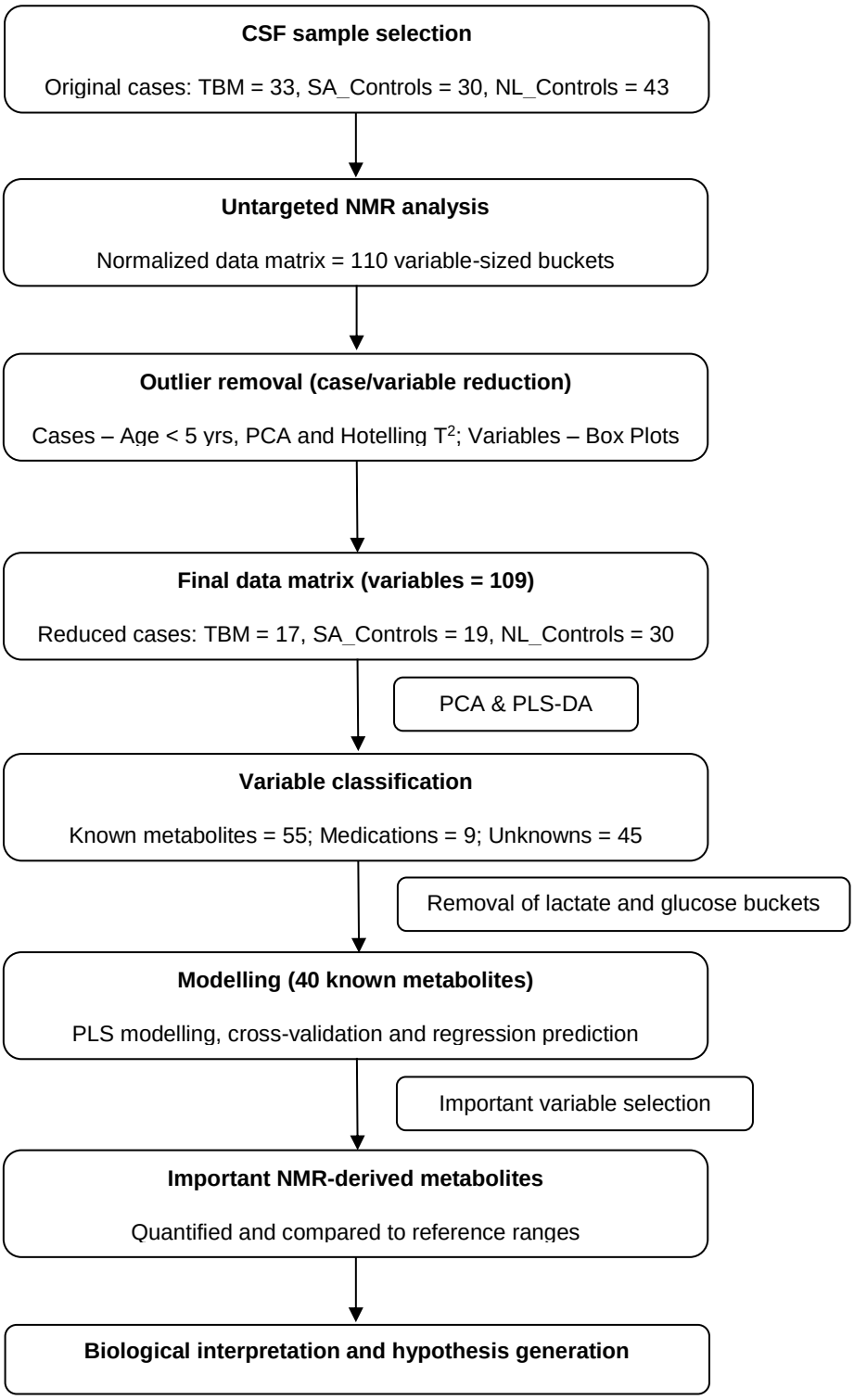


Figure 4-1: **Schematic representation of the work flow following data generation, based on statistical pre-processing, data analysis and cross-validation to identify important NMR-derived metabolites of TBM infection.**

4.3.2 Sample collection, description and storage

The three experimental groups used in this study were: 1) South African patients with confirmed TB meningitis (see comprehensive description of clinical profiles in Table A1-1 and Table A1-2 of the supplementary information (SI) — Annexure 1, and on the medication used for treatment in Table A1-3); 2) non-meningitis South African controls (SA_Controls); and 3) neurological controls from the Netherlands (NL_Controls). The first two groups comprised children between the ages of 6 months and 12 years, all of whom were originally suspected meningitis cases and referred from local clinics to the paediatrics unit at Tygerberg Hospital in the Western Cape province of South Africa. In most of these cases, a broad range of non-specific treatments (e.g., broad-spectrum antibiotics, analgesics and anti-inflammatories) were initiated prior to admission. These individuals were given a thorough assessment by a paediatric neurologist at Tygerberg Hospital, which involved an extensive description of the clinical background, including polymerase chain reaction (PCR) and culture analysis of a CSF sample obtained through a lumbar puncture. The main physical presentations and clinical symptoms of these cases, as testified by parents and observed by the respective clinicians, were compatible with non-CNS indications of TB outside the CNS, typically observed in small TB-infected children from the region. As part of the diagnostic process a detailed inspection of the clinical chemistry of the CSF was conducted and yielded a report describing the count and type of cells present, particularly immune response-related cells (Table A1-2), as well as a measure of CSF protein (mostly high: >1 g/L) and glucose (mostly low: <2.2 mmol/L or CSF:blood glucose ratio <50%) levels. The South African controls were confirmed negative for any form of meningitis, despite being ill with symptoms reminiscent of the condition. Ethical and practical considerations limited the availability of healthy controls, which could partially be overcome through a comparative analysis of a second control group with CSF collected from untreated individuals from the Radboud University Medical Centre in Nijmegen, the Netherlands. The Netherlands non-TBM control group (the third experimental group) consists of CSF samples from age-matched patients who were suspected to suffer from a neurometabolic disease. After appropriate and in depth investigations no clinical or biochemical evidence was found for such diagnosis in any of these patients. The only disease exclusion criterion applied to all cases was HIV co-infection.

The present study was approved by the Human Research Ethics Committee of Stellenbosch University, South Africa (study nr. N11/01/006).

4.3.3 Sample preparation and ^{1}H NMR spectroscopy

All CSF samples were stored at –80°C prior to analysis and transported to the NMR facility at the Laboratory for Genetic, Endocrine and Metabolic Diseases at the Radboud University Medical Centre in Nijmegen, where the NMR analyses were conducted. Sample preparation followed the standard operating procedure (SOP) as established by the Nijmegen laboratory (Engelke *et al.* 2005; Engelke *et al.* 2006; Wevers *et al.* 1995).

Each CSF sample was measured at 500 MHz on a Bruker DRX Avance spectrometer equipped with a triple-resonance inverse (TXI) probe head. Software used was: Bruker Topspin (V3.1) for data pre-processing and Bruker AMIX (V3.9.12) for binning and quantification (Ellinger *et al.* 2013). Variable-sized bucketing was used to create a data matrix of 110 variables, followed by log transformation. The NMR region representing the suppressed water signal (4.67–4.96 ppm) was excluded from the data matrix. A more detailed description of the NMR protocol and pre-treatment of the original NMR-data are included in section A1.2 and A1.3, respectively, of the SI.

4.3.4 Statistical analyses

The original TBM group was slightly reduced by excluding 4 cases which were older than 5 years of age, intended to define a more homogeneous group of infants and children. Unsupervised principal component analysis (PCA) and Hotelling's T^2 test, with a confidence level of 95%, was used to remove further case outliers for each experimental group; assessment of overall variance of variables was used to remove variable outliers. The multivariate analyses (see Figure 4-1) were done using the statistical software package The Unscrambler® X (V10.3, CAMO software AS, Norway), successfully applied in other metabolomics studies (e.g., Maddula and Baumbach 2011). Supervised partial least squares–discriminant analysis (PLS-DA) modelling was used to identify the important variables and cross-validated by leaving one sample out at a time, using the same samples for model estimation and testing. The principal advantage of this leave-one-out (LOO) cross-validation is that it is an efficient method of dealing with a small number of cases (as is typically the situation with metabolomics data) and allows for the jack-knifing approach on which an uncertainty test is based. The uncertainty test creates a number of sub-models on all the samples not kept in the cross-validation segment; for every sub-model the variations are estimated to assess the stability of the results and an overall p-value is output per validated variable based on the PLS model. Following this, a regression model is given that

plots the predicted Y-values of the PLS model for all samples, together with a deviation that expresses the uncertainty of the prediction. The deviations are estimated as a function of the global model error, the sample leverage, and each sample's residual X-variance. A small deviation indicates that the sample used for prediction is similar to the samples used for the calibration model, whereas predicted Y-values with high deviations are less reliable. This predicted regression model is used to assess how well the PLS model is able to correctly classify cases as TBM or non-TBM.

The online metabolomics suite, Metaboanalyst 2.0 (www.metaboanalyst.ca) (Xia *et al.* 2009), was used to create the quantitative VIP values [Variables Important in Projection (VIP) are conventionally used in metabolomics studies to identify important metabolites] for the PLS-DA analyses, as well as to calculate the univariate measures [fold change (d-values) and t-test (p-values)] for the final reduced data matrix.

4.4 Results and discussion

4.4.1 Data analysis and identification of important variables

A representative NMR spectrum, covering the region 0.80–4.80 ppm and scaled relative to the internal standard peak (TSP), from each of the three experimental groups used, is given in Figure 4-2. This section of the spectrum serves to illustrate some of the discernible qualitative NMR differences associated with the CSF taken from the TBM-infected infants and small children compared to non-TBM cases. Not shown in Figure 4-2 are assignments attributed to medication found within the aromatic region of the spectra (6–9 ppm) from treated individuals, which were identified by comparison with spectra corresponding to the pure compounds, as well as unassigned peaks that have yet to be identified.

Case and variable reduction was followed as described in Figure 4-1. Unsupervised PCA of all 3 experimental groups separated the TBM cases and both control groups, with some overlap between the control groups indicating that they shared similar characteristics but were not homogeneous (see Figure A1-1). PCA of the SA_Controls vs TBM (Figure 4-3A) and NL_Controls vs TBM (Figure 4-3B) cases further illustrates the natural separation between TBM and non-TBM. Quantitatively, the two metabolites that yielded the greatest power values based upon PCA (i.e., those most responsible for the separation) were highly elevated lactate and decreased glucose in the TBM cases relative to normal reference values for CSF glucose (see Table A1-5).

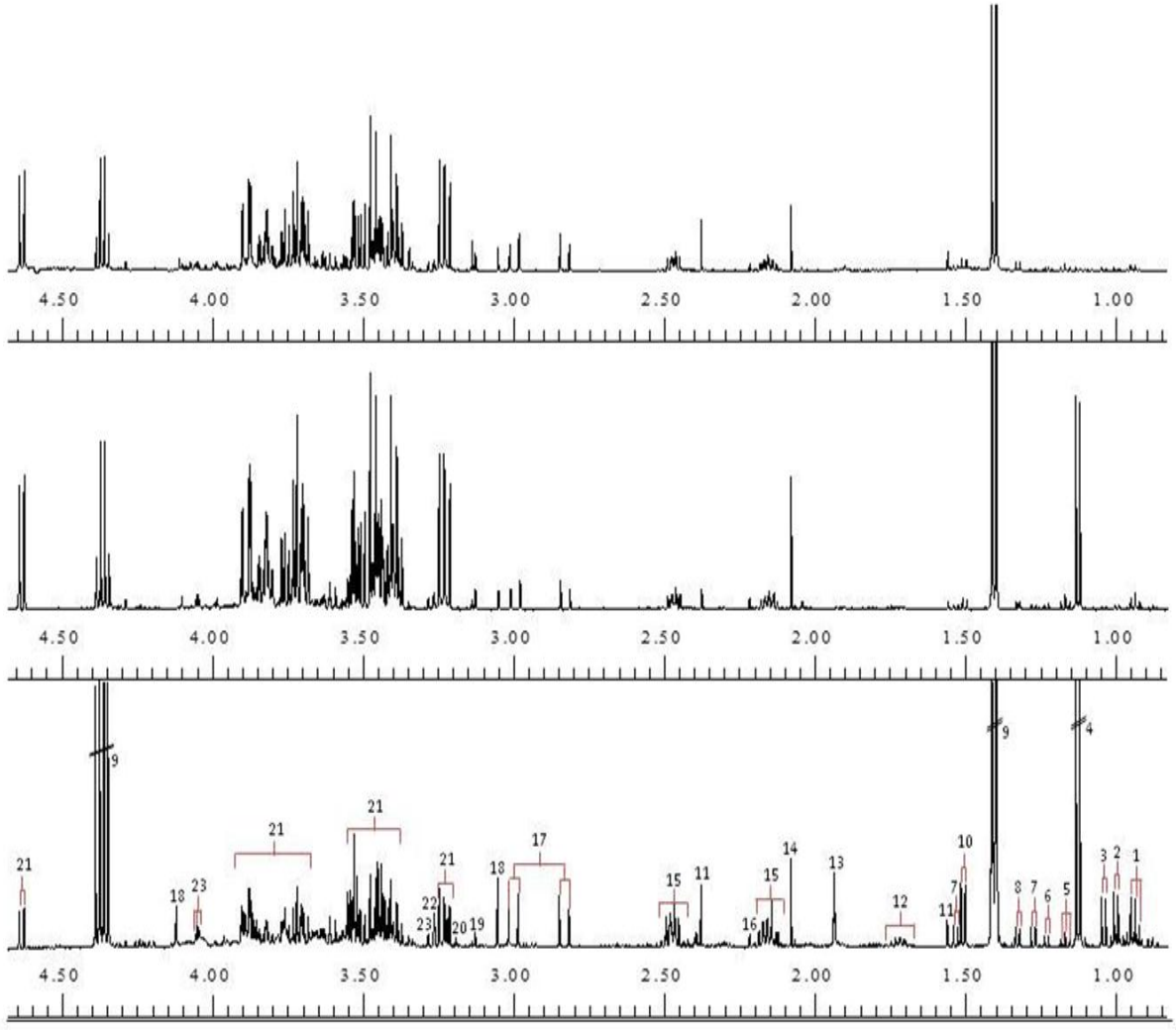


Figure 4-2: **^{1}H 500 MHz NMR spectra of CSF at pH 2.5 scaled according to internal standard peak (TSP) illustrating qualitative differences between single examples of untreated NL_Control (top), treated SA_Control (middle) and TBM cases (bottom).**

[Assignments (ppm): 1 = isoleucine/leucine (0.94t/0.95d), 2 = valine/isoleucine (1.00d/1.01d), 3 = valine (1.04d), 4 = propylene glycol (1.13d), 5 = ethanol (1.17d, 3.64q), 6 = 3-hydroxybutyrate (1.23d, 2.53q), 7 = ^{13}C lactate (1.30d, 1.52d), 8 = threonine/3-hydroxyisovalerate (1.33d/1.33s), 9 = lactate(1.41d, 4.36q), 10 = alanine (1.51d), 11 = pyruvate (1.56s, 2.37s), 12 = lysine (1.73m), 13 = unknown, 14 = acetate (2.08s), 15 = glutamine (2.16m, 2.47m), 16 = acetone (2.22s), 17 = citrate (2.91AB), 18 = creatine (3.05s, 4.10s), 19 = creatinine (3.13s, 4.29s), 20 = choline (3.19s) , 21= glucose (3.23dd, 3.3-3.9, 4.64d, 5.22d), 22 = betaine/myo-inositol (3.27s/3.27t), 23 = myo-inositol (4.05t)].[7]

[7] Multiplicity of peak(s): s = singlet, d = doublet, dd = double duplet, t = triplet, q = quartet, m = multiplet, AB = AB system.

In order to focus on the biological/metabolic information we subdivided the data matrix into three classes of variables, namely: 1) known endogenous compounds, 2) medications (exogenous), and 3) unassigned variables. To determine the more subtle characteristics that distinguish the host metabolic profile of TBM we selected the known endogenous variables (classifiable as metabolites), which could be confidently assigned given the highly specific nature of NMR technology, and in addition removed the two dominating metabolites – lactate and glucose. This filtering resulted in a reduction of the data matrix under analysis from 109 to 40 variables.

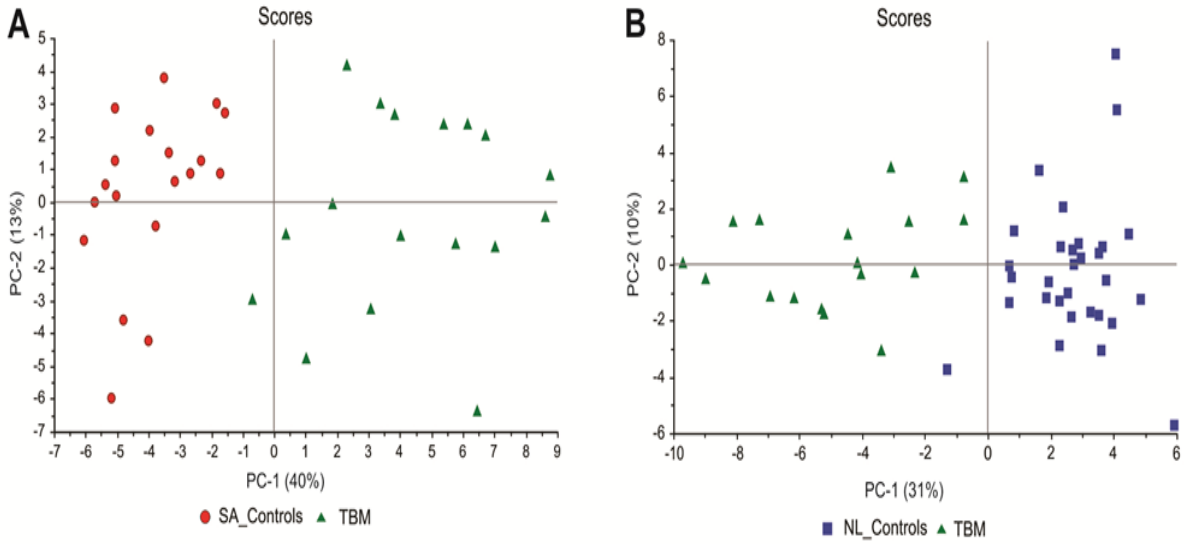


Figure 4-3: PCA scores plot of SA_Controls vs TBM (A) and NL_Controls vs TBM (B), showing natural separation between TBM and non-TBM cases.

Supervised PLS-DA modelling was performed on the reduced data set for both the SA_Controls vs TBM and NL_Controls vs TBM cases as shown in Figure 4-4. The regression model identified the important variables, marked in the corresponding correlation loadings plots, which were then cross-validated and used in a regression prediction with deviation that illustrates – for the SA_Controls vs TBM case (Figure 4-4B) and in the NL_Controls vs TBM case (Figure 4-4D) — that there is good case classification. The R^2 and Q^2 values for the NL_Controls vs TBM case were 0.80 and 0.77 respectively, while the R^2 and Q^2 values for the SA_Controls vs TBM case were 0.90 and 0.83, respectively. The fundamental requirement for PLS to yield meaningful information is the ability to select variables. In our case this is done quantitatively on the basis of the VIP value, which, together with other meaningful univariate quantitative measures (fold change d-values and t-test p-values), are given in Table 4-1 Based upon a selection criterion of VIP > 1.0, and

taking the univariate measures into consideration, we were able to select 8 significant metabolites common to both cases, namely: myo-inositol, creatinine, dimethyl sulfone and several amino acids (lysine, alanine and three branched-chain amino acids: valine, isoleucine and leucine). Using the same selection criteria, phenylalanine and tyrosine are uniquely important when comparing the NL_Controls vs TBM case, whereas choline, formate, acetate, citrate and pyruvate are uniquely important to the SA_Controls vs TBM case. Ethanol manifested as important in the NL_Controls vs TBM case but, although very small amounts of ethanol are produced endogenously by bacteria through anaerobic fermentation, it is likely that its presence was due to contamination from skin disinfection prior to lumbar puncture.

Thus a total of 16 NMR-derived CSF metabolites, which differentiated between the South African TBM cases under investigation and controls, were selected for quantification in the 17 TBM, 19 SA_Control and 30 NL_Control cases and the outcome compared to their normal reference ranges for human CSF (summarized output presented in Table 4-2).

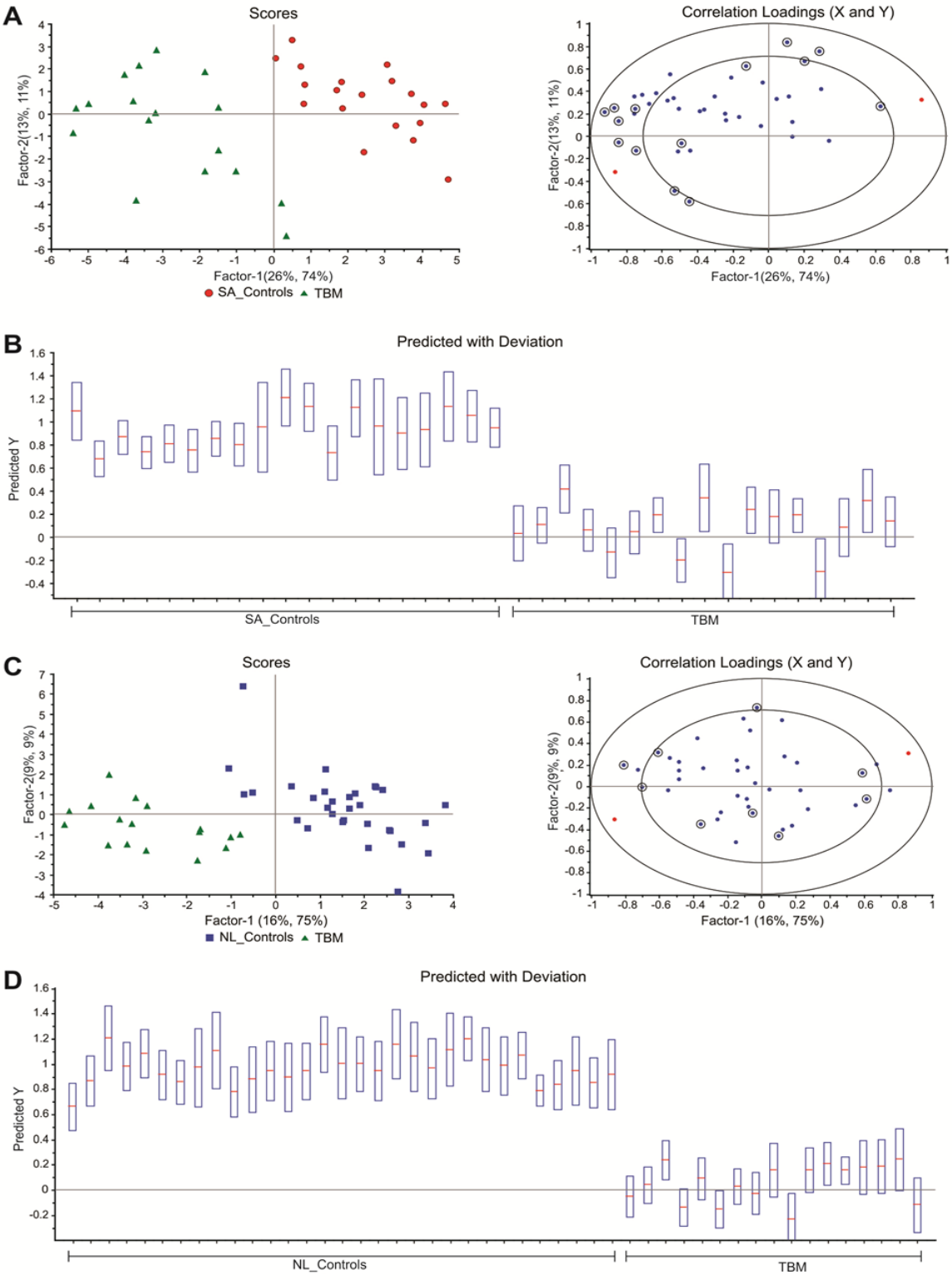


Figure 4-4: **PLS scores plot and corresponding correlation loadings with marked important variables of SA_Control vs TBM (A) and NL_Controls vs TBM (C) reduced cases. The regression prediction plot illustrates that the PLS model classifies the cases correctly in the SA_Control vs TBM (B) and NL_Control vs TBM (D) cases.**

Table 4-1: **Quantitative statistical data indicating the important metabolites that discriminate between TBM and non-TBM for both SA_Controls vs TBM and NL_Controls vs TBM cases with dominating metabolites lactate and glucose removed [The chemical shift, in ppm, of each identified metabolite is given in brackets].**

SA_Controls vs TBM	PLS	t-test	Fold change	Validation
	(VIP value)	(p-value)	(d-value)	(p-value)
choline (3.19)	1.810	<0.001	0.263	<0.001
alanine (1.51)	1.793	<0.001	0.344	<0.001
myo-inositol (3.285)	1.708	<0.001	1.605	0.007
creatinine (3.13)	1.654	<0.001	1.689	<0.001
lysine (1.73)	1.611	<0.001	0.601	<0.001
valine/isoleucine (1.01)	1.580	<0.001	0.409	<0.001
valine (1.04)	1.514	<0.001	0.417	<0.001
acetate (2.08)	1.371	<0.001	1.635	0.224
myo-inositol (4.05)	1.338	0.001	1.478	0.366
$DMSO_2$ (3.14)	1.110	0.005	1.809	0.072
formate (8.25)	1.089	0.006	0.651	<0.001
pyruvate (1.56)	1.083	0.007	0.632	0.001
citrate (2.81)	1.044	0.009	1.245	0.616
isoleucine/leucine (0.95)	1.042	0.009	0.715	<0.001
pyruvate (2.37)	1.011	0.012	0.668	0.005
tyrosine (7.19)	0.979	0.015	0.688	<0.001
2-oxoglutarate (2.68)	0.977	0.015	0.511	0.007
phenylalanine/medication (7.42)	0.947	0.019	0.666	<0.001
citrate (2.84)	0.944	0.019	1.224	0.838
threonine (1.34)	0.895	0.027	0.727	<0.001
creatinine (4.29)	0.891	0.028	1.176	0.185
creatine (3.05)	0.886	0.029	0.832	<0.001
phenylalanine (7.39)	0.790	0.053	0.681	0.010
creatine (4.10)	0.751	0.067	1.230	0.500
glutamine (2.47)	0.747	0.068	1.098	0.233
succinate (2.66)	0.699	0.089	0.671	0.016
acetone (2.22)	0.576	0.164	1.763	0.617
mannose (5.17)	0.530	0.202	1.626	0.422
ethanol (3.64)	0.516	0.214	1.145	0.877
tyrosine/medication (6.90)	0.512	0.218	0.873	0.125
acetoacetate (2.30)	0.421	0.312	0.951	0.178
citrate (2.97)	0.367	0.380	1.062	0.327
glutamine/medication (2.16)	0.287	0.492	1.049	0.090
ethanol (1.18)	0.237	0.572	1.013	0.491
betaine/myo-inositol (3.27)	0.202	0.630	0.949	0.002
3-hydroxybutyrate (2.53)	0.198	0.636	0.999	0.374
3-hydroxyisovalerate/threonine (1.33)	0.113	0.789	0.949	0.003
citrate (3.00)	0.083	0.843	0.995	0.002
carnitine (3.22)	0.031	0.940	0.952	0.026
3-hydroxybutyrate (1.23)	0.018	0.965	1.123	0.882

NL_Controls vs TBM	PLS	t-test	Fold change	Validation
	(VIP value)	(p-value)	(d-value)	(p-value)
alanine (1.51)	1.986	<0.001	0.299	<0.001
valine (1.04)	1.961	<0.001	0.320	<0.001
creatinine (3.13)	1.938	<0.001	1.860	<0.001
$DMSO_2$ (3.14)	1.851	<0.001	3.132	<0.001
myo-inositol (3.285)	1.836	<0.001	1.778	<0.001
creatinine (4.29)	1.557	<0.001	1.705	<0.001
myo-inositol (4.05)	1.497	<0.001	1.734	0.003
valine/isoleucine (1.01)	1.477	<0.001	0.449	<0.001
lysine (1.73)	1.441	<0.001	0.592	<0.001
ethanol (1.18)	1.329	0.001	0.314	0.067
tyrosine/medication (6.90)	1.274	0.002	0.620	0.002
phenylalanine/medication (7.42)	1.092	0.009	0.698	<0.001
tyrosine (7.19)	1.085	0.009	0.726	<0.001
phenylalanine (7.39)	1.074	0.010	0.578	0.003
acetoacetate (2.30)	0.825	0.051	1.130	0.164
choline (3.19)	0.801	0.058	0.770	0.125
isoleucine/leucine (0.95)	0.788	0.063	0.804	0.005
citrate (2.84)	0.730	0.085	1.284	0.121
creatine (4.10)	0.693	0.103	1.222	0.269
pyruvate (1.56)	0.649	0.128	0.858	0.026
pyruvate (2.37)	0.642	0.132	0.845	0.042
glutamine (2.47)	0.638	0.134	1.131	0.428
mannose (5.17)	0.569	0.183	4.641	0.075
creatine (3.05)	0.540	0.207	0.907	0.295
3-hydroxybutyrate (2.53)	0.431	0.315	1.019	0.794
acetate (2.08)	0.382	0.374	1.135	0.501
betaine/myo-inositol (3.27)	0.348	0.418	1.104	0.353
3-hydroxyisovalerate/threonine (1.33)	0.330	0.442	1.179	0.715
ethanol (3.64)	0.306	0.477	1.074	0.513
glutamine/medication (2.16)	0.221	0.607	0.977	0.419
formate (8.25)	0.209	0.627	0.967	0.440
carnitine (3.22)	0.206	0.633	1.043	0.857
citrate (3.00)	0.203	0.637	1.032	0.362
citrate (2.97)	0.197	0.647	1.025	0.510
threonine (1.34)	0.183	0.671	1.080	0.672
acetone (2.22)	0.172	0.689	1.681	0.752
citrate (2.81)	0.164	0.703	1.064	0.685
3-hydroxybutyrate (1.23)	0.149	0.730	1.364	0.722
succinate (2.66)	0.133	0.758	1.103	0.640
2-oxoglutarate (2.68)	0.111	0.797	1.060	0.456

4.4.2 Brain energy metabolism and the response to TBM

The clear and defining metabolites distinguishing the TBM cases under investigation from the control groups were the combination of perturbed glucose and highly elevated lactate, which is in line with the clinical characteristics of CSF from patients with BM and TBM (Leib *et al.* 1999) and is generally accepted as a key indication of disturbances in energy metabolism of many neurological disorders (Leen *et al.* 2012). Regional variation and the complement of cellular determinants participating in a synchronised way is an unique characteristic of neurological energy production. It has thus been proposed that the microvascular endothelium of the blood–brain barrier (BBB), astrocytes, neurons and extracellular matrix constitute a "neurovascular unit", which is essential to normal CNS physiology and health (reviewed by Hawkins & Davis 2005).

A basic functional component in this unit is the metabolic astrocyte–neuron coupling. The astrocytes participate through glutamate-stimulated aerobic glycolysis, resulting in the release of lactate, thereby serving as fuelling in conditions of an activity-dependent neuronal energy demand (Magistretti & Pellerin 1999b). Several experiment-based insights (reviewed by Pellerin & Magistretti 2003 and Pellerin & Magistretti 2004) lend empirical support to the original proposed "astrocyte–neuron lactate shuttle" model on energy transfer in the CNS (Bittar *et al.* 1996; Pellerin *et al.* 1998). Recent views highlight the importance of brain endothelial cells in the modular organization of the neurovascular unit (Abbott *et al.* 2006) and summarize how several pathologies of the CNS may be involved in the disturbance of this organization, including pathological conditions like TBM. The resulting profile of energy-associated metabolites in CSF from TBM patients (Table 4-1) differs significantly from profiles of the metabolites associated with the controls in several ways.

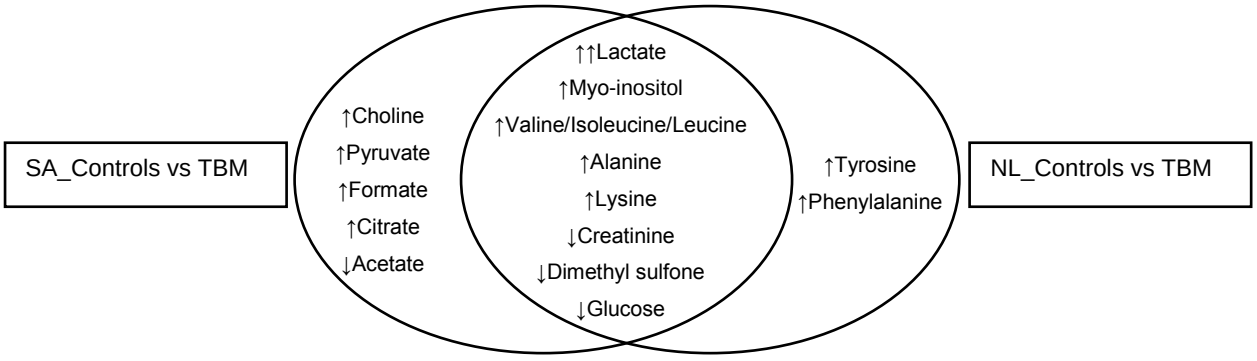


Figure 4-5: Important metabolites that distinguish between TBM and non-TBM cases based on VIP values >1.0 and significant univariate measures, common to both cases and unique to each case (increase↑/decrease↓ relevant to TBM cases).

Table 4-2: Summarized quantified data of 16 important metabolites discriminating between TBM and controls, compared with normal reference ranges.

Metabolite		Isoleucine/ Leucine	Valine	Alanine	Lysine	Phenylalanine	Tyrosine	Creatinine	Acetate	Pyruvate	Citrate	$DMSO_2$	Choline	Myo-inositol	Lactate	Glucose[c]
Chemical form (shift)		$(CH_3)_2$ (0.95t)	CH_3 (1.00d)	CH_3 (1.51d)	CH_2 (1.73m)	$(CH)_3$ (7.35m)	$(CH)_2$ (7.19d)	CH_3 (3.13s)	CH_3 (2.08s)	CH_3 (2.36s)	CH_2 (2.89AB)	$(CH_3)_2$ (3.14s)	$(CH_3)_3$ (3.19s)	$(CH)_2$ (4.05t)	CH_3 (1.41d)	α-CH (5.22d)
NL_Controls	Mean	18.18	12.69	23.20	24.75	4.74	5.62	52.34	120.15	45.70	135.30	10.70	2.11	201.36	1394.81	2838.15
	STD	9.31	5.75	12.43	13.65	5.02	5.15	24.02	78.60	30.13	48.00	6.44	1.34	203.23	619.61	933.18
NL_Controls vs TBM t-test (p-value)		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.506	0.028	<0.001	<0.001	0.003	0.002	0.306	<0.001	0.640
SA_Controls	Mean	26.44	16.06	37.28	29.52	6.11	13.60	58.96	192.98	36.00	172.90	7.75	2.10	169.85	1706.83	4561.79
	STD	11.01	9.63	13.52	13.44	9.18	19.08	16.42	50.30	26.53	45.16	4.72	1.42	47.52	419.06	1101.83
SA_Controls vs TBM t-test (p-value)		<0.001	<0.001	0.028	<0.001	<0.001	0.005	0.075	0.736	<0.001	0.015	0.094	0.013	0.400	<0.001	<0.001
TBM	Mean	54.01	69.49	99.47	86.60	37.44	31.04	47.70	183.38	82.03	206.02	4.86	4.34	177.30	7363.69	2689.05
	STD	27.55	40.05	71.92	40.72	18.44	14.81	20.29	111.22	35.60	29.47	5.35	3.39	95.59	2361.32	1216.99
Reference range		7±5/ 16±9[a]	19 ±13[a]	46 ±27[a]	29 ±13[a]	15 ±13[a]	12 ±9[a]	43 ±12[a]	58 ±27[a]	53 ±42[a]	225 ±96[a]	2 ±1[a]	3 ±1[a]	84 ±40[a]	1651 ±626[b]	2960 ±111[b]

*Concentration units are µmol/L; [a]Wishart *et al.* 2008; [b]Leen *et al.* 2012; [c]total glucose concentration calculated based on α-glucose concentration and using α:β ratio of 39:61 (Boss *et al.* 2000).

4.4.2.1 Metabolic burst

The increased catabolism of glucose through glycolysis to pyruvate, and its conversion to alanine and predominantly lactate, reflects a metabolic burst as seen in the TBM CSF profile, with the greatly elevated levels of lactate, most likely leading to immune activation in the TB-infected microglial cells (Nareika *et al.* 2005). In addition, the increased levels of many monocarboxylates such as lactate and pyruvate, the presence of branched-chain keto acids derived from leucine, valine and isoleucine, and ketone bodies such as acetoacetate, 3-hydroxybutyrate and acetate suggests functional down regulation of the neuron-associated monocarboxylate transporter-1 (MCT-1), resulting in decreased neuronal energy supply, which manifests in the coma-related clinical symptoms seen in the TBM patients. Down regulation of neuronal MCT-1 likewise promotes increased levels of these monocarboxylic acids and ketones, which could be destined for high-energy support to activated microglia operative in TBM. Some pyruvate is converted to alanine by a transaminase reaction. The ammonia necessary for this reaction depends on the presence of branched amino acids, leading to an increase in the branched-chain 2-keto acids. It should be noted that acetyl-CoA is an important precursor to the tricarboxylic acid cycle (TCA) for energy production, not only in the host, but is also one of the primary sources of energy for Mtb (Savvi *et al.* 2008). In Mtb, formate is a by-product of acetate production and may inhibit downstream mitochondrial energy production in the host, limiting macrophage activation.

4.4.2.2 Amino acids

The branched-chain amino acids (BCAAs – isoleucine, leucine and valine) participate directly and indirectly in a variety of important biochemical functions in the brain, including the provision of building blocks for the production of energy through gluconeogenesis under conditions of high-ATP demands as prevails in activated microglia. BCAAs and alanine also participate in intercellular shuttles between mitochondria in astroglia and neurons (Sweatt *et al.* 2003). Likewise, a basic aspect of neurometabolic coupling is the glutamate/glutamine link in the glycolytic catabolism in astrocytes, lactate production and its availability to mitochondrial ATP production in neurons, or proposed for microglia in TBM. Increased lysine seen in CSF from TBM patients has been associated with mental retardation and in other motor neuron diseases (Shaw *et al.* 1995); it has also been shown to form adducts with other compounds, such as acrolein–lysine, a marker of lipid peroxidation in childhood meningitis (Tsukahara *et al.* 2002).

4.4.2.3 Creatinine

A catabolic product of creatine phosphate in muscle that is produced at a fairly constant rate during homeostasis and has a slow diffusion rate across the blood–brain barrier. It is also correlated with the CSF monoamine metabolites homovanillic acid and 5-hydroxyindoleacetic acid, which are implicated in neurodegenerative conditions such as depression and schizophrenia (Agren & Niklasson 1988; Levine *et al.* 2000; Swahn & Sedvall 1988). Implications of perturbed creatinine in TBM is unclear but are likely linked proportionally to the progression of neuronal injury.

4.4.2.4 Myo-inositol

This carbohydrate is synthesized mainly *de novo* from glucose in the brain. As phosphoinositides, they participate as second messengers in numerous neurotransmitter systems and are important in signal transduction (Cordoba *et al.* 1996). Myo-inositol is a glial cell marker that has been implicated previously in activation of microglia and astrocytes (gliosis), as well as a known pathological response readily observed in neurodegenerative disease and neuroinflammation (Pears *et al.* 2005). Thus myo-inositol is known to be crucial in the initiation of the cascade of events necessary for an immune response, although the mean concentrations found in the present study do not strongly highlight this function of myo-inositol.

4.4.2.5 Choline

Choline is synthesized in small amounts in the brain by converting phosphatidylethanolamine to phosphatidylcholine – although its main source is dietary – and can be oxidized to form betaine, which is a methyl source for many reactions (e.g., conversion of homocysteine to methionine). It is an important precursor of the neurotransmitter acetylcholine, which activates muscles in the peripheral nervous system and, in excess, induces seizures (Zimmerman *et al.* 2008), a common clinical symptom in TBM. Increased choline can be related to neural membrane breakdown, reflecting neuronal loss and gliosis in the brain.

4.4.2.6 Dimethyl sulfone ($DMSO_2$)

$DMSO_2$ is a normal constituent of human CSF, occurring in concentrations of 0–25 µmol/L and is derived from dietary sources, intestinal bacterial metabolism and human endogenous methanethiol metabolism. Metabolically, $DMSO_2$ arises from oxidation of dimethyl sulfoxide (DMSO), and partly from methionine metabolism (Engelke *et al.* 2005). DMSO is a polar

amphiphilic compound that increases membrane fluidity (Hallows & Frank 1992), acts as a powerful scavenger of oxygen radicals (Rosenblum 1983), stimulates granulocytic differentiation (Watson *et al.* 1997) and induces apoptosis of neuroblastoma cells (Kruman *et al.* 1993) and macrophages (Marthyn *et al.* 1998). Decreased levels of $DMSO_2$ in CSF can thus be postulated as being a consequence of perturbed osmoregulation of CSF and depletion of DMSO in response to oxidative stress, induced apoptosis and differentiation of macrophages.

4.4.3 A hypothetical "astrocyte–microglia lactate shuttle" (AMLS)

The pathophysiological response to neural infection by Mtb provided the context for our interpretation of the resultant metabolic changes outlined above. Clinical observations in our TBM patient group include the classical profile of encephalopathy with manifestations of seizures, involuntary movements, fever, impaired consciousness, poor feeding and vomiting, indicative of neuroinflammatory responses and impaired neuronal functioning (Udani & Dastur 1970). Concomitantly, microglial cells are likely to pose an increased energy demand for phagocytosis of Mtb, most likely strengthened by anaerobic glycolysis in the activated microglia, although this response appears to be insufficient as the energy provider for an inflammatory response (Volomoueva *et al.* 2013). A present paradigm is that lactate functions as a key intermediate in conditions of increased energy demand, based on the notion that glycolytic and oxidative pathways can also be operationally linked, as opposed to alternative processes only. Lactate, the product of the anaerobic pathway, thus becomes the substrate for the aerobic pathway – reminiscent of the classical observation by Carl and Gerty Cori (Cori & Cori 1928) on the conversion of muscle glycogen to liver glycogen with lactic acid as an intermediary stage, now commonly known as the Cori-cycle. The intermediary role of lactate led to a postulated lactate shuttle (Brooks 1986), even prior to the discoveries of several classes of transport factors had been implicated in this process in the CNS (See "Effector isoforms" as discussed in section A1.4). Various proposals exist on the involvement of lactate in the astrocyte's role in cell–cell participation under neurostress conditions, enabled also through the syncytial network of neuronal cells within the expansive parenchymal domain of the CNS:

(1) Initial *in vitro* experiments indicated that lactate is an efficient energy substrate for neurons, particularly during periods of intense activity. Lactate does not cross the blood–brain barrier easily, which excludes the possibility that blood-borne lactate can be a primary source for this energy provision, although several investigations indicate that astrocytes

could release large amounts of lactate, leading to the hypothesis of an activity-dependent astrocyte-neuron lactate shuttle (ANLS) for the supply of energy substrates to neurons (Pellerin *et al.* 1998). Although the ANLS hypothesis provides an important existing paradigm on energy provision to neurons, alternative views on activated (Patel *et al.* 2014) as well as resting neurons (reviewed by Dienel & McKenna 2014) also strongly prevails, and provides considerable scope for alternative investigations on hypothesis testing.

(2) Lactate is produced continuously under fully aerobic conditions in mammalian skeletal muscle, especially during exercise when rates of glycogenolysis and glycolysis are elevated, with active skeletal muscles being capable of lactate removal, mainly via oxidation, from which an intracellular lactate shuttle (ILS) has been postulated (Brooks 2002). Investigations with primary cultures from rat cortex and hippocampus, including immunohistochemistry and immunoprecipitation techniques, indicated that monocarboxylate transporter-2 (MCT-2) and lactate dehydrogenase (LDH) are co-expressed in the mitochondria of cultured neurons, which gave support to the concept that in neurons, as in skeletal muscle, a mitochondrial ILS could be operative (Hashimoto *et al.* 2008). It has thus been proposed that lactate seems to play a critical role in neuronal survival, memory and conditions like Alzheimer's disease, although more functional evidence is needed to confirm this hypothesis (Newington *et al.* 2013).

(3) Most recently, a microglia–astrocyte–neuron lactate shuttle (MANLS) has been proposed for microglial activation with lipopolysaccharide (LPS) and interferon-γ (IFN), as well as in response to conditions of excitotoxicity. Both conditions imply the participation of a four-component constellation in which lactate is shuttling to neurons not only from astrocytes but also from microglia (Gimeno-Bayón *et al.* 2014). In this model, glucose, oxidized by astrocytes and microglia, are converted into lactate that is taken up by neurons for its complete oxidation, while glutamine and glutamate synaptic removal by microglia fuels the TCA cycle to maintain mitochondrial activity and ATP generation.

Finally, lactate not only serves as a fuel source and gluconeogenic precursor, but it also acts as a signalling molecule (Philp *et al.* 2005; Hashimoto *et al.* 2007). Hashimoto and co-workers demonstrated that lactate increased production of reactive oxygen species (ROS) and up-regulated 673 genes, many known to be responsive to ROS, in L6 myogenic cells. It thus appears that increased lactate augments inflammatory gene expression.

Against this multi-paradigm background, we speculated that the inflammatory responses and metabolic imbalances created in the CNS following the initial phases of infection by a pathogen, such as Mtb, should be to the advantage of the microglia to fulfil their immune-

protective function. For this purpose we advanced the following hypothesis: 'The host's response to neural infection results in an "astrocyte–microglia lactate shuttle" (AMLS) that operates in neuroinflammatory diseases, such as TBM.' We formulated this hypothesis on the metabolite information on TBM in children and infants (summarized in Table 4-2), with special reference to the significant increase in lactate in all the TBM cases, as well as through inductive reasoning (Goodacre *et al.* 2004) on the characteristics of the cell–cell interactions and factor isoforms, which are well-established to be operative in normal and stress-induced conditions in the CNS (see detailed discussion in section A1-4 of the SI).

Briefly, it is postulated that in TBM, lactate produced through glycolysis in astrocytes participates in the activated immune response and, in association with ketones and gluconeogenic amino acids, is collectively directed from the neurons preferentially into microglia where it enters the mitochondrial TCA cycle, contributing to oxidative phosphorylation and hence producing high levels of ATP and forms of ROS required for Mtb degradation. ROS, and a multitude of factors produced by the microglia to modulate the functions of surrounding immune cells, are toxic to neurons and the unregulated activation of microglia in response to stimulants, such as Mtb, propagate neuronal injury (Block & Hong 2005) and eventual apoptotic cell death for the over-activated microglia as well (Liu *et al.* 2001).

Activation of microglia, as implied by the AMLS hypothesis, does not, however, present a uniform process and involves intricate interactions and feedback loops between the microglia, astrocytes and neurons that hamper attempts to construct basic and linear cascades of cause and effect; TBM involves a complex integration of the responses from the various cell types present within the CNS, with microglia and the astrocytes as main players. The initiating result is the infiltration of Mtb into the subarachnoid space followed by immune activation, but several other factors, related to age, genetic background as well as societal factors from the environment and past experiences, are all expected to modulate the integrated response of this complex neuroinflammatory pathology underlying TBM. We also propose a conceptual model, encapsulating the AMLS hypothesis and providing a visual representative of theoretical constructs (see Figure A1-2), as further exploration of the pathophysiological responses in TBM aimed at fostering curiosity and stimulating new ideas in research on this serious disease.

4.5 Concluding remarks

The results from the metabolomics analysis of the CSF from TBM-infected small children, the metabolite profile of the host response from which it is derived, and the avenues opened by the AMLS hypothesis, provide for further reflection and stimulus, especially for improvement of early clinical assessment of TBM, timely introduction of treatment and for monitoring responses to applicable treatment regimes. Finally, despite its declining incidence in most high-income countries, tuberculosis shows no signs of disappearing in the near future and prevails as a major burden of disease affecting high-risk groups, such as socio-economically deprived individuals and especially infants and children. This investigation indicates that a metabolomics analysis of TBM is feasible and a potentially important complementary tool in combating this devastating disease.

Acknowledgements: Shayne Mason is a recipient of a Desmond Tutu-NRF-VU doctoral fellowship for a joint PhD study between the Vrije Universiteit in Amsterdam, the Netherlands and North-West University (Potchefstroom), South Africa. Research funding for this project is provided by the Technological Innovation Agency (TIA) of the Department of Science and Technology of South Africa.

Conflict of interest: There is no conflict of interest to disclose.

Compliance with ethical requirements: The present study was approved by the Human Research Ethics Committee of Stellenbosch University, South Africa (study nr. N11/01/006).

Supplementary information: see Annexure 1.

CHAPTER 5 TUBERCULOUS MENINGITIS IN INFANTS AND CHILDREN: INSIGHTS FROM NUCLEAR MAGNETIC RESONANCE METABOLOMICS

Shayne Mason[1], Carolus J. Reinecke[1], Regan Solomons[2], A. Marceline van Furth[3]

[1]Centre for Human Metabonomics, Faculty of Natural Sciences, Private Bag X6001, North-West University (Potchefstroom Campus), South Africa,
[2]Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 19063, Tygerberg 7505, South Africa,
[3]Department of Paediatric Infectious Diseases and Immunology, Vrije Universiteit Medical Centre, De Boelelaan 1117, 1081 HV Amsterdam, the Netherlands.

Accepted: *South African Journal of Science,* 2016, 112(3/4), (http://dx.doi.org/10.17159/sajs.2016/2015008).

5.1 Summary

Tuberculous meningitis (TBM) is a prevalent form of central nervous system tuberculosis (CNS-TB) and the most severe common form of bacterial meningitis in infants and children below the age of 13 years, especially in the Western Cape Province of South Africa. Research to identify markers for timely and accurate diagnosis and treatment outcomes remains high on the agenda for TBM, in respect of which the field of metabolomics is as yet largely unexploited. However, the national Department of Science and Technology (DST) recently established several biotechnology platforms at South African institutions, including one for metabolomics hosted at North-West University. We introduce this national platform for nuclear magnetic resonance (NMR) metabolomics by providing an overview of work on TBM. We focus on selected collaborative multidisciplinary approaches to this disease and conclude with the outcomes of an untargeted NMR metabolomics study of cerebrospinal fluid from TBM patients. This study enabled the formulation of a conceptual shuttle representing the unique metabolic plasticity of CNS metabolism towards the energy requirements for the microglia-driven neuroinflammatory responses, of which TBM is one example. From insights generated by this explorative NMR metabolomics investigation, we propose directions for future in-depth research strategies to address this devastating disease. In our view, the timely initiative of the DST, the operational expertise in metabolomics now available and the potential for involving national and international networks in this field of research offers remarkable opportunities for the future of metabolomics in South Africa and for an ever greater understanding of disease mechanisms.

5.2 Introduction

Metabolomics has become an established scientific field. It provides analytical, chemical and physiological insights into naturally occurring, low-molecular-weight organic metabolites within organisms, cells, tissues and biofluids, and helps to promote understanding of metabolite interactions. (Lindon *et al.* 2011). Explorative metabolomics investigations produce metabolite profiles and rely on the scientific method of induction to generate hypotheses in order to identify metabolites that may be indicators or biomarkers of disease or used as monitors of therapeutic responses (Goodacre 2010). Tuberculosis (TB), in its variety of manifestations – such as pulmonary and pleural TB, as well as TB within the central nervous system (CNS-TB), with tuberculous meningitis (TBM) as one of its most dangerous forms – is high on the list for such applications. Although still limited in number, metabolomic technologies have been applied successfully to diseases of the CNS. The

composition of cerebrospinal fluid (CSF), which is partially derived from interstitial fluid in the CNS, is anticipated to reflect the normal and pathological biochemical processes of the brain. In the present review, we start by addressing the pathophysiology and severity of TBM and follow with a discussion on the application of metabolomics in enriching our understanding of this infectious disease. Research on TB, in its broadest sense, has a long-standing history in South Africa, emanating from several leading research units and centres, and directed and staffed by world-renowned scholars. Their contributions warrant a review alone and are not covered here; our main focus is on TBM, the latest pioneering investigations using nuclear magnetic resonance (NMR) metabolomics, and perspectives on methodological approaches that further exploratory insights and hypotheses formulated through metabolomics studies.

Tuberculosis is caused by the bacterium *Mycobacterium tuberculosis*. Although TB commonly presents as a pulmonary disease, systemic spread of the tubercule bacilli can lead to extra-pulmonary forms of TB (EPTB), present in up to 20% of reported TB cases (Rieder *et al.* 1991), and particularly common in young children and immuno-compromised individuals (Perez-Velez & Marais 2012). CNS-TB represents up to an estimated 10% of all EPTB (and 1% of total TB) cases (Bhigjee *et al.* 2007; Rieder *et al.* 1991). According to the World Health Organization Global TB Report for 2014, there were 312 380 clinically defined new cases of TB in South Africa alone; of these, 37 709 (12%) were EPTB, which included an estimated 3 771 cases of CNS-TB. Among the various manifestations of TB, TBM is not only the most prevalent form of CNS-TB but is also the most severe complication of TB and the most common form of bacterial meningitis in children below the age of 13 years (Wolzak *et al.* 2012), especially in the Western Cape Province of South Africa (Donald *et al.* 1996; van Rie *et al.* 1999).

Recent reviews have covered various aspects of childhood TBM (Chiang *et al.* 2014; Thwaites *et al.* 2013; van Toorn & Solomons 2014). The general consensus emerging from these reports is that: (1) progress in understanding, prevention and treatment of TBM remains inadequate; (2) childhood deaths and disability as a result of TBM constitute a major societal burden; (3) although preliminary results obtained with the Bacillus Calmette–Guérin (BCG) vaccine seem to be promising for protection against TB, future trials of candidate vaccines are still needed; and (4) novel research to improve timely and accurate diagnosis and treatment outcomes remains high on the agenda. TBM is a disease that is largely unexploited by metabolomics technology, which predominantly makes use of hyphenated mass spectroscopy (MS) and NMR spectroscopy – the latter being the focus of this overview.

5.3 Pathogenesis and severity of TBM

Tuberculosis is spread via an aerosol route from persons who have active disease. Most people infected with *M. tuberculosis* have latent disease and are asymptomatic. However, a small proportion of individuals go on to develop active TB – with some developing systemic TB, such as TBM – and have significant morbidity and mortality, even though effective treatment is available.

In the late 19th century, it was thought that TBM resulted from haematogenous spread to the meninges, as a result of the frequent finding that TBM and miliary TB were occurring in the same patients (Hektoen 1896). In 1933 Rich & McCordock published their report that, in the majority of post-mortems, a single caseous focus (Rich's focus) could be found from which, when ruptured, bacilli could spread to the subarachnoid space. It was thought that this single lesion was almost always older than all the lesions found in concurrently occurring miliary TB, which set the paradigm for understanding the pathogenesis of TBM. However, this model did not fully explain the frequency of miliary TB and TBM occurring simultaneously (Donald *et al.* 2005), nor the mechanism whereby *M. tuberculosis* spreads from the lungs to the meninges and crosses the blood–brain barrier (Thwaites & Schoeman 2009). Animal models have been developed in order to understand better the pathogenesis of CNS-TB but findings are frustrated by the poor human clinical–pathological correlation (Thwaites & Schoeman 2009). Magnetic resonance imaging has detected numerous concurrent leptomeningeal granulomas in children with miliary TB, further challenging Rich's pathogenic model (Thwaites & Schoeman 2009; Janse van Rensburg *et al.* 2008). It is likely that early haematogenous spread to the brain occurs before a T-cell mediated immune response is activated. This mechanism could explain the vulnerability to TBM when T-cell mediated immunity is sub-optimal in persons infected with the human immunodeficiency virus (HIV) (Thwaites *et al.* 2013) and in persons on long-term glucocorticoid therapy (Horsburgh & Rubin 2011). The protective role of lymphocytes is essential, with the contribution of CD4+ and CD8+ T cells, along with macrophages in isolating and engulfing *M. tuberculosis*; together, these eventually lead to granuloma formation. We, and others, have shown that many of the signs, symptoms and sequelae of TBM result from an immunologically directed inflammatory response to the infection (Kashyap *et al.* 2010; Visser *et al.* 2014).

Greater understanding of the entry of *M. tuberculosis* into the CNS and the immunological mechanisms allowing survival of the bacilli is crucial for improving prevention and treatment. Fluctuation in cytokine levels affecting immunological function in patients with TBM can

directly influence the duration of the disease and its severity (Visser *et al.* 2014). In particular, the balance between pro- and anti-inflammatory cytokines may be crucial to TBM disease progression (Kashyap *et al.* 2010); this relationship is reflected by the upregulation in CSF of pro- and anti-inflammatory cytokines in patients with TBM compared to patients with other forms of meningitis (Donald *et al.* 1995; Patel *et al.* 2011). The early clinical presentation of TBM is often non-specific, with symptoms such as cough, loss of weight, fever, vomiting and malaise. As the disease progresses, meningism, focal neurological signs, and a depressed level of consciousness can occur. The timing of initiation of treatment is the most critical factor affecting morbidity, mortality, and health care costs, which emphasizes the importance of early diagnosis of TBM (Schoeman *et al.* 2002). Through a large retrospective study of 554 children with TBM, we established that 14% presented with visual impairment, 16% with hearing loss, 44% with motor deficits and 77% with intellectual impairment; only a small number (16%) did not have neurological manifestations of TBM (van Well *et al.* 2009).

Accurate prediction of outcome in childhood TBM is difficult owing to the diversity of underlying pathological mechanisms and variation in host immunological response. Multidrug-resistant (MDR) TBM in children has a poor clinical outcome and is associated with increased mortality (Seddon *et al.* 2012). In-patient mortality rates do not differ between HIV-infected and non-HIV-infected children with TBM; however, mortality after hospital discharge is greater in HIV-infected children with TBM because of HIV-related illnesses (van Toorn *et al.* 2014). Even though outcome prediction is problematic, the clinical stage of disease at the time of starting treatment is predictive of prognosis (van Toorn & Solomons 2014). Children with Stage I TBM disease are likely to lead a normal life without neurological complications, whereas those with Stage III disease have a high risk of mortality (van Toorn & Solomons 2014). Of all the TBM staging systems, the refined Medical Research Council scale has been shown to have the highest predictive value (British Medical Research Council 1948; Toorn *et al.* 2012). The criteria that define each stage are as follows: Stage I — Glasgow Coma Scale (GCS) of 15 and no focal neurology; Stage IIa — GCS of 15 plus focal neurology; Stage IIb — GCS of 11–14 with focal neurology; and Stage III — GCS <11 (British Medical Research Council 1948; Toorn *et al.* 2012). The importance of early diagnosis corresponding to an early TBM stage is emphasized by the significantly increased risk, as the disease progresses, of severe motor deficit and degree of intellectual disability in children classified as Stage IIb compared to those classified as Stage IIa (British Medical Research Council 1948). Stage III disease carries a 73% risk of the patient developing spastic quadriparesis and a mean developmental quotient of 44 (British Medical Research Council 1948). Besides the prediction of disease outcome according to TBM stage,

convulsions, headache, motor deficit, brainstem dysfunction and cerebral infarctions on neuroimaging were independently associated with poor clinical consequences of TBM in multivariate analysis (van Well *et al.* 2009).

5.4 CSF diagnostic indicators of TBM

Cerebrospinal fluid findings are integral to the diagnosis of TBM. Typical CSF findings consist of leukocytosis with lymphocyte predominance, elevated protein, and abnormally decreased CSF glucose (hypoglycorrhachia) (Thwaites *et al.* 2009). Hypoglycorrhachia is indicated by either decreased CSF to plasma glucose ratio or reduced CSF glucose, and is unaffected by HIV co-infection (Marais *et al.* 2010). A uniform research case definition for TBM identified CSF criteria to aid in the diagnosis of TBM, including a CSF to serum glucose ratio of less than 0.5 or an absolute CSF glucose concentration of less than 2.2 mmol/L and elevated CSF protein of gretater than 1 g/L.

There is potential for the use of CSF lactate in children with TBM, as CSF lactate has been shown to differentiate between bacterial and aseptic meningitis (Huy *et al.* 2010). CSF lactate levels, unaffected by serum lactate concentration, were significantly higher in adult TBM patients who subsequently died, reflecting the severity of cerebral hypoxia and therefore the overall prognosis (Thwaites *et al.* 2003). The diagnostic utility of CSF indicators considered in TBM has rarely been described, however. Studies evaluating CSF IGRA (interferon-gamma release assay) showed good sensitivity and specificity (Thwaites *et al.* 2013), but the low CSF volumes in children are a limitation when sufficient cells are required to perform IGRA (typically 5–10 mL CSF is required) (van Toorn & Solomons 2014). A study of the host immune response to *M. tuberculosis* showed the potential value of CSF interleukin-13, vascular endothelial growth factor and cathelicidin LL-37 as indicators when differentiating TBM from other forms of meningitis (Visser *et al.* 2014). Understanding the host immune response is key to a better understanding of the pathophysiology, the clinical presentation of TBM and the treatment of the disease. Novel indicators from CSF, including those from validated metabolomics studies, have the potential to be extremely useful as diagnostic tools in clinical practice.

5.5 NMR metabolomics methodology and applications

Metabolomics involves the simultaneous and comprehensive analysis of the measured responses of various biologically relevant small molecules, collectively referred to as the metabolome. The metabolome in turn is defined as the quantitative and qualitative collection of all low molecular weight molecules (metabolites, the end products of gene expression) of interest; they are found in concentrations varying in magnitude from picomoles to millimoles, they originate from endogenous or exogenous sources, and they exhibit a wide range of physico-chemical properties (Dunn & Ellis 2005). These small molecules display characteristics which allow metabolomics to avoid several challenges associated with genomics and proteomics investigations. The great chemical complexity of the metabolome, however, means that a single analytical technique is not sufficient to provide comprehensive characterization of all metabolites (Dunn & Ellis 2005). Among the techniques available, NMR spectroscopy provides an attractive alternative to MS. Although it is less sensitive than its MS-based counterparts, NMR (1) allows for a robust and global look at the metabolome – it detects all classes of metabolites – making it non-biased; (2) is highly specific, allowing for detection of novel compounds; (3) running costs are markedly lower than that of other platforms, although initial set-up is expensive; (4) requires minimal sample preparation, and hence (5) does not chemically alter nor destroy the sample under investigation. NMR thus provides an excellent early or explorative view of the metabolome under investigation (Kell & Oliver 2003).

The potential of NMR metabolomics in research on infectious human diseases has generated unique and novel insight into the underlying pathogenesis of these diseases and has unveiled new metabolic markers for disease diagnosis (Duarte *et al.* 2014). Infectious diseases relevant to Africa that have been studied using NMR metabolomics include human African trypanosomiasis (Wang *et al.* 2008), schistosomiasis (Wang *et al.* 2004), malaria (Li *et al.* 2008a), Leishmaniasis (Gupta *et al.* 1999), pneumonia (Slupsky *et al.* 2009), and HIV (Hewer *et al.* 2006), as well as TB and TBM (discussed below).

Table 5-1: Insights offered by nuclear magnetic resonance (NMR)-based metabolomics studies specific to tuberculosis (TB) and meningitis.

Pathogen	Disease	Experimental subjects	Insights	Reference
Mycobacterium tuberculosis	Tuberculosis	Adults	Patients with TB were distinguished from healthy controls. Metabolic profile of sera of patients with TB indicated significant dysregulation of metabolic pathways, validating metabolic profiling of *M. tuberculosis*-infected murine models (Shin *et al.* 2011). Potential to develop novel clinical tools for TB diagnosis or therapeutic monitoring and improve understanding of disease mechanisms.	(Zhou *et al.* 2013)
M. tuberculosis	Tuberculosis	Adults	Plasma metabolite profile of patients with TB exhibited specificity when compared to representative metabolism-related diseases (diabetes mellitus), wasting diseases (malignancy), and lung inflammatory diseases (CAP). The closest comparison to the metabolism profile of TB was that of malignant tumours.	(Zhou *et al.* 2015)
Various	Bacterial meningitis	Children and adults	NMR-based analysis of cerebrospinal fluid (CSF) is feasible and a potentially more powerful diagnostic tool than conventional rapid laboratory indicators for distinguishing bacterial from viral meningitis (as well as distinguishing between species of bacteria), with important implications for early management, reduced empirical use of antibiotics, and treatment duration.	(Coen *et al.* 2005)
M. tuberculosis	Tuberculous meningitis	Infants and children	Differentiation between TBM and non-TBM cases based on 16 NMR-derived CSF metabolites – perturbed glucose, highly elevated lactate and several energy-related metabolites, signalling components and amino acid shuttles. Postulated the following hypothesis: 'The host's response to neural infection results in an "astrocyte–microglia lactate shuttle" that operates in neuroinflammatory diseases, such as TBM.'	(Mason *et al.* 2015)
Various	Meningitis	Children	Integration of NMR spectral information with routine clinical features that are incorporated into an in-house software system which allows for the rapid differential diagnosis of meningitis (bacterial, TB and viral).	(Subramanian *et al.* 2005)

5.6 NMR metabolomics of TB and TBM

The potential of metabolomics has been well documented, specifically with regard to TB (Parida & Kaufmann 2010). Its value is in the ability to (1) address simultaneously a wide array of metabolites from various different biochemical pathways; (2) detect and isolate patterns of disturbance for additional, targeted investigations; and (3) generate hypotheses to be tested. Godreuil *et al.* believe that the new quantitative and bioinformatics approaches to the study of the interaction between *M. tuberculosis* and the infected host, and how this interaction influences the infection process, are of particular importance, as it is fully accepted in the scientific description of infectious diseases that the outcome of transmission, infection and disease is dependent on the intrinsic characteristics of both the microbes and the host.

It is expected that at different stages of the *M. tuberculosis* life cycle – for example, dormancy, latent infection and active disease – there will be a different array of host- and *M. tuberculosis*-derived metabolites (Parida & Kaufmann 2010). These metabolic profiles can best be measured globally via NMR metabolomics (Table 5-1), which has been used successfully to provide insight into the metabolic changes in host response. In a pioneering NMR-based metabolomics pilot study of the CSF of patients with bacterial meningitis, Coen *et al.* detected: (1) elevated lactate and severe glucose depletion; (2) impairment of the citric acid cycle caused by reduced production of acetyl CoA, resulting in accumulation of pyruvate and generation of amino acids via transamination; (3) elevated CSF concentrations of pyruvate and amino acids – particularly alanine, isoleucine and leucine; and (4) the presence of ketone bodies – 3-hydroxybutyrate, acetoacetate and acetone – indicating compensatory response to glucose depletion and reduced ATP levels. Notably, the role of amino acid perturbation in meningitis is further supported by other studies (Qureshi *et al.* 1998). The metabolite profile described by Coen *et al.* for BM corresponds to recent MS-based results from our research centre on CSF from a TBM patient (Mason *et al.* 2014) and to sputum from patients with pulmonary TB (Du Preez & Loots 2013). The pilot study of Coen *et al.* thus highlights the potential of NMR metabolomics in providing a global and unbiased view of metabolic or neurological perturbations in infectious diseases, particularly meningitis.

A study by Zhou *et al.* (2013) unambiguously identified 35 metabolites in sera of TB patients, of which 17 were altered. The majority of altered endogenous metabolites in the serum of TB patients were energy related – citric acid cycle intermediates, products of glycolysis, amino acids and indicators of enhanced lipid degradation – which is consistent with other studies. The metabolic processes found to be most significantly altered in TB patients were protein

biosynthesis, followed by alanine metabolism, phenylalanine and tyrosine metabolism, and ammonia recycling. Zhou *et al.* (2013) go on to state that increased glycolytic products could be an index of tissue hypoxia and extent of necrosis as the infection progresses, that enhanced lipid degradation can be correlated with caseation of host TB granulomas, and that increased formate reflects an increased requirement for nucleotide biosynthesis, all of which indicates active host inflammatory cell division. A follow-up study by Zhou *et al.* (2015) highlighted the specificity of the metabolite profile of TB patients, compared to similar diseases such as representative metabolism-related diseases (diabetes mellitus), wasting diseases (malignancy), and lung inflammatory diseases (CAP). Consistent with their previous study, they found in the plasma: increased levels of lactate, pyruvate, lipids and ketone bodies, and decreased levels of glucose, glutamate, glutamine, glycerophosphocholine, very low-density lipoproteins and branched-chain amino acids. While Zhou *et al.* (2015) could distinguish TB patients statistically, the greatest statistical overlap occurred with the plasma profiles of patients with malignancy, indicating that, to some extent, the metabolism of TB mimics that of tumour cells. NMR metabolomics has also been used to elucidate the mechanism of action of specific medication, such as the second-line drug for TB, D-cycloserine, which is primarily used to treat MDR-TB (Halouska *et al.* 2007).

We have recently used a gas chromatography–mass spectrometry (GC-MS) approach to investigate the analytical repeatability of a CSF sample collected from a TBM patient for metabolomics analysis (Mason *et al.* 2014) in order to develop a new method for evaluating repeatability in generating metabolomics data. This study clearly indicated the severe disease state of the TBM patient from whom the sample was obtained. We subsequently embarked on a comprehensive untargeted proton magnetic resonance (^{1}H NMR)-based metabolomics analysis on lumbar CSF samples from three experimental groups: (1) South African infants and children with confirmed TBM; (2) treated, non-meningitis South African infants and children as controls; and (3) age-matched untreated controls from the Netherlands (Mason *et al.* 2015). Our primary focus was on identification of metabolites that distinguish between TBM and non-TBM cases; a total of 16 NMR-derived CSF metabolites were identified. These metabolites did not include cyclopropane, a putative marker for TBM defined in 2005 (Subramanian *et al.* 2005). The defining two metabolites were the combination of perturbed glucose and highly elevated lactate, common to some other neurological disorders. Glucose, the primary source for energy production, is utilized in a rapid fashion, resulting in a significant increase in CSF lactate, both of which reflect a metabolic burst. The remaining 14 metabolites of the host's response to TBM were likewise mainly energy-associated indicators. Citrate and pyruvate, along with acetate in the form of

acetyl-CoA, are key metabolites involved in the citric acid cycle as part of energy production. Alanine and the branched-chain amino acids – valine, leucine and isoleucine – have vital roles as shuttling systems transporting metabolites across cell membranes; they are also anaplerotic, replenishing citric acid cycle intermediates. Myo-inositol acts as an important signalling component necessary for immune responses such as microglia activation. The remaining other metabolites also have important roles; one such is dimethyl sulfone, which is an indicator of response to oxidative stress; another is choline, an important precursor to acetylcholine, which in excess is responsible for seizures.

The most conspicuous outcome of the NMR profiling was the high concentration of lactate in the TBM patients (7.36 mmol/L) relative to the two control groups (1.39 mmol/L and 1.70 mmol/L for the Dutch and South African controls, respectively). This observation has some clinical significance. It has been established (Cunha 2006) that CSF lactate is applicable as a marker that can differentiate bacterial meningitis (>6 mmol/L) from other forms of meningitis, such as aseptic meningitis (<2 mmol/L). Thus, CSF concentrations of lactate, as well as immune biomarkers, are typically high before treatment and then decrease rapidly with therapy. In this regard a study (Genton & Berger 1990) on adults suffering from meningitis indicated that death was associated with high initial CSF concentrations of lactate and low numbers of white blood cells, in particular neutrophils. This outcome underscores the authors' proposal that CSF lactate is a good single indicator of the TBM disease state, making it better than conventional markers.

Furthermore, our NMR-profiling results (Mason *et al.* 2015) provided a basis for hypothesis formulation of the pathogenic characteristics of TBM. We speculated that the inflammatory responses and metabolic imbalances created in the CNS following the initial phases of infection by a pathogen, such as *M. tuberculosis*, should advantage the microglia (the native macrophage in the brain) by fulfilling their energy-intensive immune-protective function. This speculation was made against the multi-paradigm background of the current literature and based on (1) the shift in energy requirements as a result of the neuroinflammatory responses in TBM; (2) the notion that microglia play an important role in the neuropathogenesis of TBM; (3) recognition of the unique metabolic plasticity of cell–cell communication and regulation in CNS metabolism; and (4) the clinical profile on dysfunctional neuron activity seen in TBM. From observations and through inductive reasoning (Goodacre *et al.* 2004) on the characteristics of the cell–cell interactions and factor isoforms, we advanced the following hypothesis: 'The host's response to neural infection results in an "astrocyte–microglia lactate shuttle" (AMLS) that operates in neuroinflammatory diseases, such as TBM', represented as

a conceptual model detailed in the supplementary material of Mason *et al.* (2015) — Annexure 1.

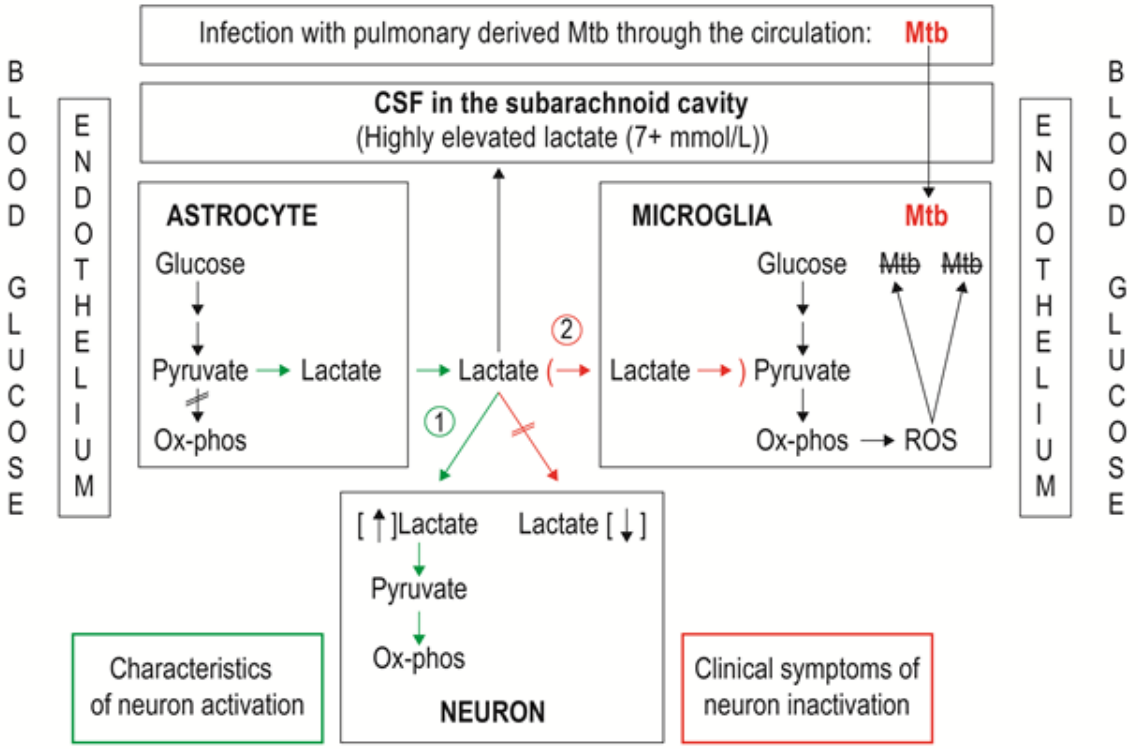


Figure 5-1: **Representation of metabolic pathways of two lactate shuttles within the central nervous system.**

(1) The astrocyte–neuron lactate shuttle (ANLS; green pathway)[58] is operative under normal physiological conditions with astrocytes responding to glutamatergic activation by increasing their rate of glucose utilisation and release of lactate in the extracellular space, making the lactate available for neurons to sustain their energy demands.

(2) The hypothetical astrocyte–microglia lactate shuttle (AMLS; red pathway) (Mason *et al.* 2015) proposed for tuberculous meningitis TBM, with astrocytes responding to signalling from *Mycobacterium tuberculosis* (Mtb)-infected microglia by increasing glucose mobilisation, leading to increased extracellular lactate, reflected in increased cerebrospinal fluid (CSF) lactate levels, available as an additional energy source for reactive oxygen species (ROS) production aimed at destroying the invading Mtb.

A simplified representation of the conceptual model is presented in Figure 5-1. Briefly, it is postulated that, in TBM, lactate produced through glycolysis in astrocytes participates in the activated immune response and is directed from the neurons preferentially into microglia, where it enters the mitochondrial citric acid cycle, contributing to oxidative phosphorylation and hence producing high levels of adenosine triphosphate (ATP) and formation of reactive oxygen species (ROS) required for *M. tuberculosis* degradation. ROS, and a multitude of factors produced by the microglia to modulate the functions of surrounding immune cells, are toxic to neurons. Moreover, the unregulated activation of microglia in response to stimulants such as *M. tuberculosis* propagate neural injury (Block & Hong 2005) and eventual apoptotic cell death for the over-activated microglia. By contrast, several investigations indicate that astrocytes could release large amounts of lactate under conditions of high energy demand by the neurons, leading to our hypothesis of an activity-dependent astrocyte–neuron lactate shuttle for the supply of energy substrates to neurons (Pellerin *et al.* 1998).

From the preceding overview we highlight the following insights that have transpired from the NMR metabolomics of TB and TBM.

Advantages of NMR technology for clinical applications

(1) Non-invasive sample collection methods are often essential in clinical medicine, for which body fluids such as urine, cord blood and serum/plasma are particularly suited, with NMR spectroscopy providing a cornerstone for their metabolomics investigations. In addition, NMR spectroscopy can work with small sample volumes.

(2) NMR analyses require a minimum of sample pre-preparation and thereby prevent the occurrence of confounding analytical artefacts.

(3) Metabolomics produces large and complex data sets that cannot be interpreted through classic reductionist methods, but require an inductive, open-minded approach, aimed at hypothesis generation, which is philosophically a useful methodological approach in uncharted fields of research.

(4) Clinically, perturbations in metabolism are often difficult to recognise as the phenotypes may show considerable variation; however, NMR technology has provided a proven track record through the identification of inborn errors of metabolism, often already manifesting in neonates.

(5) The unique advantage of metabolomics technology is that it provides a dynamic view of host functional responses in health and disease, and offers early and rapid identification of the status and progress of a disease.

Cost-effectiveness of NMR technology

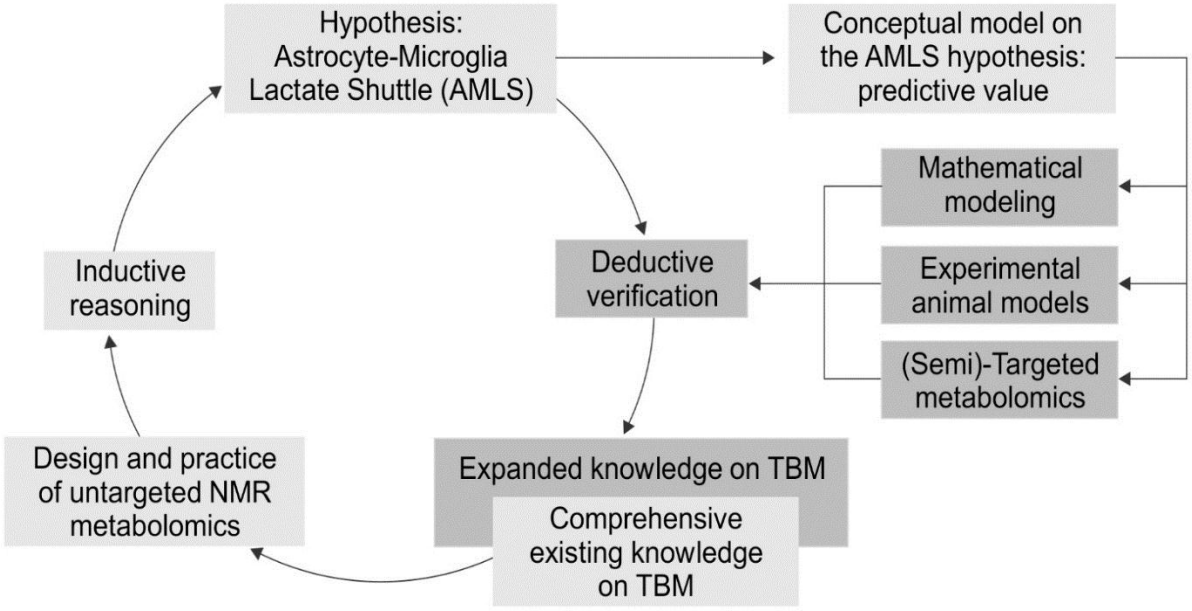
High-resolution NMR equipment, such as the instrument installed in early 2015 in a collaboration between the Technological Innovation Agency (TIA) and North-West University (NWU), is extremely expensive and requires highly skilled technical and scientific expertise for its operation. However, once an NMR facility has been established, the analytical clinical analysis is relatively inexpensive, costing only a few rands per sample to run. Moreover, the international trend in the market for miniaturisation of NMR instrumentation is well established, and is opening up new areas of study in academic and industrial settings with clear potential and appeal for future clinical practice (Lee *et al.* 2008).

Systems insights

Fields that are data rich but hypothesis poor – prime examples are systems biology and many areas of medicine and clinical practice – are best approached through data-driven, essentially inductive methods. Biological systems are inherently complex; they require an inductive approach to data generation, data analysis and modelling if we are to continue to make strides in our understanding of these phenomena in health and disease. The advent of multidisciplinary interaction between biologists/clinicians, analytical chemists and informaticists/biostatisticians in developing complex models has ushered in a new era in the mathematics of science, for the first time enabling understanding and prediction of large-scale biological systems; the 'omics' technologies are key in this process by being inductive and exploratory in their very essence. Intrinsic to this process remains the iterative relationship between the realm of ideas/hypotheses/thoughts (= induction) and that of observations/data/information (= deduction), linked in a repetitive cyclic way in which one arc is not simply the reverse of the other, but has methodological characteristics in its own right. We conclude our overview by focusing on this cyclic relationship.

5.7 Directives for hypothesis verification

Hypotheses are generalized statements derived from empirical evidence and prior knowledge of existing data, and need to be verified or refuted by deductive methods (Figure 5-2). This activity can be represented as a circular process with one half of the iterative cycle of knowledge being inductive reasoning; the second half (deduction) aims at testing hypotheses and furthering existing knowledge. This representation highlights the essential nature of exploratory metabolomics investigations, and features three approaches that could be applied for hypothesis testing: (1) computational models, (2) animal models and (3) semi-targeted metabolomics.



TBM, tuberculous meningitis; NMR, nuclear magnetic resonance

Figure 5-2: Iterative cycle of knowledge, using the astrocyte–microglia lactate shuttle (AMLS) hypothesis as an example. The existing knowledge forms the basis of inductive reasoning that leads to the formulation of hypotheses, which in turn is followed by the use of deduction to verify these hypotheses and to further existing knowledge.

Computational modelling

The extreme complexity of intracellular regulatory networks, involving genetic circuits, metabolic regulation and signal transduction, hampers the ability to analyse these functions quantitatively to achieve a holistic understanding of their role in health and disease. Complexity theory has thus become established over the last two decades as a novel discipline directed to such problems in the humanities and natural sciences, and is at present pursued in South Africa at the Centre for Studies in Complexity established in 2009 at Stellenbosch University. Modelling, in its various forms, demands key competencies in studying complexity. Models are used to describe complex systems through computer simulation based on quantitative information on many and key variables that characterize what is being investigated. An example is the computational model we recently developed – using in silico data that capture the unique immunological environment of the brain – which allows us to study the key mechanisms driving TB-linked granuloma formation in time (El-Kebir *et al.* 2013). The model was based on a known pulmonary agent-based description of TB, representing the molecular signalling networks that affect granuloma formation during infection with *M. tuberculosis.* It focused on the formation of granulomas whose structure

and function may reflect the success or failure of the host to contain infection. We envisage that computational model development, based on NMR data, may be a fruitful approach to use for understanding the dynamics involved in metabolic flux associated with the AMLS hypothesis, thereby permitting greater comprehension of the perturbations following *M. tuberculosis* infection of the microglia in the meninges.

Experimentation using animal models

Animal models have contributed significantly to knowledge of TB. However, there are certain major differences between humans and other animals in their responses to the disease. It is known that mice, for example, do not form the granulomas seen in humans and most other vertebrates with TB. Such differences caution our interpretations of data from animal models as they apply to human diseases. Nonetheless, animal models are indispensable to TB research and are well established in South African TB research centres (de Souza *et al.* 2010; Drennan *et al.* 2004). Until recently, for example, no murine model was available for experimental studies on TBM. We successfully developed a reproducible *in vivo* murine model to study the inflammatory response in TBM (van Well *et al.* 2007), and obtained useful insights despite limitations in translating the experimental observations to the human manifestation of the disease. Although the route of infection mimics the way TBM is acquired naturally, the cytokine profile in the mice does not show the clinical characteristics of the human disease. Nonetheless, the model has potential for further experimentation on *M. tuberculosis* infection because we observed that bacterial growth of TB in the CNS leads to a typical chronic inflammatory response. Such results indicate that experimental animal studies hold potential for investigating the biological changes in a *M. tuberculosis*-infected model at the biochemical level – an approach that could be expanded through appropriate knock-out models, as has been shown for cryptococcal meningitis (Lee *et al.* 2010a).

Semi-targeted metabolomics

The metabolites present in the CSF represent the actual metabolism of the CNS and the balance between blood and CSF, rendering CSF analysis indispensable in the assessment of neurological disorders. Sampling human CSF is, however, an invasive procedure and ethical and safety constraints limit the availability of such samples for TBM, especially in work with infants and children. Thus, accessing these samples is much more difficult than collecting blood or urine. Metabolomics offers promising options. The use of MS-based metabolomics of plasma from patients suffering from propionic acidemia and methylmalonic acidemia (Wikoff *et al.* 2007), as well as our GC-MS metabolomics analysis of urine samples from patients with respiratory chain deficiencies (Reinecke *et al.* 2012) and isovaleric academia (Dercksen *et al.* 2012), have provided proof of concept that semi-targeted

metabolomics can expand the range of metabolites associated with human disease. These studies also indicate the potential of metabolomics for non-invasive diagnosis and clinical evaluation of patients with TBM.

5.8 Perspective

Metabolomics studies of complex diseases are still in their infancy, although the volume of NMR- and MS-based metabolomics publications in particular is increasing exponentially.[2] As it stands, NMR metabolomics is not a standalone method of diagnosis nor, by extension, of disease management of TB or of TBM, but it undeniably offers insights for potentially improving existing knowledge of infectious diseases in ways that may complement and benefit current diagnostic and treatment methods. Recent results, such as those by Zhou *et al.* (2015), strongly indicate that NMR metabolomics can contribute to improved understanding of disease mechanisms. However, a major gap in work on infectious diseases is still the paucity of subsequent validation of clinical studies, which is necessary to reap the full rewards of metabolomics technology. Ultimately, the optimal outcome, particularly in resource-limited settings, is the validation and declaration of simple and clear biomarkers for disease diagnosis – which requires appropriate and thorough follow-ups on metabolomics studies.

Fortunately, the great potential of biotechnology was recognised by the Department of Science and Technology, which enabled the creation of several technological platforms, including the Metabolomics Platform supported by the TIA and hosted by NWU on its Potchefstroom campus. This platform consists of a wide array of gas and liquid chromatography based MS instruments, as well as a 500-MHz NMR spectrometer dedicated to investigations of biofluids and other biological specimens. Furthermore, the contractual agreement between the TIA and NWU stipulates that: '...the Institution [NWU] agrees to have an open policy that encourages use of its facilities by researchers other than those based at the Institution…'. This policy of open access to the Metabolomics Platform aligns with the TIA's objective of stimulating and exploiting technological innovation in order to improve economic growth and the quality of life of all South Africans – a view that we strongly share, with the hope that this review will further promote these far-reaching ambitions.

Acknowledgements

Research funding for this project was provided by the Technological Innovation Agency of the Department of Science and Technology of South Africa. S.M. and R.S. are recipients of a Desmond Tutu–NRF–VU doctoral fellowship for a joint PhD study between the Vrije Universiteit in Amsterdam (the Netherlands), and North-West University and Stellenbosch University in South Africa.

Authors' contributions

S.M. and R.S. wrote the manuscript. C.J.R and A.M.v. F. are the South African and Dutch project leaders, respectively, who directed the study and contributed to the writing of the manuscript.

PART 3: SEMI-TARGETED GC-MS METABOLOMICS: A WALK ON THE ANALYTICAL SIDE OF GC-MS AND APPLICATION TOWARD AN URINARY PROFILE AND BIOSIGNATURE OF TBM

CHAPTER 6 GC-MS: A 'GOLD STANDARD' IN METABOLOMICS

6.1 Introduction

One of the most widely used methods in metabolomics studies is GC-MS (Dunn & Ellis 2005; Weckwerth & Fiehn 2002) — often termed the 'gold standard'. A well-designed GC-MS metabolomics experiment is dependent on well-defined, standardized analytical protocols (Want *et al.* 2010) and reporting procedures (Goodacre *et al.* 2007) in order to produce reliable, useful data. It is important to formulate a robust experimental design and to investigate the potential influences of variation due to the analytical procedure to ensure that any perturbations detected in the metabolic profiles are a consequence of the biological response being studied and only minimally (ideally, not at all) from the analytical method.

Biological variation can be categorized into 3 types (Want *et al.* 2010):

- Induced/relevant — experimental conditions designed to induce a desired biological variation caused by a pathological condition of interest.
- Non-relevant — intra-individual and inter-individual variation that can be monitored, accounted for and, to an extent, controlled.
- Uninduced — variations in certain metabolic concentrations under identical experimental conditions.

Thus, total variation within data generated in a GC-MS metabolomics experiment consists of the sum of all induced/relevant, non-relevant and uninduced biological variations plus all technical variations, bearing in mind that a typical GC-MS metabolomics study is geared towards analysing the overall variability of a biological origin (Shurubor *et al.* 2007).

Each variation due to the experimental procedure may have a limited (often negligible) influence on the overall experimental results if well-defined quality controls are in place. However, the cumulative effects of multiple analytical variations may influence the overall data in such a way that they cannot be neglected, and hence it is usually not possible to completely avoid analytical variation in the experimental results. The recognition, understanding and minimization of potential undesirable variations are therefore necessary in a metabolomics analyst (van Batenburg *et al.* 2011; Álvarez-Sánchez *et al.* 2010; Xu *et al.* 2010b; Teahan *et al.* 2006; Kanani *et al.* 2008).

Some GC-MS studies require method validation and quality assurance by including (1) standardized mixtures and/or (2) quality control (QC) samples in the experimental analysis (Want *et al.* 2010; Gika *et al.* 2007; Pasikanti *et al.* 2008). (1) A standardized mixture is a solution consisting of a limited, but known, number and concentration of relevant compounds prepared from commercially available standards. This standardized solution is used to assess the overall performance characteristics of a GC-MS run (such as retention time stability, peak shape, signal intensity, mass accuracy, and detector response) (Want *et. al.* 2010; Gika *et al.* 2007). Ideally, the standardized sample is best used as both the first and last samples analysed in a GC-MS batch run. Comparing the two results of examining the standardized sample determines the degree of instrument-based variation within a batch run and whether or not the results of the samples are valid. (2) A QC sample is defined here as a mixture consisting of aliquots of all samples to be analysed in a batch, also known as a 'pooled sample'. This positive control sample broadly represents the entire sample set being analysed. The QC sample is used to condition the GC column to the sample matrix to be analysed and provides data to assess the analytical variability (repeatability) (Want *et. al.* 2010; Gika *et al.* 2007). Ideally, the QC sample should be injected at regular intervals throughout the GC-MS procedure. Thus, the standardized mixture and QC sample account for any changes over time (Zelena *et al.* 2009). These quality assurance methods assess the overall features and analytical variability of the experiment, test the validity of experimental results and determine the robustness of the analytical method, instrument and data within a batch run. However, the limitation inherent in the validity of a single measurement batch means that a well-defined, quantifiable measurement of repeatability that leads to an inter- and intra-batch assessment of quality of data is not provided. Repeatability here refers to the ability of an analyst to produce precise data matrices to the point where comparison of *m* data matrices generated by the same analytical method under identical experimental conditions of *n* aliquots from the same sample exhibits variation within a defined range of deviation (that is, an acceptable critical value for measurement of repeatability), so as to be considered reliable. Thus, the evaluation of the repeatability of metabolomics data can be used to assess the reliability of data generated by an analyst.

The relative standard deviation (RSD), or coefficient of variation (CV), serves as a measurable (quantifiable) assessment of variation and repeatability (Xu *et al.* 2010). This assessment is limited, however, to a single variable (such as a metabolite) within a measurement batch (van Batenburg *et al.* 2011) and does not take into consideration the repeatability of a multi-component analysis, as is the case with a metabolite profile. In addition, it is observed that the CV approaches infinity as the mean tends to zero, a phenomenon termed the "Horwitz trumpet" by Hall & Sellinger (1989). This becomes an

issue as metabolomics data typically deal with very low metabolite concentration values which exhibit high CVs that can be incorrectly interpreted. A rapid, visual assessment of qualitative repeatability is sometimes presented in the form of a graphical overlay of chromatograms/raw data generated from repeat samples (Teahan *et al.* 2006). Such overlays are, however, complex to interpret, making assessment of repeatability misleading as the data are not normalized and do not provide a quantitative measurement of error.

6.2 Assessing GC-MS and the analyst

Each bias introduced by virtue of conducting the GC-MS procedure may have only a slight, often negligible, influence on the overall data individually; however, the accumulation and interaction of multiple analytical biases may greatly influence the overall data. Complete avoidance of bias is often impossible but recognition, understanding and minimization of potential biases are important when interpreting the experimental results.

In the following chapter (Chapter 7) we implement a well-known statistical method in a novel fashion — smoothing of raw GC-MS chromatograms (Figure 6–1).

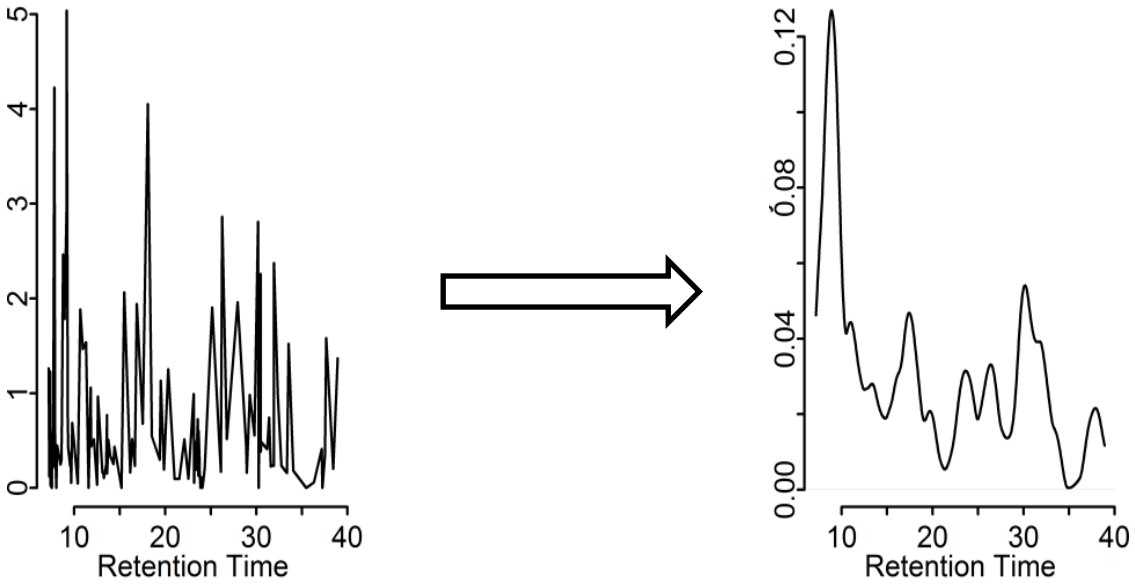


Figure 6-1: Application of smoothing to a raw GC-MS chromatogram (left) to create the density representation (KEMREP-smoothed form) of the chromatogram (right) that is easier to interpret.

The aim was to determine qualitatively if repeat GC-MS analyses of different biofluids were repeatable by different analysts, and quantitatively by adding confidence intervals. This new approach — termed KEMREP — provides compelling proof that the technical variation inherent in metabolomics data generated by GC-MS can be considered minimal if the analyst is conscientious and experienced.

6.3 Working towards a urinary biosignature for TBM

In Chapter 1, during the formulation of the biological question, the term 'biomarker' was defined. It was noted that Denoroy *et al.* (2013) advocated three levels of biomarker validation:

1) Validation of the scientific hypotheses derived from the analysis of the metabolite perturbation.
2) Technical validation by means of measuring the same perturbation by using a different technology platform or a similar platform using different settings.
3) Validation of the device used for measuring the perturbation.

The first two levels require additional samples, time and high-quality scientific output in order to fully validate our GC-MS-based results, and there was insufficient sample (volume and type) and time available to do this. However, we could partially validate our GC-MS-based results based upon level 3: *validation of the device used for measuring the perturbation* — by determining an analyst's ability to produce reliable GC-MS-based metabolomics data — using KEMREP. This was done by assessing repeatability and reproducibility among different analysts, GC-MS instruments, laboratories and types of samples (CSF and urine). This novel method (Mason *et al.* 2014) proved the reliability of the metabolomics data generated by GC-MS.

For the purposes of our study we used a pooled urine QC sample, as suggested by Want *et al.* 2010, and applied our KEMREP method to determine that our GC-MS-based method was repeatable (results discussed in Chapter 9) and that the metabolomics data generated by GC-MS are reliable. Hence, discriminating variables determined in Chapter 8 can be related to biological variation and the biological interpretation of a putative biosignature for the TBM paediatric population under study could be done with confidence.

CHAPTER 7 KEMREP: A NEW QUALITATIVE METHOD FOR THE ASSESSMENT OF AN ANALYST'S ABILITY TO GENERATE A METABOLOMICS DATA MATRIX BY GAS CHROMATOGRAPHY–MASS SPECTROMETRY

Shayne Mason[1], Gontse P. Moutloatse[1], A. Marceline van Furth[2], Regan Solomons[3], Mari van Reenen[4], Carolus Reinecke[1] and Gerhard Koekemoer[4]

[1] Centre for Human Metabonomics, Faculty of Natural Sciences, Private Bag X2006, North-West University (Potchefstroom Campus), South Africa,

[2] Department of Paediatric Infectious Diseases–Immunology and Rheumatology, Vrije Universiteit Medical Centre, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands,

[3] Department of Paediatrics and Child Health, Tygerberg Hospital, University of Stellenbosch, PO Box 19063, Tygerberg 7505, South Africa,

[4] Statistical Consultation Services, Faculty of Natural Sciences, Private Bag X2006, North-West University (Potchefstroom Campus), South Africa.

Accepted: *Current Metabolomics,* 2014, *2*(1):15–26.

7.1 Abstract

The analytical procedures required to generate a quantified metabolomics data matrix include many and widely different potential sources of error, complicating the generation of reliable data. The methods generally used to assess the precision of such data all have distinct merits but some clear limitations as well. In this paper we describe KEMREP (kernel method for the assessment of repeatability and reproducibility), a new method with the advantage and focus aimed specifically at a qualitative assessment of the precision by which analysts generate a metabolomics data matrix. Repeatability and reproducibility were assessed on gas chromatography–mass spectrometry (GC-MS)-generated metabolomics data matrices produced by and between analysts and across laboratories, using cerebrospinal fluid (CSF) and urine as biological samples for analysis. KEMREP provides a visual overlay of the smoothed and scaled versions of the data from repeated samples for a direct and easy qualitative assessment of repeatability or reproducibility of a distinct chromatographic region (univariate dimension) or for the experiment as a whole (multivariate dimension). The KEMREP method can also be extended by the imposition of confidence bounds which provide lower and upper limits that indicate quantitatively whether the experiment was repeatable or reproducible at a predefined input coefficient of variation (CV). KEMREP is thus a novel approach which supplements existing methods of assessment of reliability of metabolomics data; provides a benchmark for assessing the quality of practical work performed by analysts; monitors the sequence of data pre-treatment steps; and tests the robustness of an experimentally designed protocol for metabolomics.

7.2 Introduction

The terms “metabolome” (Oliver *et al.* 1998) and “metabolomics” (Fiehn 2002) have been well defined since they were first coined and evolved into an established field of research, which analytically culminates in producing a comprehensive qualitative and quantitative two-dimensional (a matrix) or three-dimensional (a tensor) distribution of the metabolites present in a given biological system (Álvarez-Sánchez *et al.* 2010). In this paper we focus on the qualitative assessment of the precision by which analysts generate a metabolomics data matrix by gas chromatography–mass spectrometry (GC-MS) using KEMREP, a non-parametric kernel approach used to estimate probability density through data smoothing. The main contribution of this approach is to assess the ability of analysts to generate a specified metabolomics matrix, examined by and between analysts and across laboratories. This approach proved to be valuable to assess, in-house, the technical skills of researchers

new to metabolomics analysis, and for periodical monitoring of such technical skills in analysts as a standard laboratory quality control practice. Finally, we propose how a quantitative conversion of the qualitative assessment using the KEMREP method can be introduced as well.

Analytical precision is of central importance in the generation of metabolomics matrices from biological samples (Álvarez-Sánchez *et al.* 2010; Yin *et al.* 2013), capturing qualitative and quantitative information on compounds related to one or more complex biological perturbations under investigation. Samples may contain as many as hundreds of metabolites, which constitute the columns of the matrices. In most metabolomics investigations, these data points are designated as features, containing mainly true metabolites but also contaminants, artefacts and noise that should be eliminated in the process of data pretreatment (van der Greef *et al.* 2007; Bruce *et al.* 2009). The metabolites whose concentrations are quantified are molecules of low molecular weight (<1500 Da) and products of cellular, metabolic and regulatory or pathological processes (Yin *et al.* 2013; Guy *et al.* 2008; Reinecke *et al.* 2012). Their levels can be regarded as the response of biological systems to influences that are endogenous (e.g., genetic, homeostasis or age related) or exogenous (e.g., nutrition, infection or medication related). A metabolomics analysis can thus be described as a data-driven process to characterize the quantitative composition of these low molecular weight chemicals (Fiehn 2002), with the objective of identifying those metabolites of interest in relation to the endogenous or exogenous perturbation under investigation. However, a need in current metabolomics is the further refinement and development of several aspects of metabolomics data generation and interpretation (Zhang *et al.* 2012; Benton *et al.* 2012; Worley *et al.* 2013). A key underlying aspect of all metabolomics investigations is the requirement that the analyst should produce data of a high quality (Ioannidis 2005), which is the central aspect addressed in this paper.

The analytical procedures required to generate a quantified metabolite profile include many and broadly different potential sources of error (van Batenburg *et al.* 2011; Xu *et al.* 2010), which complicates the generation and repeatability of metabolomics data; here, repeatability refers to the ability of a single analyst to produce consistent metabolite data, when using the same analytical method under identical experimental conditions. Apart from repeatability, reproducibility is another factor that affects the integrity of metabolomics matrices. Reproducibility here refers to assessment of repeatability of data generated analytically by more than a single independent source (i.e., evaluation of variation among laboratories, analysts, instruments and/or batches). Repeatability and reproducibility are thus two key terms used in assessing the overall variability of data and are essential in the validation of

experimental results (Ioannidis 2005; Lay *et al.* 2006), which requires the need to account for and ultimately eliminate or reduce unwanted analytical variability and focus on the perturbation of interest in order to produce statistical results that are biologically sensible (van den Berg *et al.* 2006; Sysi-Aho *et al.* 2007).

The relative standard deviation (RSD), or coefficient of variation (CV), are standard quantitative measures used in metabolomics investigations that give insight into different sources of variation contributing to the total variation in the data and may be applied to monitor repeatability. The assessment of variability using CVs is, however, limited to single components within a measurement batch (Urban *et al.* 2010); it also has the limitation of being unreliable at very low concentrations. Several univariate approaches (Xu *et al.* 2010; Smilde *et al.* 2009; Rocke & Lorenzato 1995) have been proposed to define measurement errors encountered in metabolomics experiments and to improve the assessment of the data; indeed, a low-level multivariate detection model (van der Voet *et al.* 1998) was advanced to assess the probability that certain peaks in GC-MS data are generated by the analytes, rather than by interferences. Van Batenburg *et al.* (2011) recently presented a further development of the univariate Rocke–Lorenzato model. It could be successfully applied to comprehensive metabolomics data, although it retained the limitation of describing the measurement error variance of single metabolites as a function of their responses.

Despite their limitations, all existing methods to assess precision of analytical data have distinctive merits in illuminating some specific aspect of the data under consideration. KEMREP has the advantage that it proposes an easily applicable tool for the convincing qualitative assessment of repeatability, as well as reproducibility. The design is based on kernel density estimation and produces as output a smoothed chromatographic profile of the original data. Any GC-MS metabolomics data matrix generated from a single biological sample, subdivided into several identical aliquots for repeated analyses, may be used to evaluate the KEMREP method. The approach focuses on analytical variability, whereas the contribution of variability due to biological and/or environmental factors is nullified.

We selected a sample from two biofluids (urine from a healthy adult and cerebrospinal fluid (CSF) from a patient infected with, and treated for, tuberculous meningitis (TBM)) to illustrate the application of KEMREP. We used aliquots from the same sample to be analysed by the same or different analysts, using the same experimental protocol but generating the data on different GC-MS platforms, to address the issue of reproducibility. The GC-MS instruments chosen were almost identical, but used for different purposes in our laboratory: one was a multi-purpose platform, used primarily for general postgraduate student teaching and research purposes, whereas the other was dedicated to routine organic acid analysis for

diagnosis of inherited metabolic disorders in our clinical-chemistry laboratory (www.pliem.co.za). The outline that we followed is presented in the flow diagram shown in Figure 7-1 and discussed below.

The practice of using a limited number of technical replicate samples to assess repeatability and reproducibility is a well-established and recognized approach in this kind of study (see Table 7-1). Such technical replicates are multiple samples, pre-processed by exactly the same protocol and used for consecutive injections into the GC-MS, to assess repeatability and/or reproducibility.

Achieving repeatability in metabolomics data is especially demanding (Fukusaki & Kobayashi 2005), which reinforces the need to include qualitative assessments of repeatability and reproducibility at various stages of an experimental design in order to produce validated and reliable results, as addressed through the KEMREP approach presented here. Against this background, our protocol followed the typical experimental and theoretical procedures for organic acid profiling (Reinecke *et al.* 2012; Goodacre *et al.* 2007; Brown *et al.* 2005), presented in detail in the supplementary information (SI) — Annexure 2. This enabled us to succeed in devising an easy and practical approach for the qualitative assessment of the repeatability and reproducibility of the comprehensive analytical process used to generate a metabolomics matrix. We believe that this new method has the potential to be widely used in metabolomics practice, and will lead to its further development and refinement. To support our claims and foster such initiatives, we include a detailed statistical description of the new method, as well as the original data, computer code for KEMREP and instructions on how to use the code, which is executable using the free statistical software program R (available for free download at http://www.r-project.org). This will enable fellow researchers to evaluate the appropriateness of the method for application on different metabolomics data sets.

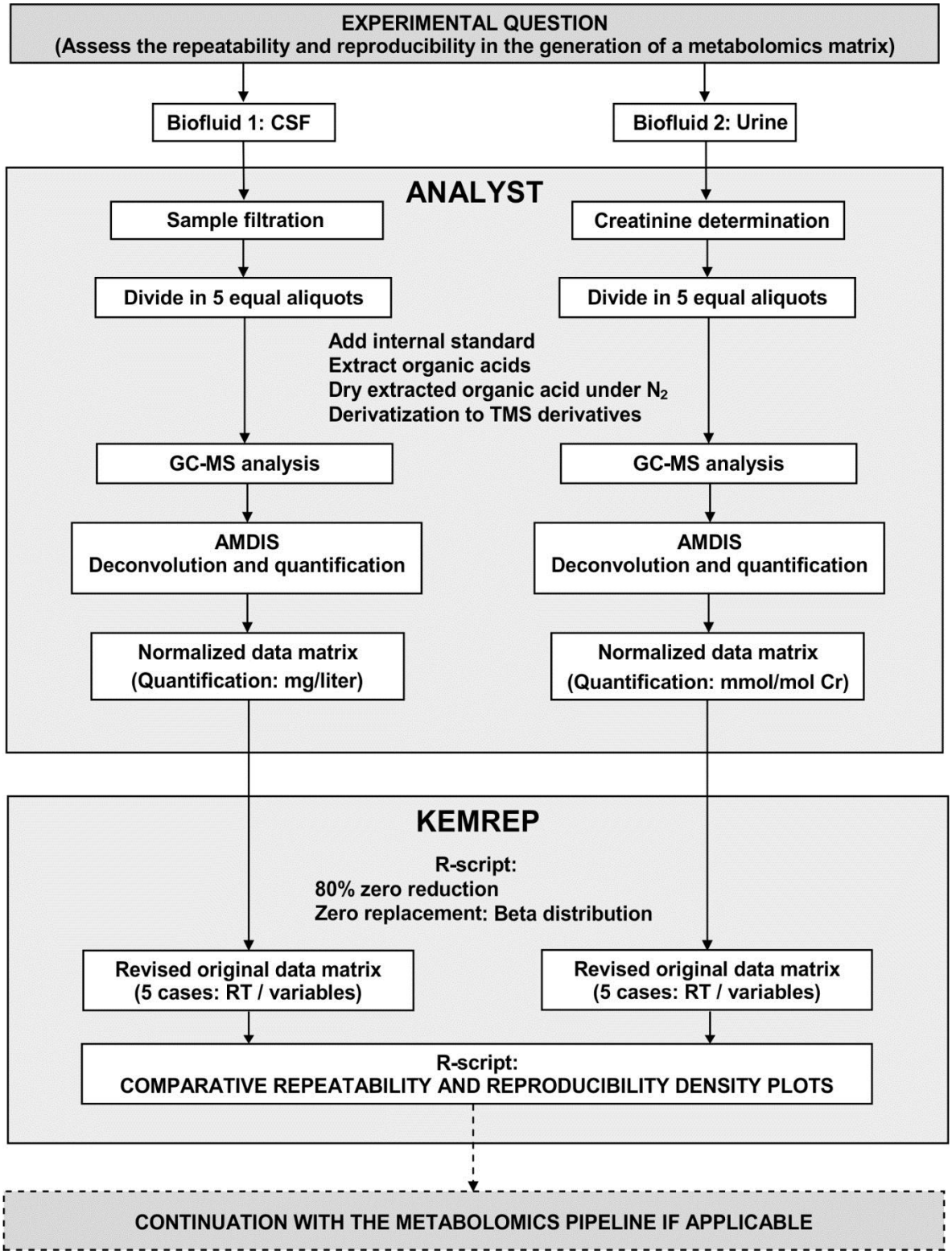


Figure 7-1: **Flow diagram illustrating experimental design focused on analytical procedures of the metabolomics pipeline followed in evaluating KEMREP.**

Table 7-1: Selection of metabolomics-based studies that use technical replicates to assess repeatability and/or reproducibility.

	Technical repeats (n)	Repeatability/Reproducibility measure
Jiménez *et al.* (2002)	5 – negative controls 4 – positive controls 4 – low concentration 4 – high concentration	RSD ratio in 11 qualitative chromatographic methods
Shepherd *et al.* (2007)	4 – short term 9 – long term	CV
van Liempd *et al.* (2013)	6 – single pooled 2 x 3 – double pooled	CV
Tredwell *et al.* (2011)	3 x 6 different analytical methods (total = 18)	ANOVA, correlation analysis, pairwise rank correlations and CV
Tanaka *et al.* (2007)	6	CV
t'Kindt *et al.* (2009)	5 – pooled quality control	CV

7.3 Materials and methods

7.3.1 Experimental design

The metabolomics pipeline, ranging from sampling to the point of new knowledge generated through the metabolomics approach, is one that has been well described not only from a holistic viewpoint (Fukusaki & Kobayashi 2005; Goodacre *et al.* 2007; Brown *et al.* 2005), but also from specialized, focal aspects (Hendriks *et al.* 2011; Liland 2011; Katajamaa & Orešič 2007; Castillo *et al.* 2011). Our design focuses on the analytical procedures in the pipeline, ranging from sample preparation to data processing, with particular attention being given to the analyst (Milman 2005) (see flow diagram in Figure 7-1). By using the same sample for repeat analysis we eliminate the interference of biological variation, thereby avoiding the initial sampling aspect of the pipeline. Similarly, by proceeding to the point precluding the

multivariate statistical analysis (data mining), we remove the need to quantify and apply biological interpretation to the data and focus on the analytical assessment, which is the prime goal of this paper.

7.3.2 Sample selection: background and assumptions

Two distinctly different types of biofluid were used in this comparative investigation. The first sample was cerebrospinal fluid (CSF) obtained from a patient suffering from tuberculosis meningitis (TBM). The patient was a young African male (4 years and 10 months old at the time of sampling), who was admitted to Tygerberg Hospital, Department of Paediatrics & Child Health, University of Stellenbosch Medical School, South Africa, with clinical symptoms indicative of meningitis (persistent headache, cranial nerve palsy, sleepiness and strabismus) and assigned a Glasgow Coma Score of 13/15 (Schutte & van der Meyden 1998). The patient was treated with a broad-spectrum range of medication immediately upon admission after which biofluid samples were used for a definitive diagnosis. TBM stage II was confirmed by polymerase chain reaction (PCR) and microbiological culture analyses. At the time of sampling, direct ventricular drainage of CSF from the brain was performed to alleviate intra-cranial pressure, yielding approximately 20 ml of CSF. This sample, designated for analysis as well as assessment of repeatability, was divided into 1.0 ml aliquots, transported to our laboratory and stored at –80ºC. After approximately 8 weeks in storage, five of the CSF aliquots were removed from the freezer and used for metabolomics analysis by one analyst, as described here.

The other sample used for the evaluation of KEMREP as an assessment tool was an early morning urine sample obtained from a single, healthy African adult female (27 years old) in an overnight fasted state under no form of medication. The urine sample can thus be considered as being uncomplicated by exogenous influences (i.e., a typical control sample).

7.3.3 GC-MS analysis of organic acids

All matrices generated by the different analysts consisted of five aliquots of a single sample, whereas the number of variables differed from case to case. The variables in the original data set are considered features (components associated with peaks in the chromatogram defined as molecular entities with unique mass spectrometric *m/z* values at unique time points, or retention times, in a chromatogram) (Tautenhahn *et al.* 2010). These compounds are later redefined as metabolites – chemical intermediates in the enzyme-catalyzed reactions of metabolism identified by the use of a database to link the retention time of a particular component to the accurate mass and structural formula derived from the model

components included in the database (Reinecke *et al.* 2012) or as defined substances – intact non-biological chemical components (e.g., derivatization by-products) ranging from those identified in the original chromatograms to some remaining in the final metabolomics data matrix, regarded as artefacts present in the original data (Wasim & Brereton 2004).

In the case of the CSF analysis, each of the five aliquots was thawed and filtered using a 10,000 Da MWCO Centrisart-1 centrifugal unit by centrifugation for 15 min at 3000 rpm to remove any interfering proteins and bacteria or bacterial macromolecules present. Organic acids in the CSF were extracted using a standard operating procedure (SOP) as described previously (Reinecke *et al.* 2012). The five prepared aliquots were analysed on two similar GC-MS instruments designated GC-MS-1 (used primarily for general postgraduate student teaching and research) and GC-MS-2 (used for dedicated routine organic acid analysis for metabolic diagnostic purposes) in two independent laboratories (see SI — Annexure 2 — for detailed SOP and complete apparatus configuration). Both instruments had similar, but not identical, GC temperature oven programs, and so yielded different retention times and total run times. Similarly, the split ratios for each GC-MS apparatus differed marginally. These slight differences in configuration simulated a real-world situation in which two independent laboratories with the same type of instrument, measuring the same subset of the metabolome, have similar instrument settings but differ slightly in function. They also provided the opportunity to test the robustness of our proposed method of assessment of repeatability and reproducibility of our experimental data with respect to the SOP, GC-MS instruments and analyst. In the case of the urine analyses, the organic acids were extracted and derivatized by the same procedure as used for the CSF (Reinecke *et al.* 2012) and analysed on GC-MS-1. The Automated Mass Spectral Deconvolution and Identification System software program (AMDIS, version 2.66 from the National Institute for Standards and Technology) was used for spectral deconvolution and identification of features.

Prior to application of KEMREP, the data were normalized through quantification and then transformed by using a shifted logarithmic transformation. The final quantified concentrations of metabolites from the CSF are expressed as mg/liter and as mmol/mol creatinine for urine (Figure 7-1). These are relative values by being volume-normalized (CSF) and creatinine-normalized (urine), respectively.

7.3.4 Development of the KEMREP method

7.3.4.1 Model description

The KEMREP method aims to provide the user with a visual overlay of the information measured (i.e., smoothed chromatograms) from sample replicates in such a way that the user can quickly and easily assess repeatability or reproducibility in a distinct chromatographic region or for the experiment as a whole. This is accomplished by presenting a plot consisting of a smoothed and scaled version of the data with bounds that satisfy repeatability or reproducibility. We selected the data of the complex CSF sample used in our analyses and implemented some data reduction methods to assist in the explanation of the development of the statistical aspects of KEMREP.

The index is used to represent a specific reproducibility item and the replicates of reproducibility item $\boldsymbol{i}$ will be denoted by $\boldsymbol{j}$. Suppose there are $\boldsymbol{I}$ reproducibility items in total, then $\boldsymbol{n_1, n_2, \cdots, n_I}$ represent the number of samples (cases) for each reproducibility item. If $\boldsymbol{p}$ variables have been detected, then $\boldsymbol{t_{i,j,1}, t_{i,j,2}, \cdots, t_{i,j,p}}$ are the observed retention times for reproducibility item *i,* replicate *j* so that $\boldsymbol{t_{i,j,k} > t_{i,j,k-1}}$ for $\boldsymbol{k} \in \{\boldsymbol{2, 3, \cdots, p}\}$. The variable $\boldsymbol{t}$ is then used to represent a continuous retention time (i.e., $\boldsymbol{t}$ can assume any value between the smallest and greatest retention times observed in the data). Since the variables are measured on different value scales, and are all used simultaneously, some form of transformation is required to make the scales more comparable, yet retain the variation in which we are interested. For this purpose we use the shifted log transformation, with shift parameter set at one. The shifted log-scaled value is defined as $\boldsymbol{X_{i,j,t_{i,j,k}}}$ for reproducibility item *i,* replicate *j*, at a fixed observed retention time $\boldsymbol{t_{i,j,k}}$, where $\boldsymbol{k = 1, 2, \cdots, p}$.

The KEMREP method obtains the smoothed chromatogram for each replicate using an accumulation of the observed normalized and transformed values (the data used to illustrate the development of the KEMREP method are plotted in Figure 7-2A). Let $\boldsymbol{\theta_{i,j,k} = \sum_{q \in \{t_{i,j,1}, t_{i,j,2}, \cdots, t_{i,j,k}\}} X_{i,j,q}}$ for $\boldsymbol{k = 1, 2, \cdots, p}$, be the accumulated log-scaled values, which are those of each sample accumulated in order of increasing retention time. These accumulations are performed for each sample *individually*. An example of such an accumulation is presented in Figure 7-2B. It should be noted that although this accumulation resembles a distribution function, it contains no frequency information. The only information contained within the "distribution" function is that of normalized transformed value accumulations as time elapses. Using linear interpolation (or monotone increasing spline interpolation – see Fritsch & Carlson (1980)), we subsequently obtain the deterministic

monotone increasing function $\boldsymbol{H(t)}$ that can be used to map any retention time to an appropriate cumulative transformed value. Consider, as an example, a case of where the retention time versus the cumulative normalized and transformed values is plotted for five replicates of the same reproducibility item. If the experiment is repeatable, the overlapped cumulative values should all be similar.

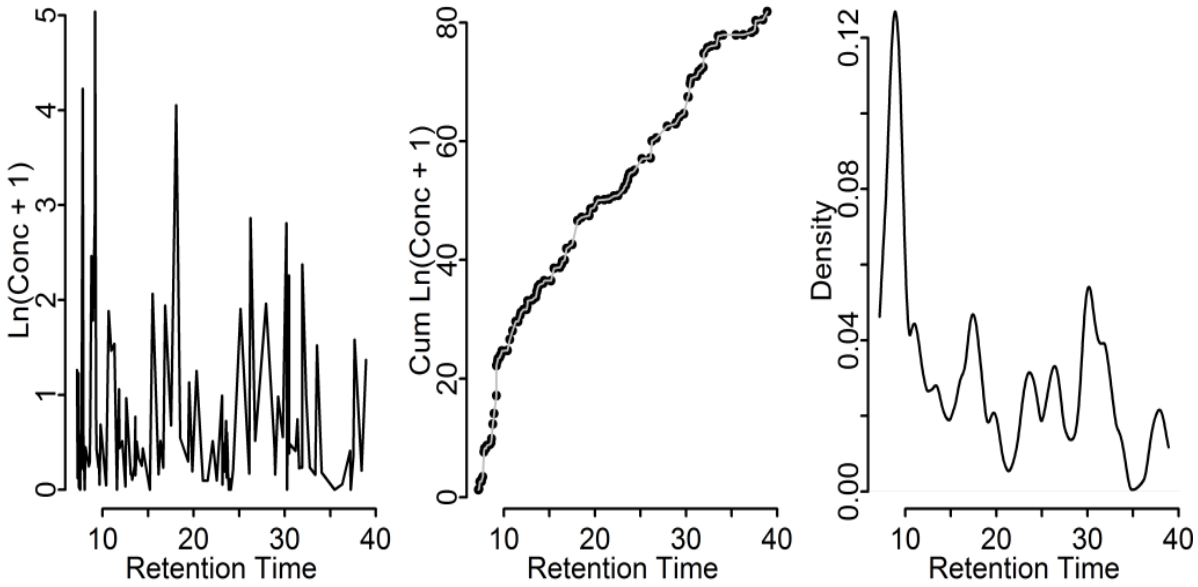


Figure 7-2: **A: Lineplot of retention time versus the normalized log-transformed data for a single sample (case). B: Accompanying scatterplot of retention times versus the cumulative normalized log-transformed data. The line $\boldsymbol{H(t)}$ represents the linear (or monotone spline) interpolation function. The arrows represent an interpolated retention time. C: Density representation of the chromatogram (i.e. the KEMREP-smoothed chromatogram).**

To find the smoothed chromatogram representation of the data (see Figure 7-2C), we set up a fine, uniformly spaced grid (of size $\boldsymbol{G}$) from the smallest normalized transformed value to the total accumulated transformed value for each sample. As an example, consider Figure 7-2B, where this grid is constructed on the y-axis. We then use the inverse interpolation function $\boldsymbol{H^{-1}(\cdot)}$ to find numerically the interpolated retention time values using the finely spaced grid as input (i.e., the inverse). This is illustrated visually (for one grid point) using arrows in Figure 7-2B. The function $\boldsymbol{H(t)}$ can therefore be considered as a model that can be used to generate data (retention times) given the functional form of the model. Note again that this function is *not* a probability model because the dependent variable contains no

information that resembles frequencies (as it is constructed from a single replicate), but instead presents information concerning the intensity of the machine activity. The result of the interpolation is that we map the bivariate data (i.e., retention time and transformed value) to univariate data (i.e., retention time), in such a way that a higher density of retention times can be expected in areas where large values are observed, and a corresponding lower density in areas with less machine activity or where lower values were recorded. A histogram can now be constructed to visualize the retention time density that resulted from the function $\boldsymbol{H^{-1}}(\cdot)$ and we opt for a kernel density estimator, which at a point $\boldsymbol{t}$ for data points $\boldsymbol{T_1, \cdots, T_G}$ is defined as

$$\hat{\boldsymbol{f}}_{\boldsymbol{h}}(\boldsymbol{t}) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{G}} \boldsymbol{K}\left((\boldsymbol{t} - \boldsymbol{T_i})/\boldsymbol{h}\right)/(\boldsymbol{Gh}),$$

where $\boldsymbol{K}(\cdot)$ is the so-called kernel function and $\boldsymbol{h}$ is the bandwidth. In the current application we use the bounded biweight kernel and select the bandwidth according to the method of Sheather & Jones (1991). The biweight kernel is defined as

$$\boldsymbol{K}(\boldsymbol{x}) = \frac{\mathbf{15}}{\mathbf{16}}\left(\mathbf{1} - \boldsymbol{x}^{\mathbf{2}}\right)^{\mathbf{2}} \mathbf{1}_{\{-\mathbf{1} \le \boldsymbol{x} \le \mathbf{1}\}},$$

where $\mathbf{1}_{\{-\mathbf{1} \le \boldsymbol{x} \le \mathbf{1}\}}$ is an indicator function; a review of kernel methods can be found in Wand & Jones (1995). This density estimate closely resembles the chromatogram but is smoothed and normalized to integrate to one. The density estimate for the example data is presented in Figure 7-2C. For a repeatable experiment an overlay of these density estimates should show little variance.

Alternatively, the density estimate can be viewed as a smoothed and normalized estimate of the derivative of the cumulative function $\boldsymbol{H}(\boldsymbol{t})$. The amount of smoothing is controlled via the kernel density estimator bandwidth parameter (i.e., the retention time window width ($\boldsymbol{h}$)) as well as the size of the grid specified from the smallest transformed value to the total accumulated transformed value. If $\boldsymbol{h}$ is decreased then the resulting density estimator will exhibit more roughness, and vice versa. Conversely, if the grid size ($\boldsymbol{G}$) is increased then the resulting density estimator will exhibit more roughness, and vice versa. Data-driven selection of the bandwidth parameter $\boldsymbol{h}$ is discussed above. The effect of various grid sizes $\boldsymbol{G}$ is displayed in Figure 7-3. The result of increasing the grid size is that a rougher density estimate is obtained that more closely resembles the original data (see Figure 7-2A). One reason for this behaviour is that the data-driven bandwidth $\boldsymbol{h}$ is inversely related to the number of data points (i.e., $\boldsymbol{h} \propto \boldsymbol{G^{-0.2}}$) used to construct the density estimate. Hence, an increase in the number of data points (obtained by increasing the grid size $\boldsymbol{G}$) will result in

smaller bandwidths $\boldsymbol{h}$, and consequently rougher density estimates. For practical purposes we propose using $\boldsymbol{G}$ = 1000.

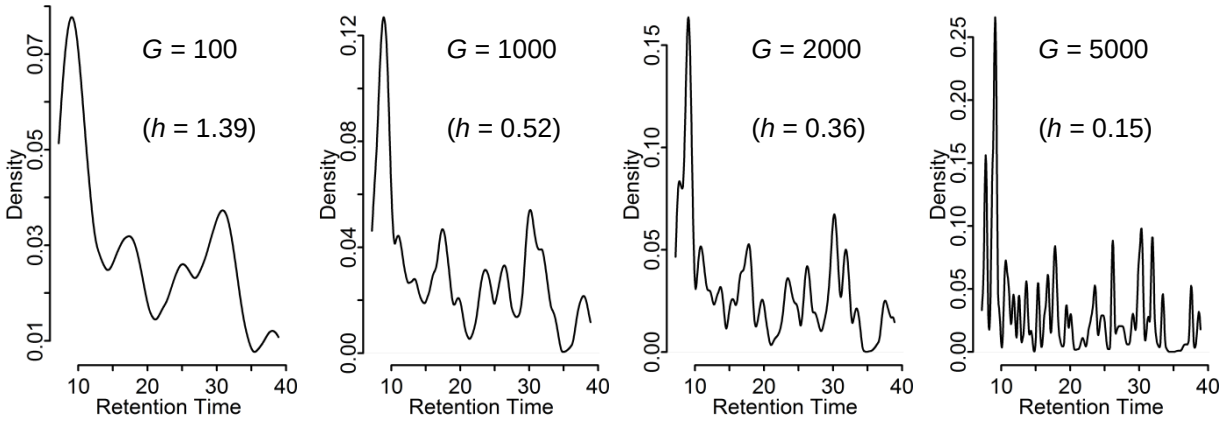


Figure 7-3: Density estimates for various grid sizes (*G* = 100, 1000, 2000 and 5000) for an individual sample (case). The bandwidth (*h*), calculated using the method of Sheather & Jones (1991), is indicated in brackets.

7.3.4.2 Quantitative bounds for smoothed chromatographic profiles

The kernel density estimates for each of the replicates can now be obtained and overlaid to assess their repeatability or reproducibility visually. However, reference bounds are required to interpret the amount of variation among the various density estimates. These reference bounds can be easily determined using a parametric bootstrap approach, where the bootstrap data are generated from a null distribution that satisfies repeatability or reproducibility. The bootstrap is a technique that can estimate population parameters and distributional properties of statistics by substituting the population mechanism used to obtain the parameter with an empirical equivalent. These estimates can be obtained analytically, but they are mostly obtained through the use of re-sampling and Monte Carlo methods carried out on a computer. Bootstrap estimates of standard errors, as well as confidence intervals, can be derived. The reader is referred to Efron & Tibshirani (1993) for a thorough discussion of the bootstrap method.

Consider a single variable $\boldsymbol{X_{i,k}}$, where $\boldsymbol{i}$ denotes the reproducibility item and $\boldsymbol{k}$ the variable index. It is reasonable to assume that this variable should be normally distributed, with mean $\boldsymbol{\mu_{i,k}}$ and variance $\boldsymbol{\sigma^2_{i,k}}$, if reproducibility item $\boldsymbol{i}$ can repeat the measurements successfully. The variance $\boldsymbol{\sigma^2_{i,k}}$ can be expressed in terms of a predefined relative standard deviation (*r*), i.e. $\boldsymbol{\sigma^2_{i,k}} = \boldsymbol{r^2\mu^2_{i,k}}$, hence

$$\boldsymbol{X_{i,k} \sim N\left(\mu_{i,k}; r^2 \mu_{i,k}^2\right)}. \quad (1)$$

In the context of reproducibility the allowable variance can be constructed by using the following null distribution (dropping the reproducibility index $\boldsymbol{i}$ for $\boldsymbol{X}$):

$$\boldsymbol{X_k \sim N\left(\mu_k; r^2 \min(\mu_{1k}^2, \mu_{2,k}^2, \cdots, \mu_{I,k}^2)\right)}. \quad (2)$$

Thus, a variable can be considered reproducible if it originates from a normal distribution with mean and variance specified according to equation (2). The two distribution restrictions outlined above define the null distributions satisfying repeatability and reproducibility, respectively. Hence, for each variable a random number from the appropriate null distribution can be generated independently. The KEMREP method can then be applied to the simulated data, where the associated retention times used are the average retention times of the observed data. This procedure is then repeated for a bootstrap replication number of $\boldsymbol{B}$, after which a 95% confidence interval for the true density satisfying repeatability or reproducibility can be constructed at each retention time, $\boldsymbol{t}$, where the densities were calculated. These confidence bounds are then superimposed on the density overlays generated by applying the KEMREP method to the observed data. If all the observed density estimates are contained within the lower and upper bounds, we conclude that the experiment was repeatable or reproducible at a predefined input relative standard deviation ($\boldsymbol{r}$), where $\boldsymbol{r}$ is specified as an input parameter to restrict the variance (see equation (1) and equation (2)). If this is not the case, the user can increase the predefined relative standard deviation $\boldsymbol{r}$ until all the density estimates are contained within the wider bounds, in which case we can interpret the suitability of the $\boldsymbol{r}$ value necessary.

It should be noted that the KEMREP method involves transformation of the data using a shifted log transformation. Since the repeatability and reproducibility null distributions are normal, it is possible (especially if the value of $\boldsymbol{r}$ is increased) to generate data for which the natural logarithm does not exist even after adding the shift constant of one. It is therefore important to restrict the data that can be generated from these null distributions to the interval $[\mathbf{0}; \infty)$ or to use a different transformation function. To restrict the data to the interval $[\mathbf{0}; \infty)$, we generate the null data according to:

$$\boldsymbol{X = \Phi^{-1}\left(CU + \Phi(-m/s)\right)s + m},$$

where $\boldsymbol{m}$ and $\boldsymbol{s}$ are the mean and standard deviation, respectively (obtained from the null distribution), $\boldsymbol{C = \int_0^\infty \phi((x-m)/s)/s\, dx}$, $\boldsymbol{\Phi}(\cdot)$ and $\boldsymbol{\phi}(\cdot)$ are the standard normal distribution and density function, respectively, and $\boldsymbol{U}$ is a random number from the standard uniform

distribution. The values of $\boldsymbol{m}$ and $\boldsymbol{s}$ can be calculated from the observed data and depend on whether we are testing for repeatability or reproducibility (see equations (1) and (2) above).

It should also be noted that the null data are independently generated from the null distributions of each variable. This assumption of independence should be improved upon by generating the null data from a multivariate normal distribution. The mean and variance vectors can be determined by the repeatability or reproducibility restrictions (see again equations (1) and (2) above). The off-diagonal entries of the covariance matrix can be estimated from the observed data. In the current application, we do not sample from a multivariate normal null distribution, since the sample size is small, which might increase the influence of individual data points when estimating the covariance structure.

7.4 Results and discussion

The first results deal with intra-individual repeatability (the same analyst) and inter-laboratory reproducibility through the use of two different GC-MS apparatuses (see left panel in Figure 7-1). The derivatized substances, extracted from the five identical aliquots of CSF, were separated on the GC columns of GC-MS-1 and GC-MS-2, respectively, producing the typical chromatograms generated in such experiments (Figure 7-4). A rapid visual qualitative assessment of the repeatability of the raw data generated by GC-MS through the overlaid chromatograms of multiple repeats (Yu *et al.* 2007) is complex to interpret (Mok and Chau 2006), making assessment of repeatability misleading as the data are not normalized.

A total of 398 features were identified from the analysis using GC-MS-1, compared with 239 from GC-MS-2. This difference could be related to the slower temperature ramp or the diverse analytical procedures performed (carry-over from previous, unrelated analyses) on GC-MS-1 compared with the dedicated use of GC-MS-2. A longer run time may reduce the risk of missing components in the experimental sample (i.e., false negatives), but it also increases the risk of false positives and multiple hits for the same features. This is evident from the fact that AMDIS identified more features in our data matrix for GC-MS-1 than for GC-MS-2, using a customized library. While AMDIS is good at peak deconvolution and feature identification, the program is limited by less accurate quantification (Lu *et al.* 2008), emphasizing the need for data reduction, manual curation and repeatability and reproducibility (data validation) assessment methods when metabolomics data are analysed.

Application of an 80% detection filter (Bijlsma *et al.* 2006) substantially reduced the number of features to 192 and 115 for GC-MS-1 and GC-MS-2, respectively. The features thus

removed are considered as artefacts and not repeatable (i.e., unreliable) and consequently were removed from any further analysis. Data correction through manual curation (Williams *et al.* 2012) was also applied by combining those features with similar retention times and MS spectra (i.e., correction of errors due to software inaccuracies); this action subsequently reduced the respective number of features to 175 and 112. The final data matrix generated consisted of the 95 features common to both apparatuses.

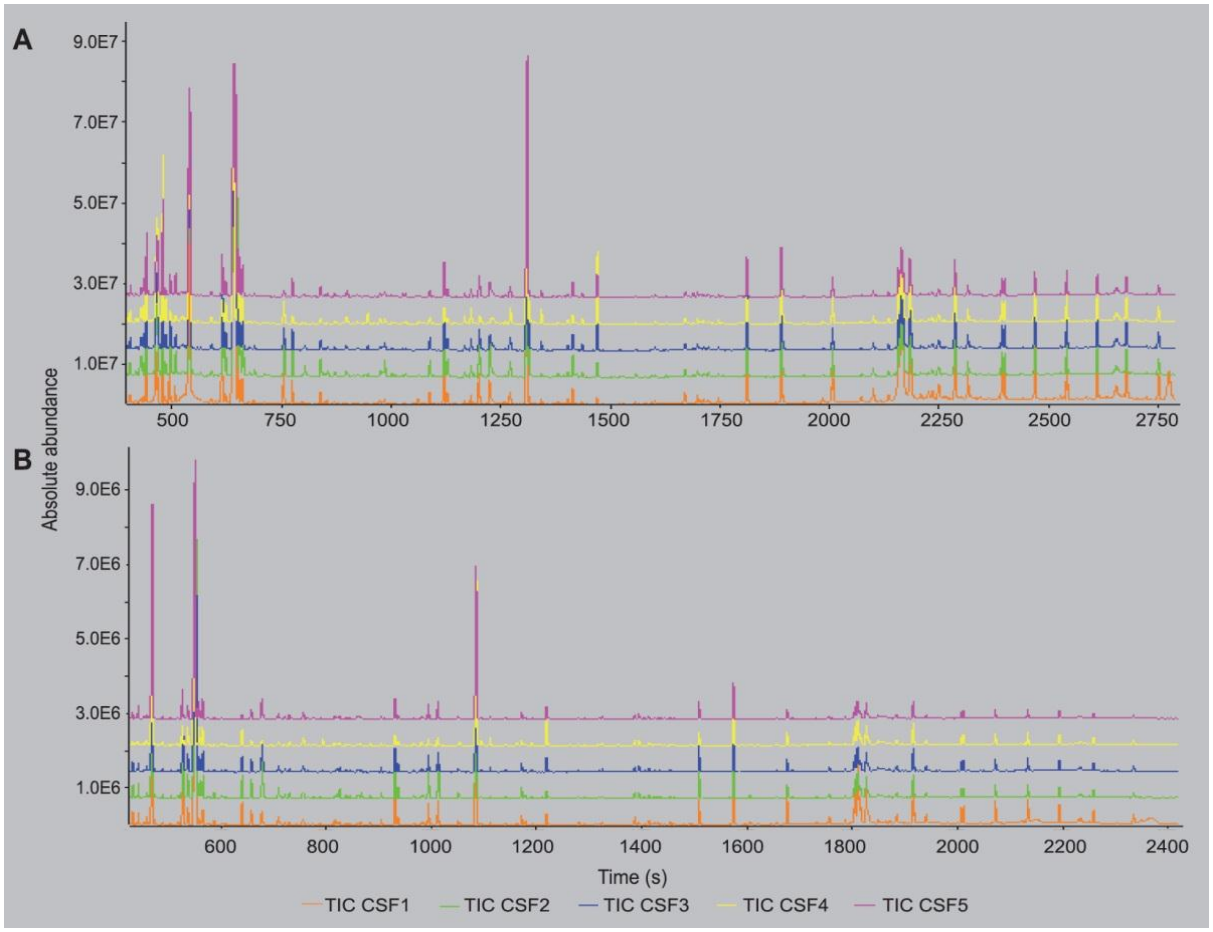


Figure 7-4: Overlay of 5 non-normalized chromatograms generated by GC-MS analysis of 5 aliquots of one TBM-infected CSF sample performed on A: GC-MS-1 (398 features) and B: GC-MS-2 (239 features).

The normalized relative concentrations and the respective retention times of the 95 common features identified from the data generated from GC-MS-1 and GC-MS-2 were accordingly used as a basis for the application of the KEMREP method to the metabolomics data from the CSF sample. The outcome of four such applications to represent the data of the chromatograms generated from the CSF sample using GC-MS-1 and GC-MS-2 is shown in Figure 7-5.

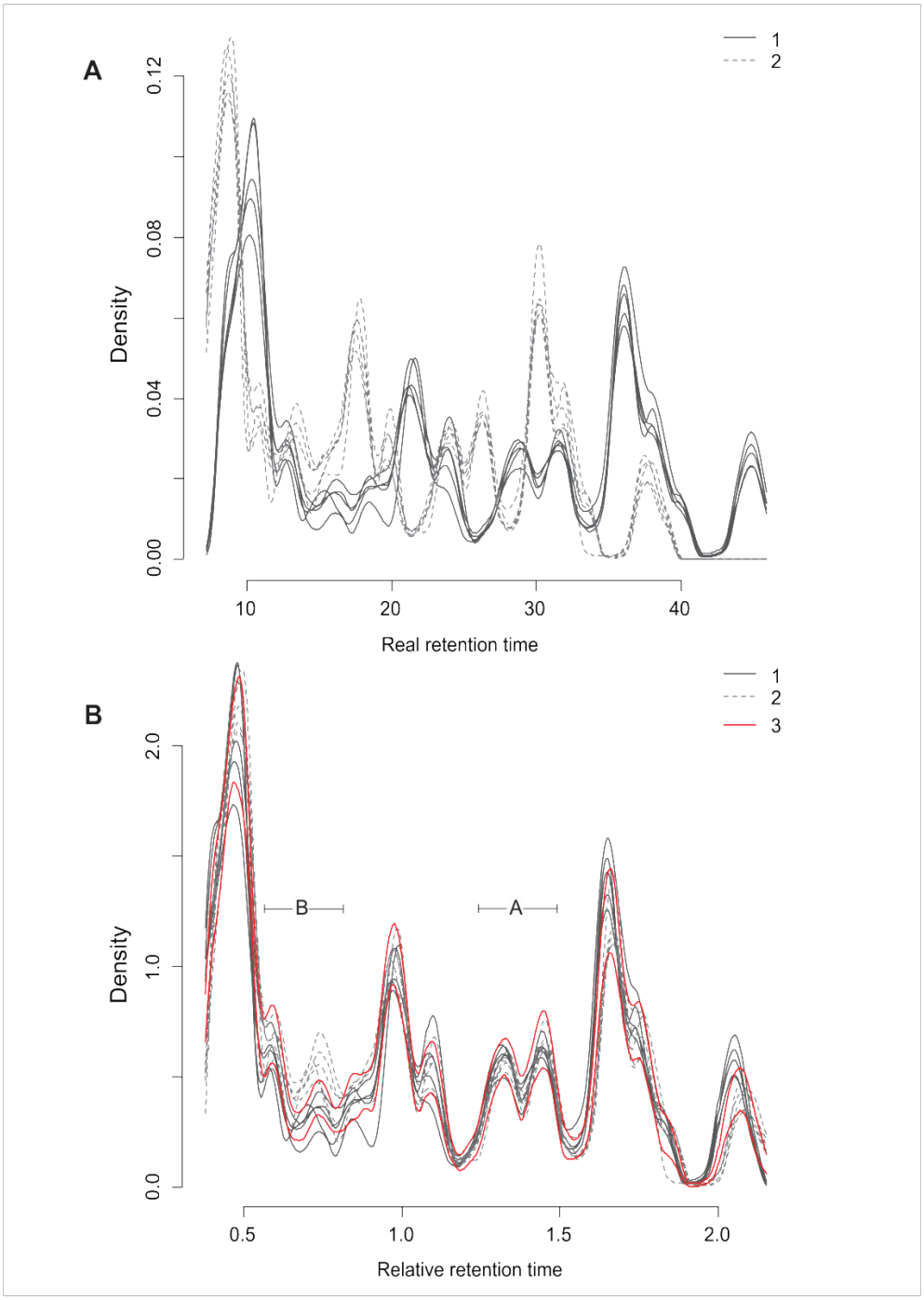


Figure 7-5: **KEMREP output of overlays of smoothed chromatograms of 95 common features from 5 repeat aliquots of 1 CSF sample.**

A: The differences in retention time between GC-MS-1 (—) and GC-MS-2 (---) are illustrated; however, these differences are removed by using relative retention times.

B: A confidence bound (—) is added to allow for a quantitative measure (A illustrates a region of good reproducibility, whereas region B highlights a region of poor-reproducibility).

Figure 7-5A presents the overlay of five chromatograms, smoothed by the KEMREP method, for GC-MS-1 (**—**) and GC-MS-2 (**---**). Visual and analytical differences in retention time between the two sets of results were immediately evident. The GC oven-temperature ramp was set at 1°C per minute more for GC-MS-2 than for GC-MS-1, thereby making the total run time for GC-MS-2 several minutes shorter. This difference gave rise to a more extended range of real retention times for GC-MS-1 (~44 minutes) than for GC-MS-2 (~40 minutes). A qualitative assessment of repeatability is made by visually inspecting the smoothed density plots individually. A well-repeated experiment is indicated by how closely the overlaid smoothed chromatograms lie on top of each other. Reproducibility is assessed by visually determining if the overall density shape is retained across density plots. Based on our definition of repeatability, an assessment of our data generated from a single CSF sample by a visual inspection of the density plots (Figure 7-5A) allows us to declare that our results are overall qualitatively repeatable for the reduced data set.

The assessments of the results, however, are all *qualitative* in nature. As indicated in the section on the development of the KEMREP model above, this method also allows for a *quantitative* assessment of repeatability and reproducibility, through the use of simulated data where the associated retention times used are the average retention times of the observed data. This is then repeated through a bootstrap procedure, after which a chosen confidence interval (e.g. 95%) for the true density satisfying repeatability or reproducibility can be constructed at each retention time point where the densities were calculated. These confidence bounds are then superimposed on the density overlays generated by applying the KEMREP method to the observed data, while the value of the confidence interval used may function as a benchmark of the improvement in the precision of the analyses. The resulting density plots, with a reproducibility null distribution specified at a CV of 30% for the 95 common variables derived from GC-MS-1 and GC-MS-2, are shown in Figure 7-5B. These results demonstrate that the overall reduced, common experimental data are reproducible from two independent GC-MS instruments. In most of the regions covering the smoothed profiles, the observed density estimates are contained within the lower and upper bounds, providing a quantitative basis to conclude that the experiment was repeatable or reproducible at the predefined input of a CV of 30%, substantiating the applicability of the KEMREP method for assessing repeatability or reproducibility for the experiment as a whole.

The region indicated as A in Figure 7-5B is an example of a chromatographic region of good reproducibility. Regions of variability exceeding the predefined bounds can also be identified, for example region B highlighted in Figure 7-5B. By utilizing the univariate measure of CV, we see that in the selected region B that CV ranges from approximately 23.6% to 87.3% and

from 37.2% to 76.3% (see SI: Table A2-6) for GC-MS-1 and GC-MS-2, respectively. It is evident, as is expected, that the metabolites of lower concentration exhibit a higher CV. Further inspection of the data reveals that the feature identified as lactic acid is of considerably higher magnitude than surrounding eluted features. Although lactic acid is known to be derivatized in an irreproducible manner (Meyer *et al.* 2011), the inconsistently repeatable levels of lactic acid in the replicates are due to inconsistent and incomplete peak integration of lactic acid by the AMDIS software. Graphically, KEMREP shows that this region is poorly-reproducible as the smoothed plots for both apparatuses fall outside the confidence bounds. Herein lies a benefit of KEMREP, as these observations are quickly and clearly detected using KEMREP, which are not so evident when dealing only with CV.

The final results presented here again deal with intra-individual repeatability (the same analysts were involved) but also with inter-individual reproducibility (different analysts) using an identical biological sample (urine), the same reagents and SOP, and the same GC-MS apparatus (GC-MS-1). The focus is thus now on the analysts in assessing repeatability and reproducibility (see right panel in Figure 7-1). The three analysts in question were new postgraduate students, interested in doing a metabolomics study for their first postgraduate degree.

The outcome of the application of the KEMREP method in these cases is shown in Figure 7-6. Figures 7-6A, 7-6B and 7-6C indicate the smoothed chromatographic profiles after the first attempt of the three analysts to produce a metabolomics data matrix by using the same SOP. Figures 7-6D, 7-6E and 7-6F represent the corresponding results of a second round of analysis by each analyst. A visual, qualitative assessment of both sets of results clearly indicates insufficient repeatability in the case of each of the analysts, as well as between them (reproducibility), during the first analysis attempted. This observation could easily and convincingly be communicated to all analysts. Thus, for the second attempt some key features of organic acid analysis had to be adhered to meticulously by the three analysts. These included (1) the use of freshly prepared reagents; (2) glassware and Hamilton syringes cleaned in exactly the same manner; (3) the analyses done in the same laboratory and using the same GC-MS equipment, and, finally, (4) generating and analysing the raw data using the same AMDIS settings. The qualitative repeatability and reproducibility profiles in all cases improved substantially at the second attempt, with analyst 1 (Figure 7-6D) producing the superior result. These results clearly substantiate the applicability of the KEMREP method to highlight the central role of the analysts in research or routine experiments (Milman 2005), and for a way to communicate effectively the qualitative status of metabolomics data generated by such experts.

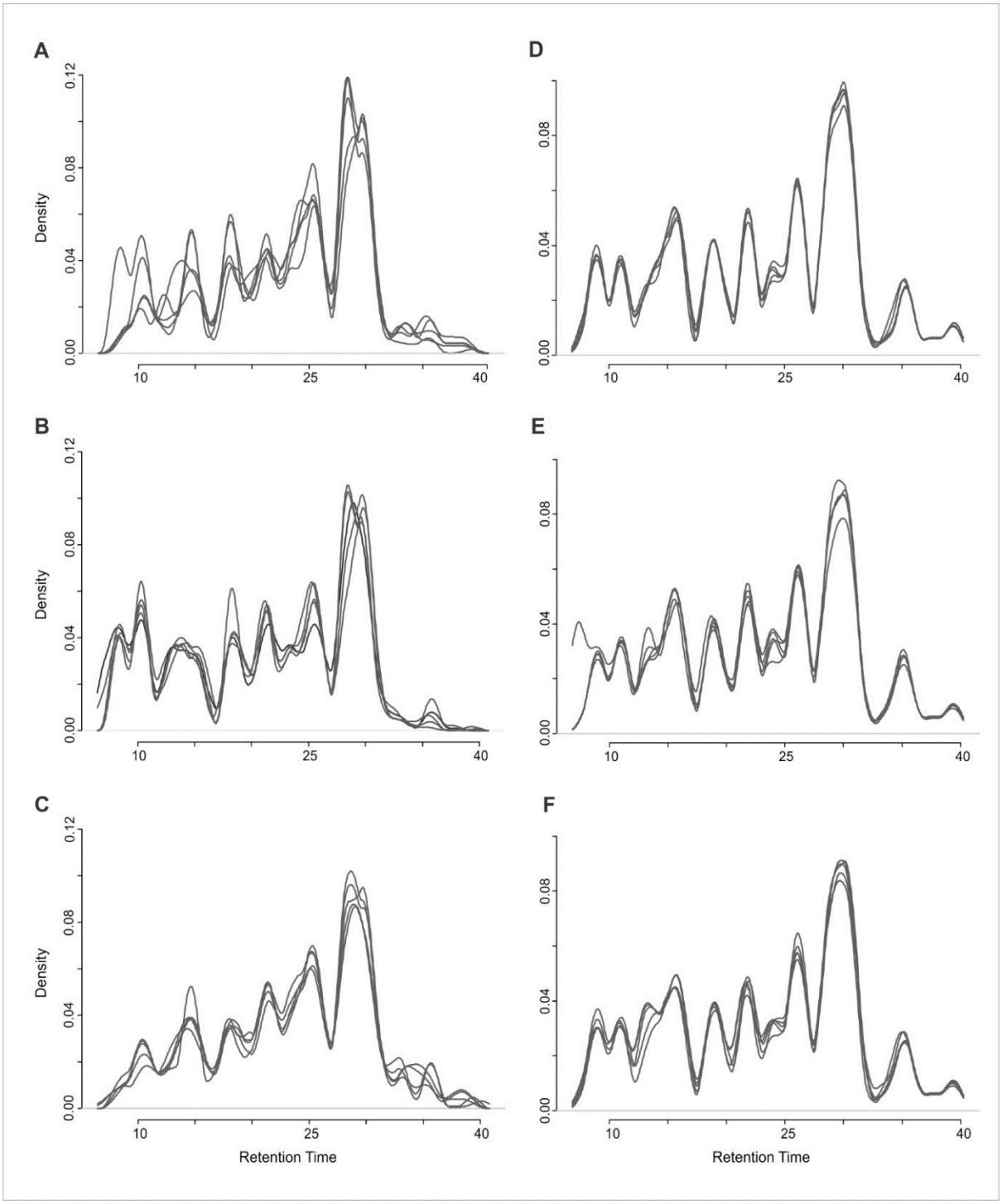


Figure 7-6: **Overlays of density plots of 5 repeat aliquots from a single urine sample. A, B & C: First attempt to generate the GS-MS data by analysts 1, 2 and 3, respectively. D, E & F: Second attempt to generate the GS-MS data by the same analysts, respectively. All analyses were conducted using the identical reagents, SOP and the same GC-MS apparatus.**

Finally, we propose that a GC-MS metabolomics data matrix generated from a single biological sample, subdivided into several identical aliquots for repeated analyses, from any other biological sample or analytical method of analysis, may be used to evaluate the KEMREP method. To illustrate this statement we include here the application of the KEMREP method to determine amino acids in 10 aliquots of a control urine sample and a freeze-dried, commercially available plasma sample spiked with amino acids, including a quantitative analysis (SI). The confidence bounds, as well as the CV profiles of these analyses, give a further indication of the precision of the analyses done by the analyst and the applicability of the KEMREP method.

7.5 Conclusions

KEMREP, as developed and described here, has been shown to be a novel statistical method for an easy and direct qualitative assessment of the repeatability and reproducibility of metabolomics data matrices, based on its application to real metabolomics data derived from different biofluids. The method can also be extended by the imposition of confidence bounds, which provides lower and upper limits that indicate quantitatively whether the experiment was repeatable or reproducible at a predefined input coefficient of variation (CV). Its application indicates several advantages in the evaluation of complex metabolomics data, emphasizing the potential wide application of KEMREP in metabolomics investigations. This paper thus highlights the ability of KEMREP to supplement existing methods in assessing the reliability of metabolomics data generated by GC-MS. We are confident that the KEMREP method has the potential to be extended as: (1) a benchmark for determining the quality of performance of analysts, (2) a means of monitoring experimental protocols, and (3) of evaluating the analytical robustness of experimental design in metabolomics.

Conflict of interest: The authors confirm that this article content has no conflicts of interest.

Acknowledgements: S. Mason is a recipient of a Vrije Universiteit-NRF Desmond Tutu Doctoral Scholarship from the National Research Foundation (NRF) of South Africa. Funding for the project from the Technological Innovation Agency (TIA) of the South African Department of Science and Technology and from North-West University is likewise acknowledged.

Supplementary information: see Annexure 2.

CHAPTER 8 A PUTATIVE URINARY BIOSIGNATURE FOR DIAGNOSIS AND FOLLOW-UP OF TUBERCULOUS MENINGITIS IN CHILDREN: OUTCOME OF A METABOLOMICS STUDY DISCLOSING HOST–PATHOGEN RESPONSES

Shayne Mason[1], A. Marceline van Furth[2], Regan Solomons[3], Ron A. Wevers[4], Mari van Reenen[1] and Carolus J. Reinecke[1]

[1]Centre for Human Metabolomics, Faculty of Natural Sciences, Private Bag X6001, North-West University (Potchefstroom Campus), Potchefstroom 2531, South Africa,
[2]Department of Paediatric Infectious Diseases–Immunology and Rheumatology, Vrije Universiteit Medical Centre, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands,
[3]Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 19063, Tygerberg 7505, South Africa,
[4]Translational Metabolic Laboratory, Department Laboratory Medicine, Radboud University Medical Centre, Geert Grooteplein 10, 6525 GA, Nijmegen, The Netherlands.

Keywords: tuberculous meningitis (TBM); urine; gas chromatography–mass spectrometry (GC-MS); metabolomics.

Submitted: *Metabolomics,* 2016

8.1 Abstract

Tuberculous meningitis (TBM) is a severe manifestation of tuberculosis, presenting with high morbidity and mortality in children. Existing diagnostic methods for TBM are invasive and time-consuming, often causing complications due to delayed treatment. The need for highly sensitive and selective diagnosis thus remains high on the TBM agenda. Our aim was to investigate metabolomics data from children to identify metabolites with characteristics to define a potential diagnostic biosignature for children with TBM through a non-invasive means. Urine samples were selected for this study, using three paediatric groups: patients with confirmed TBM (12 cases), patients clinically suspected with meningitis but later confirmed to be negative (18 cases), and an age-matched control group (30 cases). Metabolomics data were generated through gas chromatography–mass spectrometry analysis and revealed six groups of metabolites that characterized TBM — three groups reflected the host response and three groups apparently related to a host-microbial response. We proposed a global metabolite profile reflecting the potential diagnostic host-microbial metabolites and identified methylcitric, 2-ketoglutaric, quinolinic and 4-hydroxyhippuric acids — designated as SUM-4 — as a Mtb–host biosignature. SUM-4 proved to have diagnostic ability and prognostic insight into our TBM patients. This study is the first to illustrate holistically the metabolic complexity of TBM-confirmed cases and acts as a proof-of-concept that a biosignature of urinary metabolites can be used in the diagnosis and assessment of prognosis of TBM patients that could be developed and validated through future studies to generate a medical algorithm with potential for diagnosis and monitoring of treatment strategies.

8.2 Introduction

Tuberculosis (TB), the disease caused by the *Mycobacterium tuberculosis* (Mtb) bacterium, remains one of the major contemporary pandemics, with tuberculous meningitis (TBM) being one of its most severe manifestations. Infants and children are particularly affected, with a case-fatality rate being reported to be as much as 50 % and with a high potential for severe disability in survivors (Wolzak *et al.* 2012). A definitive diagnosis of TBM can be made only in patients with signs and symptoms of central nervous system (CNS) disease through analysis of cerebrospinal fluid (CSF) and identification of Mtb by molecular and/or microbial methods. This combination of analytical criteria is the gold standard for diagnosis, but requires methods that are invasive and time-consuming to apply and complications often arise in delayed treatment, especially in cases of multidrug-resistance or co-infections. Thus, the

need for highly sensitive and selective diagnosis of TBM, that uses child-friendly sampling techniques and can be applied in resource-limited settings, remains high on the TBM agenda (Graham *et al.* 2014).

The ability to diagnose infectious diseases using urinary antigens proved successful in the cases of malaria, influenza and bacterial meningitis (Marcos *et al.* 2001). In diagnosing tuberculosis using urinary antigens, lipoarabinomannan (LAM), a mycobacterium-specific lipopolysaccharide component of the bacillus's cell wall, became an important focus. A commercial ELISA test to detect LAM in urine (MTB ELISA Test®, Chemogen, Portland, OR, USA) is now available as Clearview® TB ELISA, Inverness Medical Innovations, Inc., Waltham, MA, USA. Initial assessments were encouraging but more recent evaluations showed that the current commercial urine LAM ELISA test is not useful for independent diagnosis of pulmonary tuberculosis (Reither *et al.* 2009). Evaluation of the diagnostic value of the Clearview® TB ELISA test, using urine samples from a cohort of paediatric patients suspected of TBM, likewise proved to be of little value (Blok *et al.* 2014). Thus, the lack of applicable diagnostic tests is a major hindrance to clinicians guiding TBM treatment. Meanwhile, the recommended treatment regime for TB consists of an intensive initial phase using several drugs, continued by a period in which mostly only two drugs are prescribed (Principi & Esposito 2012), and one clinical institution, for example, advocates the use of four drugs over six months (van Toorn *et al.* 2014). These relatively aggressive approaches are generally followed for children suspected of having TBM, and emphasize the need for a more efficient means of diagnosing, treating and monitoring the progress of the disease in these cases.

Metabolomics investigations of CSF (Coen *et al.* 2005), serum (Zhou *et al.* 2013) and sputum (du Preez & Loots 2013) revealed alterations in several metabolic pathways which seem to be implicated in the pathophysiology of TB and meningitis. In a unique non-metabolomics study of CSF, cyclopropane was proposed as a biomarker for TBM (Subramanian *et al.* 2005). Using proton magnetic resonance (^{1}H NMR) spectroscopy, it was shown, through a metabolomics study of CSF, that decreased glucose and highly elevated lactate are defining metabolites which distinguished TBM from controls. Those compounds and a further 14 significant energy-associated metabolites led to the postulation of an “astrocyte–microglia–lactate–shuttle” to be operate in TBM (Mason *et al.* 2015). However, none of the important distinguishing metabolites from these metabolomics studies seemed to possess sufficient specificity and selectivity, or to have been validated to qualify as potential biomarkers for TBM. The search for biomarkers which satisfy these criteria therefore continues.

By contrast to their application to infectious diseases, an extensive array of clinical chemical tests of high specificity and sensitivity based on mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy is well established for the screening of newborns (Wilcken *et al.* 2003) and diagnosis of monogenic disorders (Engelke *et al.* 2007; Scriver 2012). Urine proved to be the biofluid of choice for most of the diagnostic analyses. Global metabolic profiling expanded the scope for identifying and refining biomarkers of inherited diseases (Wikoff *et al.* 2007; Reinecke *et al.* 2012; Dercksen *et al.* 2013), and for infectious conditions with complex manifestations (Wikoff *et al.* 2008; Mason *et al.* 2015); most recently, a gas chromatography–mass spectrometry (GC-MS) metabolome analysis of urine (Das *et al.* 2015) demonstrated differences between clinically confirmed active pulmonary TB patients (n = 21), non-TB subjects (n = 21) and healthy controls (n = 11) in terms of their metabolic profiles.

Against this background, we formulated the following working hypothesis: the metabolism of patients with an infectious disease constitutes a complex dynamic system of fluctuating responses of dysregulation which induce variability in affected metabolites, subsequently reflected in patterns of correlations which are discernible in the patients' biofluids, identifiable through metabolomics technology.

To develop these ideas more specifically, we first assumed that the metabolite profile of meningitis patients differed between those suffering from Mtb infection in the CNS and those from non-TB related meningitis. Second, it was anticipated that metabolite profiles reflect the innate defence response of the host as well as those associated with the Mtb pathogen. Third, we postulated that the TBM-related metabolite profile was not constant but fluctuated in relation to the clinical condition of the patient. Further, variations in the metabolite profiles will be manifest in their concentrations in urine and create patterns of correlations, thereby demonstrating their diagnostic potential.

To explore these ideas experimentally, we undertook a metabolomics investigation of urine samples from confirmed TBM patients (n = 12) and non-TB meningitis patients as sick controls (n = 18; abbreviated as non-TBM), using matched infants and children (n = 30) as healthy controls. We show that the metabolomics approach revealed urinary metabolites that distinguished paediatric TBM from controls as well as from non-TBM patients. We subsequently validated a biosignature for the patient group, demonstrating high potential clinical value for TBM diagnosis and monitoring of treatment. Insights into TBM risks were assessed by comparing TBM patients, who eventually became healthy with no clinical consequences, with those who developed severe disabilities, following the disease. This

investigation justifies future extended large-scale studies to further validate the clinical utility of the putative biosignature revealed in this study.

8.3 Materials and methods

8.3.1 The study population and sampling

The study population consisted of infants and children (<13 years) from the Western Cape province of South Africa. In this particular area TB is endemic and recent poor weight gain or weight loss is helpful in the early diagnosis of childhood TBM (Donald *et al.* 1985). In a study on childhood TBM covering a period of twenty years, 90 % of patients from this area had either sub-acute weight loss or poor weight gain prior to clinical presentation (van Well *et al.* 2009). The patients came from local clinics, where they received broad spectrum antibiotics, and were referred to Tygerberg Hospital, Cape Town, on suspicion of meningitis. According to the medical records, most childhood TBM admissions to the hospital were in a malnourished, sometimes ketotic, state and routine management in the paediatric unit also required correction of the nutritional deficits of the children.

Urine samples from patients in this investigation were classified as confirmed TBM (group 1: TBM; n = 12), based on clinical and CSF characteristics, as summarized in the supplementary information — Annexure 3 [(SI) Table A3-1]. The samples were collected at the time of admission to hospital, when most of the patients were regarded as moderately to critically ill. The first sample was taken prior to the institution of a high-intensity TBM treatment (van Toorn *et al.* 2014) and a second urine sample was collected after 2 weeks of intensive hospitalized treatment for TBM. The long-term (6 month) recovery outcome of all 12 TBM patients was recorded; 6 appeared to have normal outcomes and 6 presented with neurological disabilities. According to clinical records, the outcome of normal development was linked to early diagnosis for TBM whereas the subgroup with disabilities was diagnosed at a more advanced stage of the disease. Controls included in the study were urine samples from paediatric patients suspected for meningitis, but later confirmed to be meningitis negative (group 2: non-TBM; n = 21) and age-matched infants and children (group 3: controls; n = 31), suspected for a metabolic disorder, but later proven normal.

For the assessment of the proposed biosignature for TBM, we used one further group of experimental subjects: middle-aged females (n = 17) suffering from fibromyalgia syndrome (FMS), a disorder quite unrelated to TBM. Patients were selected by specialist physicians

according to the American College of Rheumatology 1990 criteria for diagnosing FMS, including application of pressure to the tender points defined for FMS diagnosis.

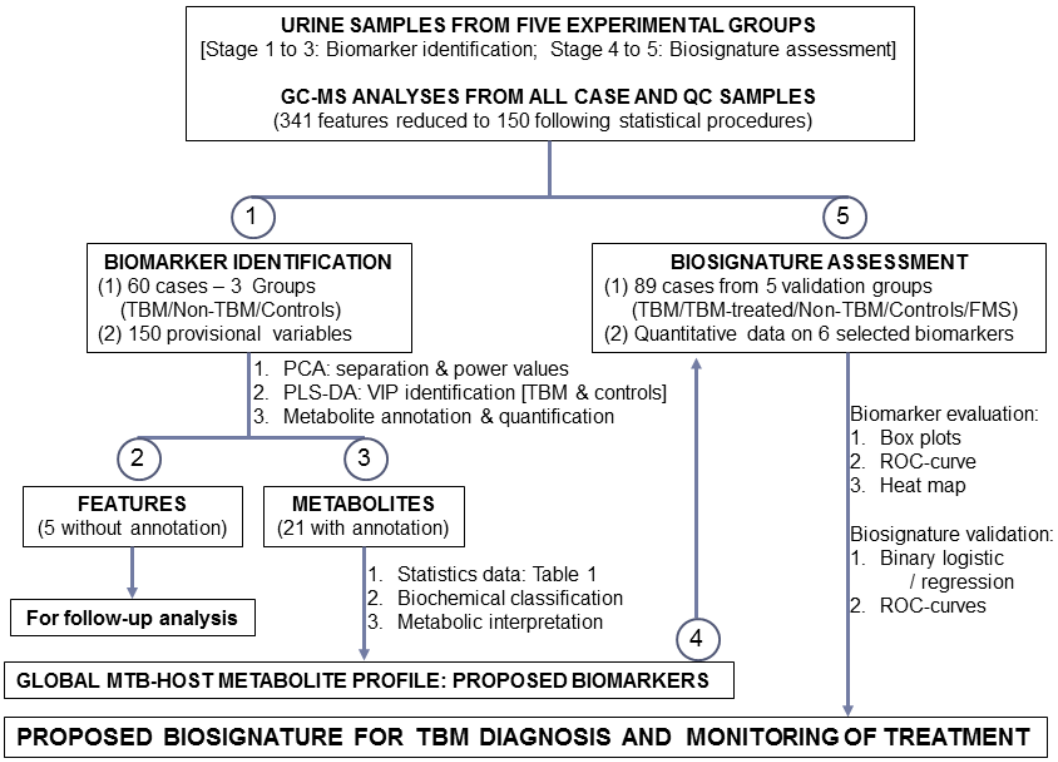


Figure 8-1: Flow diagram of experimental procedures followed to define the biosignature.

8.3.2 Quality control and analytical procedures

Organic acid and amino acid profiles of urine collected from our sample groups were analysed by GC-MS (see SI — Annexure 3 — for detailed descriptions) which proved to be both qualitatively and quantitatively repeatable and reproducible, based on the methodology of Mason *et al.* 2014. Quality assurance was confirmed by including a pooled quality control (QC) sample per analysis. Each run on the GC-MS consisted of one blank (B), followed by 3 QC samples whereas the remainder of the run was made up of blocks of approximately 9 randomly selected samples (S) with a QC sample between blocks. Each run ended with a repeat of 3 QCs, providing the following batch of a typical GC-MS run:

B|3QC|-9S-|QC|-9S-|QC|-9S-|QC|-9S-|QC|-9S-|QC|-9S-|QC|-9S-|3QC

Analysis of the QC samples showed no discernible drift or trend, and consequently no batch corrections were needed. Coefficient of variation (CV) values for QCs samples were examined after shifted log transformation; however, no data reduction was necessary.

8.3.3 Statistical analyses

Statistical analysis was performed on the organic acids data matrix, which consisted of 341 original features recorded for 64 samples. All variables in each group that did not contain values in at least 50% of the cases (i.e. more than 50 % zero values) were removed from the original data matrix, a process known as zero filtering. Variables in the reduced data matrices were followed by manual curation and classification based on the Human Metabolome Database (Wishart *et al.* 2012), with any non-biological variables (e.g. contaminants, medication and derivatization artefacts) being excluded from further analysis, leaving a biologically heterogeneous group of 150 features. This selection compared well with the expanded urine metabolome of 179 metabolites (85 quantified), identified through GC-MS (Bouatra *et al.* 2013). Next, the data were scaled using a shifted log transformation with shift parameter equal to one. The Hotelling's T^2 statistic from a principal component analysis (PCA) model was used to detect outliers, after which outliers were excluded from further analysis. Univariate statistics, specifically the Mann-Whitney (MW) test (*p* values and effect sizes) and fold change (FC) ratios, were produced for the untransformed data. Pairwise comparisons between groups identified features which differed for the three experimental groups. Zero replacement was performed for the untransformed data from the tail of a fitted beta distribution not exceeding the minimum observed value for each feature. After zero replacement, the data were again scaled using a shifted log transformation (with shift parameter equal to one) and mean centred. Unsupervised (PCA) and supervised (PLS-DA) models were fitted to the zero-replaced, transformed and centred data to identify combinations of features which differentiated between the groups (variables of influence on projection (VIP) > 0.9 and/or power values > 0.5). The next section describe the separations found between the groups and lists the features responsible for the separations.

The following statistical packages were used in the analysis of the metabolomics data:

i. MATLAB with Statistics and PLS Toolbox Release (2012). The MathWorks, Inc., Natick, MA, USA; together with notBoxPlot.m developed by Rob Campbell (http://www.mathworks.com/matlabcentral/fileexchange/26508-raacampbell13-notboxplot).

ii. SAS Institute Inc. (2015). The SAS System for Windows Release 9.3 TS Level 1M0, Copyright© by SAS Institute Inc., Cary, NC, USA.

SPSS Inc. (2015). IBM SPSS Statistics Version 22, Release 22.0.0, Copyright© IBM Corporation and its licensors (http://www-01.ibm.com/software/analytics/spss/).

8.4 Results and discussion

8.4.1 Data generation and analysis

Figure 8-1 shows the flow diagram of data generation and analysis towards biomarker identification, biosignature definition and assessment. Data analyses and identification of important metabolites were done according to standard statistical procedures used for metabolomics data as described above. The outcomes of these analyses are presented in detail in the supplementary information — Annexure 3, and the key result from these analyses is the list of important metabolites is shown in Table 8-2.

The outcome of the initial GC-MS analysis of samples from the TBM, non-TBM and controls groups yielded 341 provisionally annotated variables. Variables were reduced (Figure 8-1: stage 1) by manual curation, zero reduction (50 % rule) and application of a biological filter by removing contaminants — variables derived from medication and artefacts formed during the derivatization process. From this, 150 variables were available for modelling the metabolite profile with respect to TBM.

Dendrograms from Euclidean and Ward cluster analysis indicated complete separation between TBM and control groups, near-complete separation between TBM and non-TBM groups and no class separation between non-TBM and control groups (see Figures A3-1 (a), (b) and (c)). A heat map of the variables for the TBM patients and controls clearly distinguished these experimental groups and substantiated these findings (Figure A3-2).

Outlier detection was performed for all three experimental groups using a 95 % confidence region in a Hotelling's T^2 analysis (Figure A3-3). No outliers were observed in the TBM group, but two were detected in both the non-TBM and control groups. The four cases were regarded as true outliers as they occurred in both the Hotelling's test as well as in the score plot PCA. All four outliers were therefore removed, resulting in a data matrix of 60 experimental cases (TBM: 12; non-TBM: 19; controls: 29) and 150 variables.

Unsupervised PCA was performed to evaluate the overall variability of the three experimental groups (Figures A3-4(a-c)). Complete natural separation was observed between the TBM patients and controls (Figure A3-4a), almost complete separation between the TBM and non-TBM patients (Figure A3-4b), but no separation between non-TBM and control groups (Figure A3-4c). The various metabolites responsible for the differences observed through the PCA for the TBM and control groups were highlighted through loading plots with the respective power values of selected metabolites included in Table 8-1.

Subsequently, a supervised PLS-DA was performed to maximize the discrimination between the sample groups and to identify the metabolites which discriminated between the groups. The PLS-DA analysis largely matched the unsupervised PCA observations and is presented in Figure A3-4d (TBM patients and controls), Figure A3-4e (TBM and non-TBM patients) and Figure A3-4f (non-TBM patients and controls), again indicating a clear differentiation between the patient and two control groups. The VIP parameters of the PLS-DA model were used to identify metabolites which discriminated best between the TBM and the control groups (Figure A3-4d). A total of 26 metabolites were identified as being the most important distinguishing variables on the basis of univariate (FC: TBM/control > 1.0 with a significant difference at p value < 0.05 based on Wilcoxon–Mann-Whitney test) and multivariate analyses (PLS-DA VIP > 0.9). The one variable that did not comply with all the selection criteria was 3-hydroxybutyric acid (p = 0.0835) but is nevertheless included in the list, given its very high FC value of 34.02.

No unambiguous annotation to the metabolic status of 5 of the 26 variables could be made (Figure 8-1: stage 2). From tentative structural information, we speculated that these variables might originate from the gut microbiome. Future detailed analytical-chemical studies are required, however, as one or more of these non-annotated variables may be important to improve our proposed TBM biosignature. Table 8-1 lists the remaining 21 variables (including 3-hydroxybutyric acid) which could be annotated unequivocally as metabolites (Figure 8-1: stage 3), ranked according to the VIP values. Also included in Table 8-1 are: (1) 5 metabolites that did not by definition (PLS-DA VIP > 0.9) contribute to the separation seen in the profiles of the TBM patients and controls, but with FC (TBM/control) > 1.0 and p < 0.05; (2) the mean concentrations of the selected variables for the non-TBM controls. The niche associated with the 21 metabolites was classified in two major categories: first, indicators of dysfunctional host metabolism and second, indicators of host–microbe response in TBM. We subsequently identified biomarkers for this disease from these two categories, based on pathophysiological interpretation of TBM and through inductive reasoning.

Table 8-1: Metabolites that contributed to the separation between the TBM and the control groups, along with their respective statistical significance, ranked according to VIP values. Concentrations of all variables are µmol/mmol creatinine. *The data on two amino acids ([tryptophan] and [glutamic acid]) are included in the table, but were obviously not incorporated in the data matrix for the organic acid statistical analyses. For reference purposes we also report the mean values for the organic acids of the non-TBM.

Annotated metabolite	PLS-DA VIP	M-W *p*-value	FC	PCA power	Mean TBM	Mean Control	SD TBM	SD Control	Mean Non-TBM
4-Hydroxyhippuric acid	14.66	<0.0001	14.28	0.54	72.31	5.06	55.17	9.2	6.5
Quinolinic acid	7.5	<0.0001	16.43	0.57	18.48	1.12	17.77	1.89	1.21
Phenylacetylglutamine	7.01	<0.0001	51.54	0.51	17.37	0.34	20.73	1.43	0.55
Pyroglutamic acid	4.92	<0.0001	6.82	0.51	8.72	1.28	2.65	2.13	1.22
Hippuric acid	4.7	0.0002	3.01	0.27	235.38	78.33	124.99	77.84	65.24
2-Ketoglutaric acid	4.53	<0.0001	17.86	0.30	22.55	1.26	48.06	2.82	1.1
4-Cresol	4.5	0.0018	5.75	0.27	43.94	7.64	42.14	12.89	19.31
Glycerol	2.85	0.0028	11.4	0.21	18.72	1.64	23.36	1.29	10.12
3-Hydroxyisovaleric acid	2.5	<0.0001	3.73	0.34	48.2	12.9	44.46	9.38	17.57
Uracil	2.29	<0.0001	6.07	0.28	8.56	1.41	16.57	3.27	1.02
2-Hydroxybutyric acid	2.26	<0.0001	2.27	0.44	3.64	1.61	2.13	6.37	1.61
3-Hydroxyburyric acid	2.39	0.0835	34.01	0.16	35.96	1.06	72.47	3.69	48.5
Isocitric acid	2.04	0.0189	1.29	0.17	12.88	10.02	13.74	17.24	15.25
4-Hydroxyphenyllactic acid	1.62	0.0011	2.73	0.39	9.88	3.62	8.64	4.77	2.57
Succinic acid	1.3	0.0077	-2.67	0.32	23.47	62.58	16.02	52.1	45.49
Indole-3-acetic acid	1.22	0.0028	1.72	0.12	7.98	4.64	10.01	17.97	2.12
Vanillylmandelic acid	1.11	0.0006	1.7	0.34	11.03	6.49	3.15	6.77	4.68
4-Hydroxyphenylpyruvic acid	0.99	<0.0001	>100	0.39	1.92	0.02	1.97	0.09	0.13
Methylcitric acid	0.94	<0.0001	9.97	0.53	1.59	0.16	0.74	0.19	0.22
2-Methyl-3-hydroxybutyric acid	0.92	0.0001	2.66	0.36	4.57	1.71	2.44	1.57	1.92
Monostearylglycerol	0.92	<0.0001	4.24	0.42	2.14	0.51	1.88	1.48	0.06
Other:									
Acetoacetic acid	0.25	0.0274	>100	0.11	0.88	0	1.96	0.02	16.72
Kynurinic acid	0.08	0.0063	>100	0.17	0.39	0	0.5	0	0
5-Hydroxyindoleacetic acid	0.07	0.0012	18.58	0.23	0.35	0.02	0.38	0.07	0.1
[tryptophan]*	1.37	0.0030	2.02	0.45	33.04	16.33	20.27	10.47	36.80
[glutamic acid]*	0.79	0.0238	1.48	0.05	19.22	12.93	9.10	6.38	32.40

8.4.2 Indicators of dysfunctional host metabolism in TBM

By comparing the clinical data on the TBM patients with the metabolite changes observed in the control group, we could identify 12 metabolites as being those most likely generated by dysfunctional host metabolism in response to Mtb infection. It should be noted that lactic acid is not included in the list summarized in Table 8-1, although it has been shown to be significantly increased in the CSF of TBM patients (Mason *et al.* 2015). The urinary lactic acid values did, however, correlate significantly (Spearman's correlation coefficient r = 0.42, p < 0.0001) with 2-hydroxybutyric acid, a known ketosis marker. Lactic acid, however, cannot be classified as being solely from the host, given known and anticipated interrelationships between host and Mtb metabolism.

We classified the 12 host metabolites into three categories to enable biomarker identification, being those relating to (i) lipolysis and ketosis (5 metabolites), (ii) perturbed energy metabolism (5), and (iii) liver damage (2). An example of a metabolite from each of these three groups is shown as a box-like plot in Figures 8-2(a-c). These plots provide comprehensive graphical representations of quantitative data. The raw data are scattered over a 95 % confidence interval (red) and 1 standard deviation of the mean (blue). The statistical significance of differences (i.e. *p*-values) is indicated for every group relatively to the other, shown in the upper section of each set of graphs.

8.4.2.1 Lipolysis and ketosis

In the clinical records of the TBM patients several were noted as critically ill on admission to hospital, presenting with ketosis and other clinical symptoms due to malnourishment, vomiting and diarrhoea. Ketosis is the process by which the body, suffering low available carbohydrate reserves, converts stored fat into energy through lipolysis, here indicated by the increased concentration of glycerol and monostearylglycerol (Table 8-1) in samples from the TBM patients relative to controls. Prominent indicators for ketosis included in Table 8-1 are: 2-hydroxybutyric acid (Kamerling *et al.* 1978), 3-hydroxybutyric acid (Figure 8-2a) and 2-methyl-3-hydroxybutyric acid (Landaas 1975). In addition, acetoacetic acid (VIP = 0.30; p = 0.027; FC = 214.6), an accepted clinical indicator of ketosis, was notably present in urine samples of TBM patients (Table 8-1: Other).

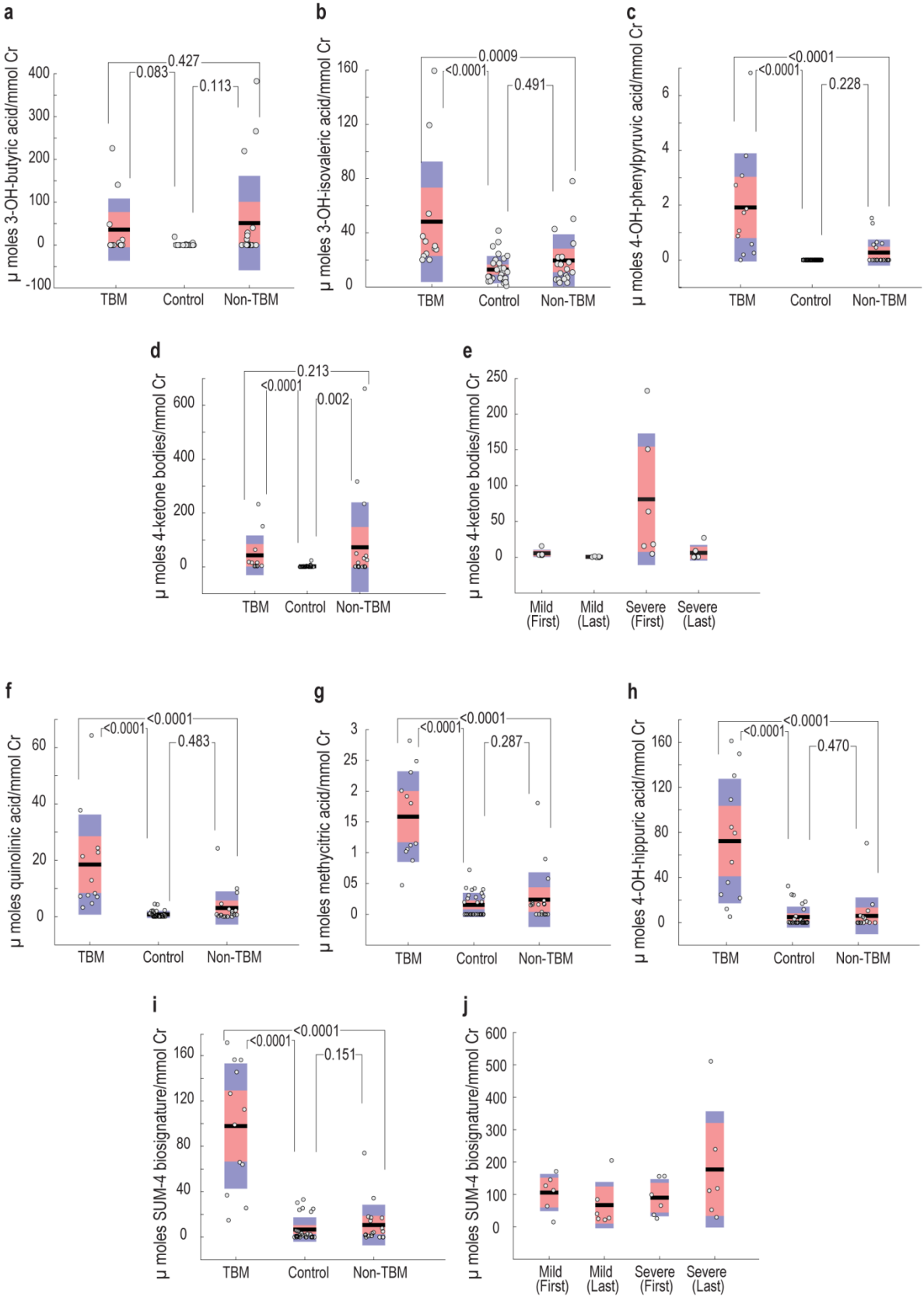

a
μ moles 3-OH-butyric acid/mmol Cr
400
300
200
100
0
-100
0.427
0.083
0.113
TBM
Control
Non-TBM
b
μ moles 3-OH-isovaleric acid/mmol Cr
160
120
80
40
0
0.0009
<0.0001
0.491
TBM
Control
Non-TBM
c
μ moles 4-OH-phenylpyruvic acid/mmol Cr
6
4
2
0
<0.0001
<0.0001
0.228
TBM
Control
Non-TBM
d
μ moles 4-ketone bodies/mmol Cr
600
400
200
0
0.213
<0.0001
0.002
TBM
Control
Non-TBM
e
μ moles 4-ketone bodies/mmol Cr
250
200
150
100
50
0
Mild (First)
Mild (Last)
Severe (First)
Severe (Last)
f
μ moles quinolinic acid/mmol Cr
60
40
20
0
<0.0001
<0.0001
0.483
TBM
Control
Non-TBM
g
μ moles methycitric acid/mmol Cr
3
25
2
15
1
05
0
<0.0001
<0.0001
0.287
TBM
Control
Non-TBM
h
μ moles 4-OH-hippuric acid/mmol Cr
160
120
80
40
0
<0.0001
<0.0001
0.470
TBM
Control
Non-TBM
i
μ moles SUM-4 biosignature/mmol Cr
160
120
80
40
0
<0.0001
<0.0001
0.151
TBM
Control
Non-TBM
j
μ moles SUM-4 biosignature/mmol Cr
600
500
400
300
200
100
0
Mild (First)
Mild (Last)
Severe (First)
Severe (Last)

Figure 8-2: Plots showing important metabolites reflecting perturbed host metabolism seen in TBM and on the state of ketosis and on the Mtb-host response in different groups of patients and controls.

(a) to (e): Ketosis expressed as a biosignature comprising of the sum total of the concentrations of four host ketosis biomarkers — 2-hydroxybutyric, 2-hydroxyvaleric, 3-hydroxybutyric and acetoacetic acids. Results on the important biomarkers for TBM patients are shown relative to the controls and non-TBM cases for 3-hydroxybutyric acid (a), an indicator for ketosis; 3-hydroxyisovaleric acid (b), an indicator of perturbed energy metabolism; 4-hydroxyphenylpyruvic acid (c), an indicator of perturbed liver function. The ketosis biosignature is shown in (d) for untreated TBM patients relative to the controls and non-TBM cases, and in (e) for the two TBM-subgroups (with mild and severe clinical conditions, respectively, affecting their long-term outcomes) on admission to hospital (TBM–first) and after two weeks of intensive treatment (TBM–last). The statistical significance of differences (i.e. *p* values) is indicated for every group relative to the other, indicated in the upper section of the applicable set of graphs.

(f) to (j): Important urinary metabolites reflecting host–microbe responses seen in TBM expressed by the Mtb–host biosignature (SUM-4) in respect of different groups of patients and controls. Results on the important host-microbe biomarkers seen in TBM patients relative to the controls and non-TBM cases are shown for quinolinic acid (a), an indicator of activated tryptophan catabolism; methylcitric acid (b), that indicates the Mtb methylcitrate cycle; 4-hydroxyhippuric acid (c), a biotransformation product from the gut microbiome. The biosignature of four Mtb–host metabolites (SUM-4) for the different patient groups and controls is shown in (d) and in (e) for the two TBM subgroups (with mild and severe clinical conditions respectively affecting their long-term outcomes) on admission to hospital and after two weeks of intensive treatment. The statistical significance of differences (i.e. *p* values) is indicated for every group relative to the other, indicated in the upper section of the applicable set of graphs.

8.4.2.2 Perturbed energy metabolism

The increased levels of ketones and hydroxyl acids derived from the branched-chain amino acids leucine (3-hydroxyisovaleric acid – Table 8-1 and Figure 8-2b), isoleucine (2-methyl-3-hydroxybutyric acid — Table 8-1, also indicated as a ketosis marker) and valine (3-hydroxyisobutyric acid; VIP = 0.41; p = 0.154; FC = 1.94) suggests a link to the functional down-regulation of the neuron-associated monocarboxylate transporter-1 (MCT-1) in the CNS (Pierre & Pellerin 2005). Down-regulation of MCT-1 results in decreased neuronal energy supply, which manifests in the coma-related clinical symptoms, noted for some of the TBM patients (Mason *et al.* 2015). Further changes in the concentration of several intermediates — isocitric, 2-ketoglutaric and succinic acids of the tricarboxylic acid cycle (TCA); Table 8-1 — indicate perturbations in this central pathway of energy metabolism. It is important to note, however, that energy production through the TCA not only operates in the host, but is also one of the primary sources of energy for Mtb (Savvi *et al.* 2008), as is discussed below. Finally, a relationship between urinary levels of vanillylmandelic acid, designated as significant in Table 8-1, is long known to occur in stress conditions (Brantley *et al.* 1988), and might relate to the stress experienced by sick paediatric patients in a hospital environment.

8.4.2.3 Liver damage

A significant rise in 4-hydroxyphenylpyruvic acid (Table 8-1) was observed relative to the control group (Figure 8-2c), whereas the increased presence of 4-hydroxyphenyllactic acid (Table 8-1) and 4-hydroxyphenylacetic acid (VIP = 0.14; p = 0.40; FC = 1.40) was likewise observed, defined as non-specific elevations caused by affected liver function (Blau *et al.* 2014). A key aspect of TBM treatment is the application of high doses of a variety of antibiotics, known to affect liver function in patients with pulmonary tuberculosis (Lal *et al.* 1972). The proposal of Lal and co-workers (1972) that patients receiving sustained antibiotic treatment should be kept under close clinical-chemistry observation to monitor liver function therefore warrants attention also for TBM patients receiving treatment.

We have thus highlighted host metabolic responses to Mtb infection, but none of the 12 metabolites observed seemed to qualify on quantitative or qualitative grounds as obvious TBM-specific biomarkers.

8.4.3 Indicators of host–microbe response in TBM

We consequently propose classification of the remaining 9 important metabolites listed in Table 8-1 into three groups of host–microbe responses to TBM: (i) Mtb-induced host

tryptophan metabolism (2 metabolites), (ii) Mtb-associated metabolites (3), and (iii) gut microbiome catabolism and biotransformation (4). A representative metabolite from each of these groups is shown in Figures 8-2(f-h).

8.4.3.1 Mtb-induced tryptophan metabolism

Tryptophan is an essential amino acid and the precursor of serotonin, a brain neurotransmitter. Increased urinary levels of tryptophan and its metabolites have been reported for several mental conditions or pathophysiological states of the CNS. The mean concentration of tryptophan in the urine of TBM patients (33.0 µmol/mmol creatinine (Cr) — a standard method of amino acid analysis is described in SI) — which exceeds the normal age-related reference values of 2–29 µmol/mmol creatinine (Venta *et al.* 2001), differed significantly (p = 0.0030; FC = 2.02) from that of controls (16.3 µmol/mmol creatinine), as was observed for indole-3-acetic acid and 5-hydroxyindole acetic acid (Table 8-1), which are products from the tryptophan/serotonin precursors.

The significant increase in quinolinic acid in the TBM patient group (Table 8-1 and Figure 8-2f) indicates an activated kynurenine pathway, supported by increased urinary excretion of kynurinic acid (VIP = 0.08; FC > 200; p = 0.006) and the microbial metabolite N-acetylanthranilic acid (VIP = 0.15; FC = 2.1; p = 0.025). It has been suggested that activated brain microglia and macrophages release quinolinic acid, thereby affecting neuronal function (Heyes *et al.* 2001). Some studies have implicated the kynurenine pathway and quinolinic acid as being involved in HIV-associated neurocognitive disorders (Kandanearatchi & Brew 2012). Moreover, a retrospective study of CSF from 100 sick Malawian children, including 3 cases with TBM, revealed elevated levels of quinolinic acid (Medana *et al.* 2003). More recently, it was found that Mtb infection induced marked up-regulation of indoleamine 2,3-dioxygenase-1 (IDO-1) expression in human lung macrophages *in vitro* (Blumenthal *et al.* 2012). IDO-1 catalyses the initial rate-limiting step in tryptophan metabolism, thereby regulating L-kynurenine availability for the formation of downstream metabolites, such as quinolinic acid (Nishizuka *et al.* 1970). Finally, it should be noted that an involvement of vitamin B6 in tryptophan metabolism should likewise be taken into account. Pyridoxal 5′-phosphate, the active form of vitamin B6, acts as cofactor for two key enzymes of the kunurenine pathway: kynurenine aminotransferase and kynureninase (Van de Kamp & Smolen 1995). Results on bacterial meningitis caused by *Streptococcus pneumoniae* provided evidence that attenuation of apoptosis by vitamin B6 is multi-factorial, including down-modulation of inflammation, up-regulation of the neuroprotective brain-derived neurotrophic factors and prevention of the exhaustion of cellular energy stores (Zysset-Burri *et al.* 2013).

Thus, the marked presence of quinolinic acid in urine samples from all 12 TBM patients (range 3.2–64.3 µmol/mmol creatinine), compared to the reference urinary value of 3.9 (0.9–15.1) µmol/mmol creatinine (Bouatra *et al.* 2013), as well as kynurenic acid observed in 6 of the TBM samples (range 0.04–1.19 µmol/mmol creatinine) and in none of the controls, clearly qualifies these metabolites from the kynurenine pathway as potential biomarkers for TBM diagnosis and disease progress during treatment.

8.4.3.2 Mtb–host related metabolites

Methylcitric acid is a human metabolite that becomes elevated in body fluids of patients with monogenic disorders, such as propionic acidaemia, methylmalonic aciduria and multiple carboxylase deficiency. We speculate that the significantly increased (p = 0.0001) methylcitric acid (Figure 8-2g) observed here is most likely to have originated from the well-characterized methylcitrate cycle of Mtb (Muñoz-Elías *et al.* 2006; Savvi *et al.* 2008). Mtb is predicted to survive on alternative carbon sources when persisting within the human host. Such carbon sources are partly derived from the catabolism of odd- and branched-chain fatty acids, branched-chain amino acids, and cholesterol, which generates propionyl-coenzyme A as a terminal, three-carbon source, and acetyl-CoA as a two-carbon source for Mtb's energy production. Furthermore, in Mtb, propionyl-CoA and acetyl-CoA are oxidized by the methylcitric acid (MC) and TCA cycles, respectively.

The TCA in Mtb is unusual in that 2-ketoglutarate decarboxylase (Kgd) and succinic-semialdehyde dehydrogenase substitute for the host's 2-ketoglutarate dehydrogenase (KDH) and succinyl-CoA synthetase (Muñoz-Elías *et al.* 2006). It was suggested by Tian *et al.* (2005) that in Mtb the TCA cycle may be operative in the half-cyclic mode characteristic of microbes adapted to low-oxygen conditions, leading to 2-ketoglutaric acid and glutamic acids via the oxidative branch and succinic acid via the reductive branch of the TCA cycle. These researchers indicated that Mtb lysates contained glutamic acid, estimated to be in the high millimolar to molar range, suggesting that Mtb might accumulate glutamic acid during growth as a reserve for use under more stringent conditions in the host. Glutamic acid is also known to cyclize spontaneously to become pyroglutamic acid, which could relate to the noteworthy presence of both glutamic and pyroglutamic acids (Table 8-1), which were observed in the urine samples of the present TBM patient group.

The observed uracil in the urine of the TBM patients relative to controls (Table 8-1) remains speculative. Uracil has been associated with the gut microbiome (Jacobs *et al.* 2008) or may pertain to the uracil excision repair mechanism that is ubiquitous in all domains of life (Maizels & Scharff 2004). Mtb has an increased risk of accumulating uracil residues, given

the G + C-rich content of its genome and occupying a niche during TBM within the host macrophages, where it is exposed to reactive nitrogen and oxygen species. The latter are two major causes of cytosine deamination (to uracil). A Mtb-specific excision mechanism (Kumar *et al.* 2011) may potentially account for uracil increases seen in urine from TBM patients.

Thus, from the discussion on Mtb-induced tryptophan metabolism and the confluent mode of the host and Mtb energy cycles, we propose that metabolites from these Mtb-related and host-perturbed pathways are indicators of choice for a TBM biosignature. Interestingly, in this regard, we observed a positive correlation (Spearman's r = 0.61 p = 0.002) between quinolinic and methylcitric acid concentrations in all TBM samples collected both before and after treatment.

8.4.3.3 Gut microbiota catabolism and biotransformation

Four metabolites listed in Table 8-1 are indicators of metabolism of gut microbiota and the host's response to these compounds: hippuric acid, 4-hydroxyhippuric acid (Figure 8-2h), phenylacetylglutamine and 4-cresol. The hippuric acid links to a drug-like phase II metabolic response of the host to products generated by the microbiota, indicating a highly active glycine-conjugated GLYAT-biotransformation system in TBM. Notably, it was suggested that the GLYAT-complex regulates systemic levels of amino acids that are also utilized as neurotransmitters in the central nervous system (Badenhorst *et al.* 2014).

A study was reported recently on 11 active pulmonary TB patients, who were followed up longitudinally while being treated and whose urine metabolic profiles were compared with those of 11 healthy controls (Das *et al.* 2015). Several metabolites, including 4-hydroxyhippuric acid, showed a treatment-dependent trend in the urine metabolome of follow-up samples; subjects declared as clinically cured showed a similar metabolic profile to those of asymptomatic healthy subjects. Das *et al.* (2015) proposed that the deregulated tyrosine–phenylalanine axis reveals a potential target for diagnostics and intervention in TB. Additionally, an experimental animal investigation on autoimmune encephalomyelitis (EAE), a mouse model for multiple sclerosis, showed that gut bacteria can promote neurological inflammation (Lee *et al.* 2010b) — suggesting that modulation of gut bacteria may provide targets for therapeutic interventions in extra-intestinal inflammatory diseases.

Phenylacetylglutamine and 4-cresol indicate well-known gut microbial co-metabolites, which were increased in the present TBM patient group (Table 8-1). In general, the category of metabolites is regarded as key for the identification of potentially important associations

between changes of bacterial community structure and dynamics of host metabolic patterns that can be used to refine existing and develop new hypotheses on the relationships between dysbiosis and disease (Li *et al.* 2008b).

8.4.4 From metabolomics to the clinic

To be relevant for clinical practice, a biomarker should provide convincing information related to the clinical question regarding a particular patient. Key requirements for a biomarker thus include improvement in the timing and accuracy of diagnosis, minimizing the invasive procedure needed for the final diagnosis and usefulness in monitoring disease progression and efficacy of treatment (Mamas *et al.* 2011). The focus for finding novel diagnostic tools in infectious diseases has been proposed to shift from single disease-specific biomarkers towards bioprofiles or biosignatures, comprising a well-defined set of reliable molecular indicators (Parida & Kaufmann 2010). The developmental path from high-throughput metabolomics data to validated biosignatures has, however, proved to be complex and slow. Rational design of a host biosignature should include elucidation of the relationship of differentially perturbed metabolites within functional clusters and requires validation of candidate metabolites and their combination in biosignatures. We base our proposal for a biosignature for TBM diagnosis on the dysfunctional metabolism generated by the disease that emerged from the present and other metabolomics investigations. We therefore propose two biosignatures that answer different clinical issues in TBM: a biosignature indicating the level of ketosis in clinically neglected TBM, and a Mtb–host biosignature specific for meningitis due to Mtb infection.

8.4.4.1 A biosignature for ketosis

The biomarkers comprising this signature are not new, but illustrate the value of metabolite profiling in disclosing the critical clinical condition of some TBM patients on admission to hospital; especially relevant in resource-poor settings. The proposed ketosis biosignature consists of equal contributions in respect of the concentrations of four metabolites: 2-hydroxybutyric, 2-hydroxyvaleric acids (Asano *et al.* 1988) (VIP = 0.80; FC = 2.73; p = 0.0002), 3-hydroxybutyric acid, and acetoacetic acid. The first two presented as VIPs in the investigation (Table 8-1), whereas the latter two showed highly significant values for at least two of the statistical parameters (Table 8-1) and are universally accepted clinical markers of ketosis.

The mean values relative to the controls was higher for the non-TBM patients than in the TBM group (Figure 8-2d). Although the data distribution is clearly skewed by a number of

outliers and a digression between mean and median values for the non-TBM group, the difference remained significant when tested with non-parametric methods. Figure 8-2e shows the results for the TBM group at admission and after two weeks of treatment. This reflected their poor clinical condition at admission as well as following the intensive antibiotic TB treatment. Mean values of the ketosis biosignature after treatment were reduced to control levels. Notably, Figure 8-2e indicates a striking difference between the ketotic state of two subgroups comprising the TBM group as a whole: three TBM patients in the severely ill group presented with very high values for the ketosis biosignature, whereas the mean of this subgroup as a whole remains above the normal controls as well as that of the subgroup with a mild condition.

Thus, the information disclosed by the ketosis biosignature derived from the metabolomics patients concurs with the clinical observations for the respective patient groups and supports the sub-division of mild and severe general clinical state of the patients observed by the paediatricians at the admission of the patients.

8.4.4.2 Exploring a potential Mtb–host biosignature for TBM

We next identified metabolites with potential for a Mtb–host biosignature based on the information in Table 8-1 and on the integrated metabolic response in TBM, as represented schematically in Figure 8-3.

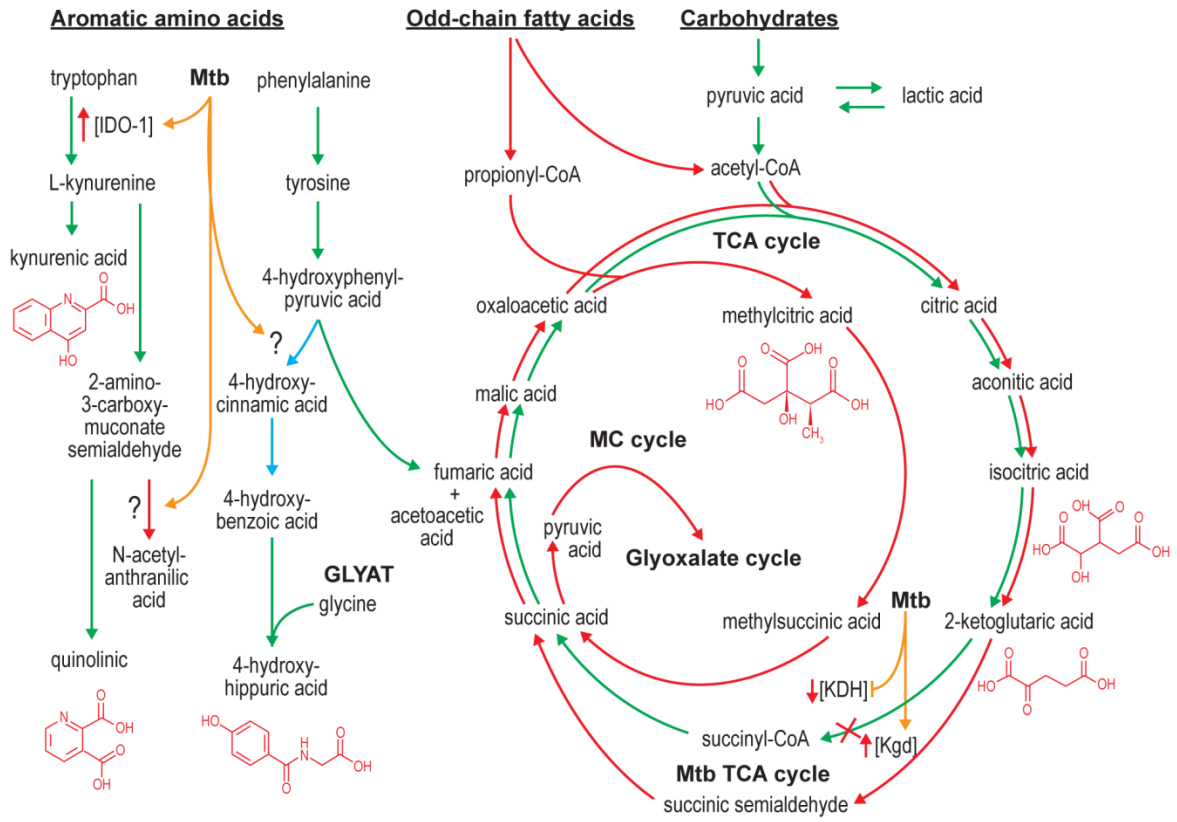


Figure 8-3: A global metabolite profile related to TBM.

Host metabolic pathways for tryptophan, phenylalanine and carbohydrates, including the TCA cycle, are shown in green; predicted pathways for Mtb are shown in red and for the gut microbiome in blue. Orange pathways are proposed signalling events induced by Mtb, leading to up-regulated (↑) and down-regulated (↓) host enzymes or to a postulated (?) event. Mtb catabolism of alternative carbon sources includes odd-chain fatty acids, which generates propionyl-CoA as a three-carbon source, which is metabolized via the methylcitrate (MC) cycle. Acetyl-CoA is catabolized in the TCA cycle, which is unusual in Mtb (Mtb TCA cycle), in that 2-ketoglutarate decarboxylase (KDH) and succinic-semi-aldehyde dehydrogenase substitute for 2-ketoglutarate dehydrogenase (Kgd) and succinyl-CoA synthetase, thereby generating succinic acid that may also function as substrate for the Mtb glyoxalate cycle. Glycine conjugation in phase II detoxification of metabolites from the gut microbiome is catalysed by glycine N-acyltransferase (GLYAT; E.C. 2.3.1.13). The structural formulas for the seven metabolites comprising the SUM-7 discriminator are shown in red.

It is well established that invasive pathogens exploit and assimilate nutrients from their hosts in order to ensure their survival. In TBM, Mtb survives in CNS macrophages and microglia, with their TCA cycle as the core energy providing metabolic pathway, and benefiting from precursors from carbohydrate, lipid and amino acid catabolism. Genomic analysis and enzyme measurement of cell lysates (Muñoz-Elías *et al.* 2006) indicated that citrate synthase, aconitase, isocitrate dehydrogenase, succinate dehydrogenase, fumarase and malate dehydrogenase are present in Mtb but that the pathogen lacks 2-ketoglutarate dehydrogenase (KDH) activity; Mtb instead encodes 2-ketoglutarate decarboxylase (Kgd) and succinyl-CoA synthetase that produces succinic semialdehyde as a stage towards succinic acid, thereby completing the Mtb TCA cycle. In addition, odd-chain fatty acids are degraded via the β-oxidation cycle to yield acetyl-CoA (substrates for the TCA cycle) and propionyl-CoA units, which are further oxidized through an MC cycle (Muñoz-Elías *et al.* 2006).

Key metabolites illustrated in Table 8-1, reflecting potential Mtb-based metabolic aberrations, are methylcitric, isocitric and 2-ketoglutaric acids, quinolinic acid (the second largest VIP value) and 4-hydroxyhippuric acid (presenting with the largest VIP value). These five metabolites, and kynurenic acid (Table 8-1) from the kynurenine pathway, are highlighted by structural formulas in red in Figure 8-3. The sum of a set of these 6 metabolites, designated as SUM-6, was considered as a potential discriminator between the controls (n = 31) and 4 patient groups: (1) those with confirmed TBM (n = 12), with sampling done at admission to hospital; (2) the same patients (n = 12) after two weeks of intensive treatment; (3) the non-TBM patients (n = 21); and (4) patients with confirmed FMS (n = 17). Logistic regression models were constructed to assess the effect of SUM-6 in the likelihood that subjects are not healthy (i.e. are different from the control subjects). The outcomes of these analyses are reported in Table 8-2 and discussed below.

Table 8-2: Outcomes from the logistic regression models applied to the different experimental groups.

Parameters of the logic regression models	Six potential biomarkers for TBM (SUM-6) (Comparison between controls and four experimental groups)				Putative biosignature for TBM (SUM-4) (Comparison between controls and TBM, TBM-treated and non-TBM)		
	TBM [First]	Non-TBM	FMS	TBM [Last]	TBM [First]	Non-TBM	TBM [Last]
Wald-Chi square *p*-value (R^2-value)	0.0019 (0.765)	0.0676 (0.096)	0.2509 (0.036)	0.0036 (0.709)	0.0192 (0.840)	0.2916 (0.033)	0.0428 (0.841)
-2Log L model (intercept only)	18.048 (49.576)	60.9 (64.443)	66.68 (68.029)	21.371 (49.572)	13.064 (49.572)	63.242 (64,443)	13.009 (49.572)
Odds ratio OR value (95% CI)	4.12 (2.048;13.581)	1.656 (0.98;2.95)	1.422 (0.785;2.673)	3.69 (1.818;11.301)	9.256 (2.509;165.3)	1.514 (0.721;36.05)	33.912 (3,993;>999.999)
Percentage correct classifications	96.3	66.4	63.7	94.8	97.4%	56.3%	97.4%
ROC model AUC (95% CI)	0.962 (0.914;1)	0.6742 (0.515;0.833)	0.648 (0.490;0.806)	0.948 (0.888;1)	0.974 (0.932;1)	0.594 (0.426;0.761)	0.974 (0.935;1)
ROC AUC CV (95% CI)	0.954 (0.897;1)	0.546 (0.358;0.734)	0.461 (0.277;0.645)	0.937 (0.868;1)	0.966 (0.913;1)	0.310 (0.127;0.494)	0.968 (0.923;1)
H-L Model AUC contrast *p*-value	0.271	0.0001	<0.0001	0.174	0.382	<0.0001	0.302
Hosmer–Lemeshow *p*-value	0.929	0.887	0.135	0.915	0.993	0.165	0.968

Model significance was assessed using the overall Wald Chi-square test (H0: $\beta_i = 0$ for all i) and max-rescaled R^2 values. Logistic regression allows prediction of a discrete outcome, such as group membership, based on information from a set of variables — here indicated by SUM-6. The outcome is deemed to be practically significant if Wald $p < 0.05$ and $R^2 > 0.6$, which was observed for the TBM patients on admission and after treatment ($p = 0.002$ and $R^2 = 0.7701$ and $p = 0.0097$ and $R^2 = 0.7144$, respectively). The results reported in Table 8-2 indicate that the SUM-6 values can discriminate between the controls and TBM patients on admission as well as after treatment, implying that Mtb infection still prevailed in the TBM patients as a group, in spite of being treated for two weeks. The importance of the predictor was subsequently determined through the odds ratio and the corresponding 95 % confidence interval, along with the log likelihood fit statistic as compared to the intercept only model. The -2 Log L statistic should decrease dramatically once the predictor is added to the model, which was observed for both TBM groups. The decrease was from 49.5 to 17.7 for the TBM patients on admission, and from 49.5 to 21.0 for the group following treatment.

The observations for the non-TBM and FMS groups (Table 8-2, rows 3 and 4) were distinctly different from those described above for the two patient groups. Taken together, 96.6 % and 94.9 % of the TBM patient groups (on admission and after treatment, respectively) were correctly classified for TBM, with 68.1 % and 64.2 % for the non-TBM and FMS groups, respectively, suggesting that the metabolic aberration expressed by the SUM-6 discriminator did not prevail in the two latter groups — SUM-6 correctly predicted the presence of TBM.

Next, the prediction accuracy of each model was reported and evaluated using a Receiver Operating Characteristic (ROC) analysis (Fawcett 2006). ROC curves indicating sensitivity and specificity of the separate models which discriminate the control and 4 patient groups are shown in Figure 8-4a. The area under the curve (AUC), along with the respective 95 % confidence intervals, gives a measure of how well SUM-6 could distinguish between the controls and each diagnostic group (Table 8-2). The value of AUC = 1 represents a perfect test; a cursory guide for classifying the accuracy of a diagnostic test is as follows: AUC = 0.90–1 (excellent, i.e. high sensitivity and high specificity); 0.80–0.90 (good); 0.70–0.80 (fair); 0.60–0.70 (poor); 0.50–0.60 (fail). Cross-validated AUC (AUCCV) estimates and confidence intervals were used to detect over-fitting along with a contrast statistic. A value of $p < 0.05$ indicates a significant drop in AUC when cross-validation is performed, that is, an indication of overfitting. The AUC values from ROC analysis, as well as the cross-validated values, indicated an excellent differentiation between the two TBM patient groups and the controls (TBM on admission: 0.9655 and 0.9569; TBM after treatment: 0.9483 and 0.9425). The AUC values as well as the cross-validated results for the two other groups proved to be

poor (non-TBM: 0.6842 and 0.5535; FMS: 0.6536 and 0.4663), lending further support for the discriminator potential of SUM-6 to distinguish TBM patients from controls and other disease groups.

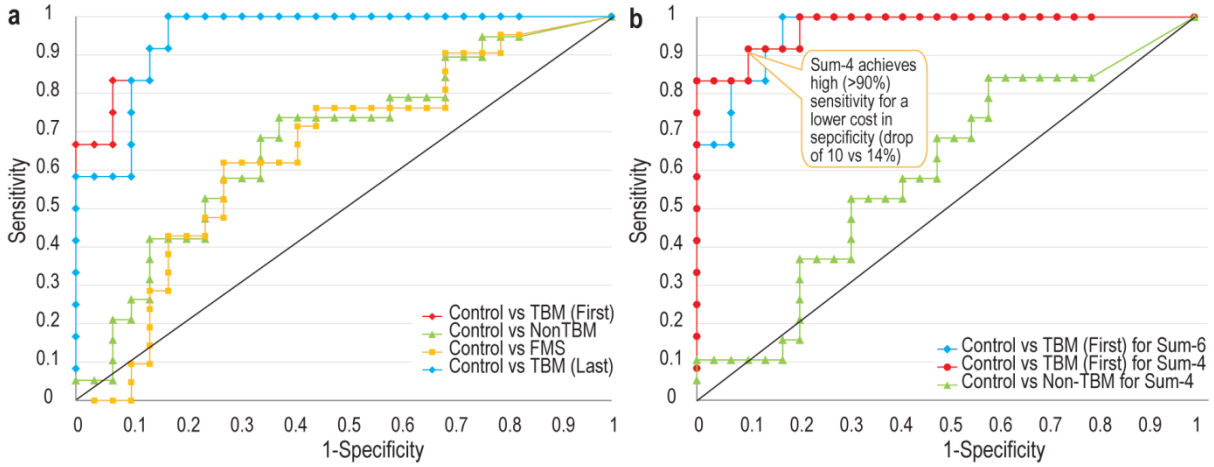


Figure 8-4: **ROC analyses for discriminating TBM, TBM (treated), non-TBM and FMS patients from controls. (a) The discriminator consisted of 6 metabolites (SUM-6) identified by multivariate, univariate and metabolic pathway analyses. Colour code: TBM at admission: red; treated TBM: blue; non-TBM: green; FMS: orange. (b) The comparison between discriminators SUM-6 (shown in blue) and SUM-4 (shown in red) for TBM, and with discriminator SUM-4 (shown in green) for non-TBM.**

Finally, the Hosmer–Lemeshow statistic was used as a goodness-of-fit measure, with larger *p*-values (closer to 1) essentially indicating a good logistic regression model fit or that a more complicated model would not necessarily give better results, as was observed for the TBM group on admission ($p = 0.8844$).

8.4.4.3 Proposal for a putative Mtb–host biosignature for TBM

We evaluated the respective contribution of each of the six biomarkers individually towards a TBM diagnostic model. For this purpose we performed (1) a ROC analysis for each of the individual biomarkers constituting SUM-6, and (2) the Hosmer–Lemeshow statistical test for goodness of fit for the logistic regression model. We did not make use of stepwise logistic regression to select predictors, because adding predictors individually would result in a slightly more complex model to apply in practice.

The outcome of the logistic regression analysis and PCA-power data (Table A3-2) indicated that quinolinic, methylcitric, 4-hydroxyhippuric and 2-ketoglutaric acids outperformed the two other components as the best discriminating metabolites for a TBM biosignature (see Table A3-3 for a summary of the model parameter estimates). We thus propose a combination of these four biomarkers as the putative Mtb–host biosignature (abbreviated here as SUM-4) for TBM. SUM-4 gave an excellent diagnostic value (AUC > 90 %) for the TBM group on admission relative to the non-TBM patients (Figure 8-4b and Table 8-2) and SUM-4 achieved high (>90%) sensitivity for a lower cost in specificity (drop of 10 vs 14 %) (TBM: AUC = 95.61 %; Wald p = 0.008).

The information gained from the putative biosignature (SUM-4) on the TBM patient group is presented in Figure 8-2i. It indicates a statistically significant difference between the TBM group at admission and controls ($p < 0.0001$) and the non-TBM ($p < 0.0001$) groups. The mean value for the non-TBM group was comparable to the mean value of the control group. Figure 8-2j shows the response of the mild and severe TBM subgroups to two weeks of intensive TB treatment. Most interestingly, the mean values of the four biomarkers for both subgroups did not differ significantly ($p > 0.05$) after treatment, although the mean values decreased for the mild and increased for the severe subgroups.

These data suggest that the Mtb–host biosignature, comprising the 4 selected biomarkers, could be useful in clinical practice to identify high-risk individuals during the early phase of treatment, if necessary to modify the treatment regime according to the individual requirements of the patient. Furthermore, the Mtb–host biosignature indicated risk of TBM similarly to the ketosis biosignature, rendering thier combination as a useful criterion to predict TBM risk for individuals, even to the extent of advising on the implementation of individualized preventive and therapeutic strategies.

8.4.4.4 Perspective

The information disclosed by the Mtb–host biosignature (SUM-4) that we derived from this metabolomics investigation is a novel potential indicator of the state of disease prevailing in the TBM patients, and supports the clinically observed distinction of the mild and severe states of the patients made by the paediatricians upon admission. Yet, this study has important limitations, most of which are typical of the exploratory nature of metabolomics studies, and influences the power of these initial metabolomics findings. First, the study was not specifically designed to identify biomarkers for TBM, and samples from only a limited number of cases and controls were available for our metabolomics data set. Second, due to its impromptu design, the availability of detailed pre-diagnostic lifestyle/dietary and treatment

data prior to admission to hospital was limited; in hindsight, it did, however, strengthen insights into the practice of health care for paediatric patients experiencing TBM in a resource-poor environment. Third, although the study defined unique associations between metabolite profiles and TBM, it was not possible to perform a comprehensive analysis including all important metabolites. Five potential markers had to be excluded from the data analysis on the basis of insufficient analytical verification of their chemical structures. By implication, once the structure of these unknown variables has been elucidated, we may achieve an improvement in the biosignature presented here. However, these limitations do not negate that the urine-derived Mtb–host biosignature indicated diagnostic and prognostic insight into the TBM patients investigated. Indeed, we believe we have demonstrated a useful approach to identifying individuals at risk of having contracted TBM that justifies validation of the signature across cohorts from different geographical locations and ethnic background to assess its robustness for clinical application.

8.5 Conclusions

A rapid, sensitive and selective diagnostic single biomarker for TBM has not been identified hitherto. The explorative GC-MS metabolomics investigation of urine from clinically ill TBM patients reported here is the first based on the metabolite profile of urine from such patients on admission to hospital. This study provided novel insights into the variables observed in these samples, from which we propose that the sum of the urinary concentrations of quinolinic and 2-ketoglutaric acids (generated by host metabolism), methylcitric acid (from Mtb metabolism) and 4-hydroxyhippuric acid (gut microbial metabolism) holds promise as a biosignature for TBM (SUM-4). These results provide proof-of-concept for the development and validation of a medical algorithm in future prospective studies, using urine analysis for the diagnosis of TBM and for a non-invasive approach to monitoring subsequent medical treatment. A distinct advantage of such a development is the relatively uncomplicated approach to obtaining valuable samples for longitudinal short- and medium-term studies to monitor the benefits of alternative TBM treatments, especially in paediatric patients.

Acknowledgements: S.M and R.S are recipients of a Desmond Tutu-NRF-VU doctoral fellowship for a joint PhD study between the Vrije Universiteit in Amsterdam, the Netherlands, and North-West University (Potchefstroom), South Africa. Research funding for this project was provided by the Technological Innovation Agency (TIA) of the Department of Science and Technology of South Africa.

Conflict of interest: There is no conflict of interest to disclose.

Compliance with ethical requirements: The present study was approved by the Human Research Ethics Committee of Stellenbosch University, South Africa (study no. N11/01/006).

Supplementary information: see Annexure 3.

PART 4: TAKING A STEP BACK AND REFLECTING:

A DISCUSSION ON THE TAKE-HOME MESSAGE OF THIS THESIS AND WHERE THIS RESEARCH CAN BE TAKEN

CHAPTER 9 DISCUSSION: INSIGHTS INTO TBM VIA METABOLOMICS

9.1 Recapitulation: motivation behind addressing the topic under investigation

According to the WHO, South Africa has the highest prevalence of TB in the world. Within the Western Cape Province, TB is endemic and as a result the occurrence of TBM, the most severe manifestation of TB, is relatively high compared to the rest of the world (Chapter 1, Fig 1-1). Hence, TBM is a serious health issue within South Africa, and particularly dangerous in infants and children. However, there can be a high survival rate with treatment following early diagnosis.

The application of metabolomics technology to the study of TB and meningitis is increasing, yet only a few related research papers have been reported on TBM specifically (Chapter 5). Using the great range of clinical knowledge, metabolomics technology and analytical and statistical expertise available across multiple academic institutions, it appeared sensible to employ metabolomics as the scientific method of choice — the context of this thesis.

Fortunately, I was able to acquaint myself with NMR metabolomics and was involved in an unrelated study whereby I could contribute by adding NMR data to the GC-MS data in the investigation on inherited respiratory deficiency diseases. That study by Smuts *et al.* (2013), presented at the end of Chapter 2, coincidently also dealt with a paediatric population. The success in defining a biosignature for these diseases provided a template for how NMR and GC-MS metabolomics data can be used in the biological interpretation of TBM. Hence, the study by Smuts *et al.* (2013) is included in this thesis as proof-of-concept that a putative biosignature can be identified by applying metabolomics technology.

9.2 Addressing the first aim of the thesis

Aim 1: to gain new insight(s) into the global metabolite profile at the site of infection of TBM by holistically analyzing CSF using untargeted NMR metabolomics which gives a broad overview of host response metabolism.

Two points of background information played an important part in addressing aim 1:

(1) It is known that "*human cases of meningitis caused by different bacterial and fungal species are relatively uncommon in the developed countries and acquisition of sufficient spectral data would require several years ...*" to warrant meaningful investigations on TBM in humans (Himmelreich *et al.* 2009). Some of the clear and defining limitations of previous studies on TBM (Chapter 2) were the lack of samples from TBM cases and their clinical descriptions, and the fact that most were pilot studies — no cross-validation nor generation of hypotheses to be tested subsequently were included. Although there is a close similarity between those studies and the present investigation, the requirements for our TBM study differed. We aspired to have a sufficient, homogeneous, well-described patient group (pg. 32) of TBM cases in a defined age group — infants and children — from the same population, as well as two matched control groups. We were in the fortunate position to gain access to CSF samples from a cohort of clinically well-characterized TBM-positive children, as well as from two control groups, for the metabolomics investigation described in Mason *et al.* 2015.

(2) We applied the traditional univariate and multivariate analyses as previously used in the investigations of respiratory metabolic diseases presented in the template paper (Smuts *et al.* 2013). In addition, I acquainted myself with the Unscrambler® X (V10.3, CAMO software AS, Norway) statistical package in order to perform my own statistical analyses. Both methods yielded virtually identical outcomes. We therefore decided to include the analysis obtained with the Unscrambler® X package in the paper published in *Metabolomics* (Mason *et al.* 2015). This decision was backed by our success with this commercial product; it is, incidently, also the most common method of chemometrics analysis used by our colleagues in Nijmegen, where the NMR analyses were performed.

Against this background, Part 2 of this thesis delves into the heart of aim 1 by means of an untargeted NMR metabolomics study on CSF (Chapter 4 – Mason *et al.* 2015). The outcome of the results required an intensive review of the literature on neuroenergetics in order to understand and formulate our novel AMLS hypothesis.

9.2.1 Shifting neuroenergetic paradigms

The classical view of neuroenergetics is that the blood supplies oxygen and glucose to the brain (Sokoloff 1989). Glucose is the primary source of energy utilized by both neurons and astrocytes. It undergoes complete oxidation via glycolysis, the Krebs cycle and oxidative phosphorylation, which ultimately produces ATP for energy-dependent reactions. Thus glucose is used in the same way by all cell types. Since neurons consume the greatest quantity of energy, metabolic intermediates (e.g., for the Krebs cycle) are diverted towards neurons. Some of the pyruvate produced by glycolysis is converted to lactate and released into the extracellular space. Under this classical viewpoint, lactate is considered a by-product with deleterious effects when in excess (Norenberg *et al.* 1987; Siesjo 1988; Bender *et al.* 1997). Astrocytes have long been thought to play a passive role in supporting neuronal function, with the neuron being the star of the show. However, the dynamic involvement of astrocytes (Ranjbar & Amiri 2015) in the forefront of neuroenergetics is now being realized, shifting paradigms (Haydon & Carmignoto 2006; Giaume 2010). The neurocentric view of neuroenergetics is evolving into a more integrated one of complementary and co-operative metabolic interactions between astrocytes and neurons.

Astrocytes constitute about 50% of the total human brain volume and their cytoarchitectural organization is such that specific sections cover 99% of the surface area of cerebral blood vessels. These astrocytes are the preferential site for glucose uptake from the blood, as well as having projections in peri-synaptic areas of neurons, providing close interaction with neuronal elements and acting as a cellular barrier between blood and neurons. The unique morphological and phenotypic characteristics of astrocytes ideally position them to sense and dynamically respond to changes in neuronal activity (Bélanger *et al.* 2011; Pellerin 2010), lending them to conduct numerous critical functions (Chen & Swanson 2003; Steele & Robinson 2012). Astrocytes therefore support neuronal activity via structural, trophic and metabolic means, suggesting a critical role in regulating neuroenergetics and homeostatic functions (Pellerin 2010). Notably, neurons rely on astrocytes to supply precursors of the Krebs cycle intermediates, or their derivatives, as the enzyme pyruvate carboxylase is present in only astrocytes but not in neurons (Hertz *et al.* 1999).

Astrocytes exhibit a higher capacity for glucose utilization, as well as greater metabolic plasticity, than neurons; these characteristics are important for homeostatic and neuroprotective functions. The high glycolytic rate of astrocytes suggests a preference for the production and release of lactate. Evidence that has emerged over the past two decades has begun to highlight lactate as a supplementary substrate for neurons, resulting in the (re)emergence of a nursing role for astrocytes (Bouzier-Sore *et al.* 2002).

9.2.2 The astrocyte–neuron lactate shuttle (ANLS) hypothesis

In 1996, Magistretti & Pellerin presented the framework of a hypothesis that they have since developed and refined to become one of the prevailing viewpoints of neuroenergetics — the ANLS hypothesis (Magistretti & Pellerin 1996). Basically, the hypothesis states that astrocytes respond to intensified neuron activity by increasing their rate of glucose uptake, glycolysis and the release of lactate into the extracellular space, as shown schematically in Figure 9-1. At the metabolic level, it begins with glutamatergic activity, a process whereby increased neuronal activity results in the release of glutamate, the main excitatory neurotransmitter in the brain, into the extracellular space along the glutamate transporter EAAT3 exclusively located in neurons. Astrocytes sense increased activity at the glutamatergic synapses, followed by glutamate uptake via the glia-specific glutamate transporters EAAT1 and EAAT2. The transport of glutamate is driven by a sodium gradient (i.e., by a Na^+-dependent mechanism), with a stoichiometry of three Na^+ ions co-transported with one glutamate, resulting in a significant increase in intracellular Na^+ concentrations in astrocytes (Magistretti & Pellerin 1999a; Pellerin & Magistretti 2004; Bélanger *et al.* 2011). Glutamate taken up by astrocytes is converted to glutamine through an ATP-dependent reaction catalysed by astrocyte-specific glutamine synthetase. Glutamine is released back into the extracellular space and taken up by neurons, where it is converted to glutamate by glutaminase (Magistretti & Pellerin 1999a; Bélanger *et al.* 2011). This reaction thereby replenishes the neurotransmitter pool of glutamate and completes the glutamate–glutamine cycle. Glutamate uptake by astrocytes stimulates glucose uptake with a stoichiometric relationship of 1:1 between uptake of glutamate and glucose. Increased concentrations of Na^+ in astrocytes activate the enzyme Na^+–K^+–ATPase, particularly the α_2 subunit. The result is the triggering of glycolysis, leading to the production and release of lactate into the extracellular space, which is then taken up as an energy substrate by neurons for oxidative-derived ATP production.

Astrocytes are therefore viewed as ‘lactate sources’ supplying the extracellular lactate pool, and neurons are ‘lactate sinks’ consuming lactate oxidatively in response to energy demands. The transfer of lactate from astrocyte to neuron can be viewed as a spatially and temporally independent process (Pellerin & Magistretti 2004). Thus the metabolic plasticity of astrocytes is likely concomitant with synaptic plasticity (Magistretti 2006). Magistretti (2000 & 2009) concisely sums up the ANLS model in terms of neurometabolic coupling in which sodium-coupled uptake of glutamate by astrocytes activates Na^+–K^+–ATPase, which triggers glucose uptake and its glycolytic processing. This results in the production and release of lactate that is used by neurons for activity-dependent energy demands. It has become

generally accepted that lactate is a pivotal component in neuronal brain energy homeostasis (Gladden 2004; Pellerein 2003; Pellerin 2010; Schurr *et al.* 1999; Schurr 2005; van Hall *et al.* 2009).

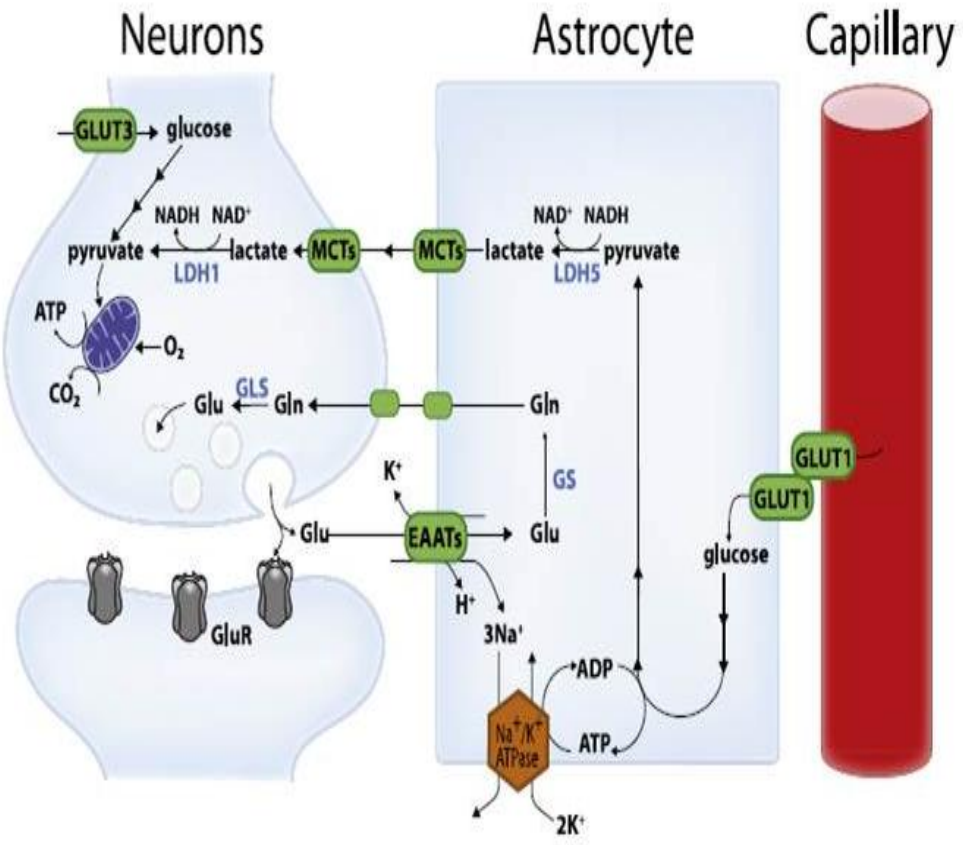


Figure 9-1: Schematic representation of the ANLS model. Glu = glutamate; Gln = glutamine; GluR = glutamatergic receptor; EAATs = excitatory amino acid transporters; GLUT = glucose transporter; MCTs = monocarboxylate transporters (with permission from Bélanger *et al.* 2011).

9.2.3 The brain in crisis

Neuroinflammation, the reaction of surrounding brain tissue to brain infection, is characterized by the synthesis of various inflammatory mediators and reactive gliosis, associated with phenotypic changes and proliferation of glia cells (both astrocytes and microglia), in response to a dynamically changing environment (Giaume *et al.* 2007). The modified phenotype of astrocytes, from basal to reactive state, results in their neglection of

their neuroprotective role — allowing excessive oxidative stress and production of reactive oxygen species (ROS) and subsequent ROS-induced neuronal damage (Pellerin 2003).

In the case of a chronic infection in the brain by a persistent pathogen, such as Mtb, the microglia strive to eradicate the scourge but unwittingly provide the habitat for the bacterium. Similar to the ANLS hypothesis, the astrocyte–microglia lactate shuttle (AMLS) concept (proposed in Mason *et al.* 2015) suggests that when the brain is in crisis due to infection, energy flow in brain metabolism is preferably shifted towards the microglia, at the expense of energy provision to the neurons (see Figure 9-2). In particular, the AMLS hypothesis postulates that in neuroinflammatory infectious diseases such as TBM, lactate, produced by glycolysis in astrocytes, participates in the activated immune response. In association with ketones and gluconeogenic amino acids, lactate is collectively directed from the neurons preferentially into microglia where it enters the mitochondrial citric acid cycle. Here, oxidative phosphorylation produces high levels of ATP and forms of ROS, such as hydrogen peroxide, required for degradation of the invading pathogen. Activation of microglia, as implied by the AMLS hypothesis, does not, however, present a uniform process and involves intricate interactions and feedback loops between the microglia, astrocytes and neurons that hamper attempts to construct basic and linear cascades of cause and effect. TBM involves a complex integration of the responses from the various cell types present within the CNS, with microglia and the astrocytes as the main players.

The bypassing of energy metabolism from the neurons effectively inactivates them in an effort to protect them from neurodestructive agents in response to chronic neuroinflammation. This neuron inactivation is evident clinically as the progression of TBM is reflected by a decline in the Glasgow Coma Score; specifically, reduced consciousness and awareness, as manifest in our TBM patients (Table A1–2). The activation of microglia thus lends itself to that of the proverbial double-edged sword — simultaneously exhibiting neuroprotective and neurodestructive properties in an attempt to save the whole at the expense of the part (Giaume *et al.* 2007; Hall *et al.* 1998).

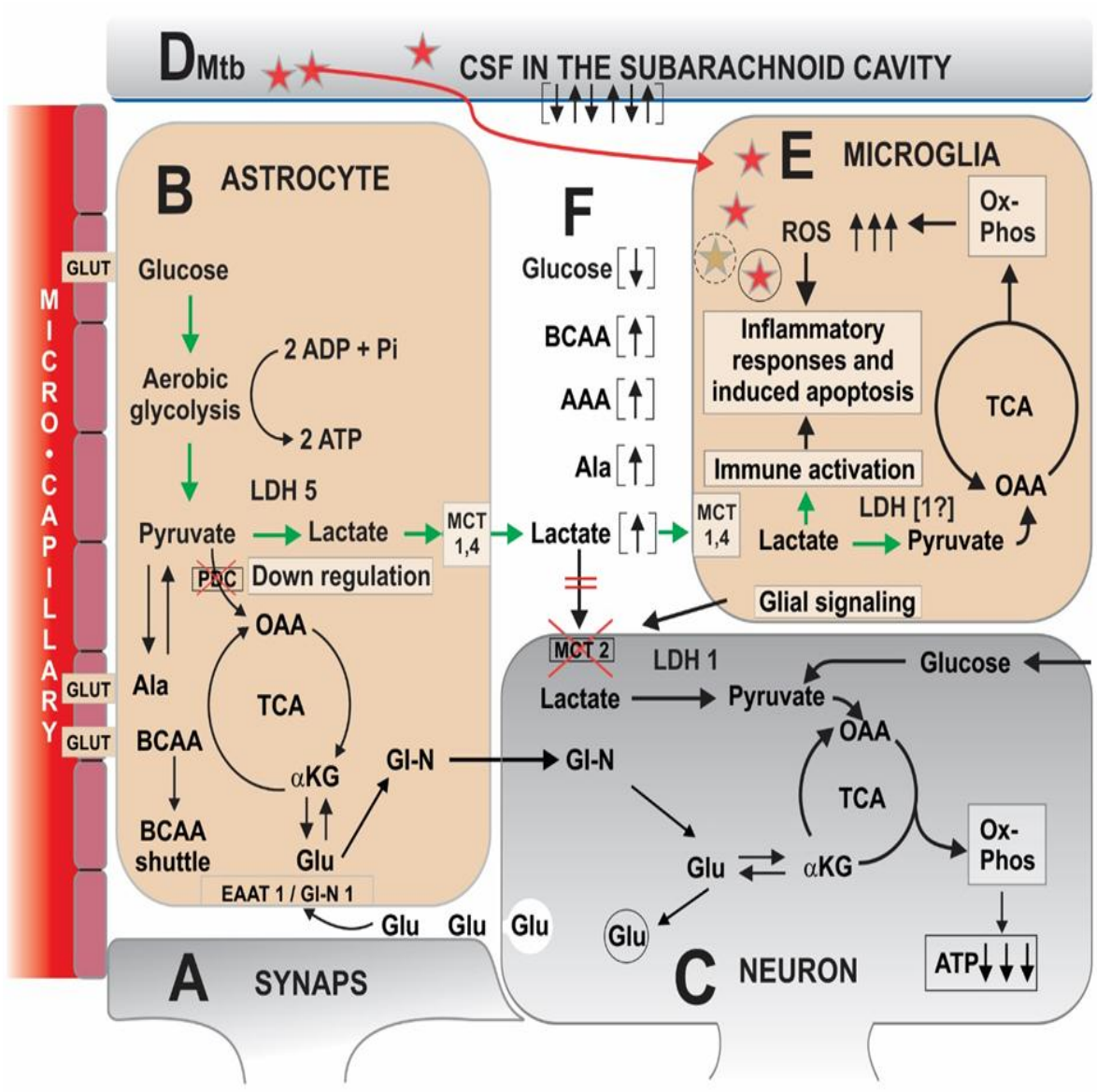


Figure 9-2: **The conceptual AMLS model (described in detail in Annexure 1 — Figure A1-2).**

Interestingly, the important CSF metabolites (VIPs) identified by our untargeted metabolomics study on CSF samples from TBM-confirmed paediatric cases was also derived in a completely independent study using exactly the same NMR data set and a novel nonparametric classification system (van Reenen *et al.* 2016). Van Reenen and co-workers used nonparametric hypothesis testing, based on minimum classification error rates as test statistics, to find statistically significantly shifted variables (the equivalent of VIPs in our study). This technique identified discriminatory and therefore informative variables, transforming error rates into *p*-values (referred to as ERp). The exploratory application of ERp on our NMR-TBM data set indicated that markers of the disease state of patients

suffering from TBM were successfully selected and used to classify patients with meningitis due to Mtb infection relative to other causes. The study by van Reenen *et al.* (2016) therefore serves as a method of mathematical modelling for deductive verification of our AMLS hypothesis (see Figure 5-2).

In summary, Part 2 begins without any hypothesis or paradigm in mind — an illustration of the so-called open mind approach to metabolomics studies (Kell & Oliver 2003), and ends with new insights which led to hypothesis formulation (Mason *et al.* 2015; Mason *et al.* 2016a) as a result of pursuing aim 1.

9.3 Addressing the second aim of the thesis

Aim 2: to investigate correlating urine samples by using the more sensitive semi-targeted GC-MS metabolomics approach — unravelling new information on the vast range of excreted metabolites as a result of the perturbation caused by TBM.

The key aspect of this aim was the unravelling of further information on TBM through urine analysis. A prerequisite for achieving this aim, however, was mastering the complex intricacies of the GC-MS metabolomics method. A novel method was therefore first developed to assess the ability of the analyst to produce repeatable metabolic data to be incorporated into a matrix, eventually to study the underlying perturbation of TBM in a patient group. For this purpose our statistical co-worker, Dr G. Koekemoer, developed the KEMREP method (Mason *et al.* 2014), which has distinct advantages:

(1) KEMREP may be applied in principle to hyphenated mass-spectrometric metabolomics data derived from any subsection of the metabolome [see Chapter 7, Mason *et al.* 2014, Figures 7-5 (on the CSF metabolome) and 7-6 (on the urine metabolome)]. The significance of this approach is its ability to provide insight into consecutive stages in the analytical process to generate metabolomics data. We demonstrate that KEMREP provides a benchmark for assessing the quality of several levels of analytical procedures used in metabolomics, thereby increasing confidence in the biological interpretation of information encapsulated in the data.

(2) Metabolomics is currently in an exponential growth phase (Goodacre 2010). The difficulty with commonly used analytical techniques in the field is that they require a high level of sophistication from analysts, and simultaneously generate complex metabolomics data, which include many potential sources of error. Our KEMREP method describes a theoretical and practical alternative for the analysis of experimental data; it offers an easy and practical new method for the qualitative (illustrated by overlays in Figure 7-5A) as well as the quantitative (illustrated by bounds shown in Figure 7-5B) assessment of the repeatability and reproducibility of the comprehensive analytical process used to generate a metabolomics matrix.

(3) KEMREP has the potential to be widely used in metabolomics practice, which should lead to its further refinement. An important refinement would be to convert the bounds (Figure 7-5B) into a figure of merit (van Batenburg *et al.* 2011), a development which requires advanced statistical insights. To foster application and advancement of KEMREP

for other metabolomics data, a detailed statistical description of the method is presented in the supplementary information (Annexure 2). Also included online (SI of Mason *et al.* 2014) is a script for the method, which is executable using the free statistical software program R.

Once satisfied with the repeatability issue in GC-MS metabolomics analysis, we embarked upon the GC-MS study of the urine samples from TBM patients. The full discussion on the perturbed metabolic profiles in TBM is presented in Mason *et al.* 2016c. It will thus not be elaborated upon here; instead there will be a shift in the focus of the discussion towards disrupted homeostasis in TBM.

9.3.1 Beyond homeostasis

Biological processes maintain stability through altered circumstances, by detecting environmental (external) and physiological (internal) changes, and activating specialized adaptive responses. The dynamic metabolic characteristics of astrocytes lend them to being particularly adept at such a task within the neuronal framework of homeostasis. Beyond homeostasis lies the comparatively new concept of allostasis. Allostasis is the extension of the concept of homeostasis and refers to the maintenance of stability by means of robust, energy-costing adaptive mechanisms in response to severe physical, psychosocial or environmental challenges. Homeostasis involves maintaining a *steady* internal environment of an organism (i.e., *feedback* mechanism); whereas allostasis is more *dynamic* in that it involves the constant evaluation of physiological needs, resulting in biochemical *adaptation* mechanisms. These two concepts seem similar and are not intended to operate independently; instead, allostasis supports homeostasis, placing emphasis on flexible adaptation with the ultimate goal of maintaining a constant stable internal environment. Frequent or chronic challenges (e.g., the gradual, persistent development of tuberculosis and the associated chronic inflammation) produce dysregulation of several major physiological systems by triggering chemical mediators of adaption that operate in a nonlinear network. The cumulative ‘wear and tear’ associated with the inability to disengage these physiological systems is referred to as allostatic load. In vulnerable biological systems an allostatic overload prevails that result in the development of disease (Danese & McEwen 2012; Logan & Barksdale 2008; McEwen 2006; McEwen 2008; McEwen & Gianaros 2010).

Since the term was first coined (Sterling & Eyer 1988), the concept of allostasis in the literature has primarily been used to describe mild, tentative perturbations that are stress related, psychosomatic and/or psychopathologocial (Blair *et al.* 2011; Danese & McEwen 2012; Logan & Barksdale 2008; Shannon *et al.* 2007; Stewart 2006; Tannenbaum *et al.* 2002; Tomiyama *et al.* 2012). As a concept, allostasis is still being developed and so is

gradually being applied to more diverse fields, such as metabolic diseases and lipidomics (Orešič *et al.* 2008). Allostasis has also been used to explain extreme glucose fluctuations as the body's inability to cope with allostatic load in the case of chronic illness, predisposing the individual to serious harm as reflected by high mortality (Rake *et al.* 2010; Stumvoll *et al.* 2003; Stumvoll *et al.* 2004). We speculate that the perturbations observed in the urine samples of TBM patients, as discussed next, are a clear indication of allostatic stress in the disease state, followed by restoration towards homeostasis and normal health if well treated (time interval T4 in Figure 9-3).

9.3.2 TBM and allostasis

9.3.2.1 Ketosis: a primary host response

Following the use of standard statistical methods (both univariate and multivariate) of metabolomics data generated by GC-MS analysis of urine from TBM-confirmed cases and controls, we discerned certain metabolic perturbations related to the disease. First, various markers of ketosis — such as 2-hydroxybutyric, 2-hydroxyvaleric, 3-hydroxybutyric and acetoacetic acids — were clearly identified in high abundance in TBM cases (Chapter 8, Fig. 8-2a). These ketotic markers were effectively restored to normality following treatment of the patients upon admission to hospital, thereby stabilizing the patient (Chapter 8, Fig. 8-2d and 8-2e). Moreover, it was observed that in the TBM cases whereby the follow-up assessment indicated mild-to-severe neurological complications following treatment, there was a strong correlation with abnormally high ketotic markers upon admission to hospital — i.e., children admitted to hospital with suspicion of meningitis and in a highly sick and ketotic state (manifesting as persistent vomiting and diarrhoea, poor feeding, and loss of weight) tended to have a poorer prognostic outcome.

9.3.2.2 Mtb-related metabolic responses

The following are the relevant Mtb-related metabolic perturbations observed in our analysis of urine:

(1) The concentration of tryptophan of TBM patients differed significantly from that of controls, as was observed for indole-3-acetic acid and 5-hydroxyindole acetic acid, which are linked to the tryptophan/serotonin pathway.

(2) The notable increase in quinolinic acid in the TBM patient group indicates an activated kynurenine pathway, supported by increased urinary excretion of kynurinic acid and the microbial metabolite N-acetylanthranilic acid.

(3) As indicated in Chapter 1, Mtb is predicted to survive on alternative carbon sources when persisting within the human host. Perturbed citric acid cycle intermediates can also be linked to the unique citric acid cycle seen in Mtb, and specifically the observed increase in methylcitric acid (Figure 8-2g), is most likely to have originated from the well-characterized methylcitric acid cycle of Mtb.

(4) Increased hippuric acid and 4-hydroxyhippuric acid (Table 8-1) indicates a highly active glycine-conjugated GLYAT-biotransformation system in TBM. Contrary to the detoxification paradigm of hippuric acid excretion, it was suggested that the GLYAT-complex regulates systemic levels of amino acids that are also utilized as neurotransmitters in the central nervous system (reviewed by Badenhorst *et al.* 2014).

More importantly, the information on key metabolites mentioned above indicated a statistically significant difference between the TBM group at admission and controls ($p < 0.0001$) and the non-TBM ($p < 0.0001$) groups (Chapter 8, Fig. 8-2i). The mean value for the non-TBM group was comparable to the mean value of the control group. Figure 8-2j shows the response of the mild and severe TBM subgroups to two weeks of intensive TB treatment. Notably, the mean values of the key metabolites for both subgroups did not differ significantly after treatment, although the mean values decreased for the mild, and increased for the severe, subgroups.

We speculate that the observations on the ketosis markers, and on the Mtb-induced host and microbe markers, reflect (1) the disease state (reaching towards allostatic overload in some patients) of the TBM patients on admission to hospital, and (2) differentiation in the restoration towards homeostasis following treatment. A conceptual model to convey the fundamental principles and systems context following Mtb infection, development of TBM and response to treatment is shown in Figure 9-3. The focus of the discussion will be limited to the period for which samples were collected from the patient groups as supplied for my investigation.

Upon admission to hospital (Figure 9-3, point B) most of the TBM patients presented with moderate to severe symptoms of ketosis. The metabolomics analysis substantiates this clinical picture through the ketosis biosignature, as presented in Mason *et al.* 2016c (Chapter 8, Figure 8-2d). Figure 8-2e shows that cases admitted to hospital with very high ketosis biomarkers tended to have a poor prognosis, leading to severe neurological complications irrespective of treatment. Indeed, in the final paper of this thesis (Mason *et al.* 2016b), in Chapter 10, we incorporate the prognostic outcome of patients into an interesting discussion of the results.

Thus, upon the patient's admission to hospital (event B), the primary objective of the clinician is to stabilize the patient (e.g., by means of a glucose drip, corticosteroids and antibiotics). We depict the improvement of the patient in Figure 9-3 by a transition from a diseased state (away from death), to a condition of allostasis (towards a return to health). However, despite the clinical improvement following two weeks of treatment, the presence of the Mtb is still reflected through the Mtb biosignature (SUM-4) — illustrated by the ROC curve in Figure 8-4a and box plots in Figure 8-2j relative to Figure 8-2i. Event C serves to indicate the point of discharge from the hospital when the patient is clinically stable, under proper treatment and is able to return home. However, the time interval from B to C varies, depending on various factors — such as severity upon admission, any subsequent complications and state of home environment (i.e., whether or not it is expected that treatment would continue). The point of discharge from hospital is thus dependent upon the clinician's informed decision, including close liaison with the adult family members or care-takers to ensure proper treatment of the patient when at home.

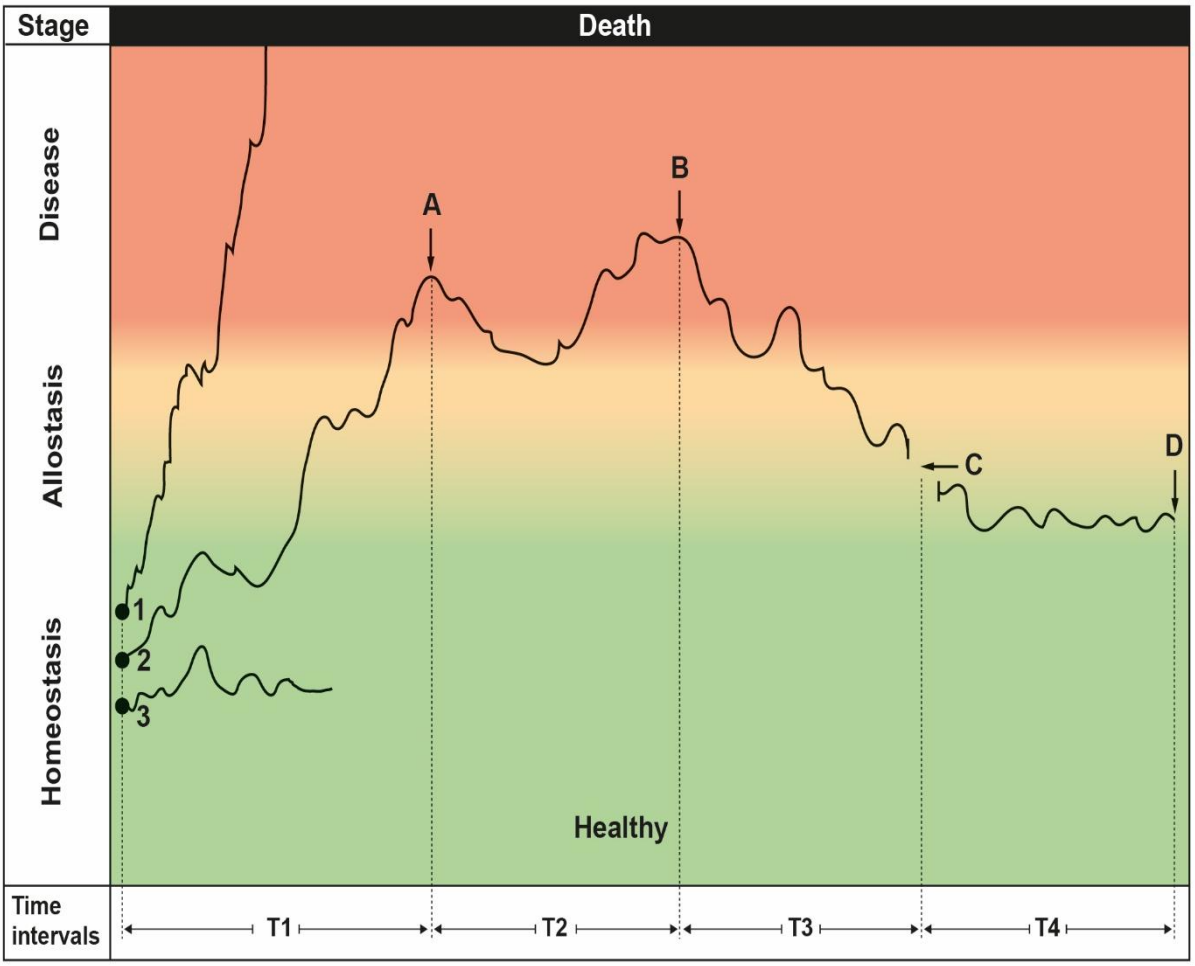


Figure 9-3: **Conceptual model illustrating the dynamic processes and the fine distinction between healthy and diseased states. The progress, or its lack, of TBM can generally be divided into four time intervals (T1–T4).**

T1: From the point of Mtb infection (beginning of T1), a TBM patient may not be able to contain the infection — severe onset with no treatment can soon lead to death (case 1). At the other extreme, self-containment can lead to latent infection (case 3). The time period here can vary between days, weeks or years in the latent scenario. The typical TBM patient examined in my investigation followed the path of case 2 — the patient exhibited symptoms of a nonspecific infection until becoming sick enough to warrant the parents taking the child to the local clinic (event A on the graph).

T2: Owing to nonspecific clinical symptoms, the patient is given adjunctive treatment (e.g., dietary interventions) and unprescribed medication (e.g., an analgesic such as paracetamol) aimed at stabilizing the child. In a typical TBM case such general treatment is not specific enough and, although adjunctive treatment might temporarily help the patient, the disease persists, eventually leading to hospitalization (event B).

T3: This is the time interval during which CSF and urine samples were collected and used in my PhD investigation (see discussion).

T4: Treatment continues at home for up to 6 months in order to eradicate the Mtb infection completely.

9.4 Addressing the third aim of the thesis

Aim 3: to apply the novel urinary metabolomics information towards developing a putative biosignature for the noninvasive diagnosis and monitoring of TBM in infants and children from the population group being studied.

This aim follows from aim 2, in that we attempted to take the new biological information gained from the perturbed urinary profile seen in TBM and determined if it had any diagnostic or prognostic capacity. It is here that we see the essential need for multidisciplinary expertise. First, each important variable identified requires to be placed in a biological context, in that we need to explain its biological origin and consequences — by invoking its biochemistry. Second, we need to have cross-validated markers of importance that can indicate selectivity and sensitivity — this requires expert statistical analysis. Finally, the proposed biosignature needs to have clinical significance and applicability in order to aid the medical expert in diagnosis and monitoring of TBM — the interface with the medical sciences.

9.4.1 Identifying a putative urinary biosignature

Following the inspection and biological interpretation of the perturbed metabolites seen in urine taken from TBM cases using GC-MS metabolomics (Chapter 8), we were able to identify several potential candidates as urinary markers of the disease state. Based upon sophisticated statistical analyses, a biosignature was isolated, consisting of 6 metabolites (methylcitric, isocitric, 2-ketoglutaric, quinolinic, 4-hydroxyhippuric, and kynurenic acids). The numerical sum of the concentrations of these metabolites — designated as SUM-6 — was shown to be a potential discriminator of the disease state. In order to assess how effective this biosignature was for classifying TBM, we used two additional experimental groups: the same TBM cases under investigation after 2 weeks of intensive TB-specific treatment, and middle-aged females suffering from fibromyalgia syndrome (FMS), a disorder unrelated to TBM. Taken together, 96.6% and 94.9% of the TBM patient groups, on admission to hospital and after treatment, respectively, were correctly classified for TBM. Moreover, after two weeks the patients were clinically stable but still had TBM, most likely indicating they were in a state of allostasis (see 9.3.2). The FMS group had a classification of 64.2%, suggesting that the metabolic aberration expressed by the SUM-6 discriminator did not prevail in FMS — but the SUM-6 value correctly predicted the presence of TBM. Further refinement of the putative urinary biosignature led to SUM-4, comprising the numerical sum of the concentrations of methylcitric, 2-ketoglutaric, quinolinic and 4-hydroxyhippuric acids. Using ROC analysis, SUM-4 gave an excellent diagnostic value (AUC > 90%) for the TBM group

on admission to hospital relative to the non-TBM patients and achieved high (>90%) sensitivity for a lower cost in specificity (a drop of 10 vs 14%).

It is also important to note that 5 non-annotated variables were initially identified as being statistically significant and may have influenced the putative potential of SUM-4 as a biosignature. This aspect will briefly be discussed in the final chapter (Chapter 10).

In summary, the KEMREP method confirms that the metabolomics data generated by our GC-MS protocol are reliable. Thus, the declaration of a urinary biosignature of TBM can be made with some confidence, yet it remains putative as it pertains only to the population under study. At this point it is important to note that a clear, unique CSF biomarker has yet to be defined for TBM. The collection of CSF is invasive, and presents a particularly high risk to infants. Herein lies the principal significance of this entire study. Based on our study group, it is possible to diagnose TBM through the metabolomics analysis of urine, a noninvasive sample collection method. This study thus opens the door for further research and validation studies, potentially leading to the development of a medical algorithm to diagnose TBM based on urine.

CHAPTER 10 LIMITATIONS AND FUTURE PROSPECTS

As is usually the case with metabolomics studies, especially with regard to diseases under investigation that have a low occurrence, there are several limitations, which provide topics for future studies. Three such issues emanated from the present study: (1) the collection of samples; (2) precise structural identification of important unknowns; and (3) the importance of chirality.

(1) **Sample collection**

The issue of samples was addressed early in this thesis (see 1.5 and 2.3), and thereafter alluded to several contexts. It is worth noting that the first two years of my PhD study involved sample collection. Approximately two new cases of TBM were recorded each month at Tygerberg Hospital, a relatively high rate compared to other hospitals, not only in South Africa but especially high with regard to resource rich countries. Experimental work therefore began only after the two years of sample collection, which yielded 33 cases of TBM, and 30 cases for the non-TBM control group. Now, a larger sample base (of over 100 samples) of the experimental group being studied would have been ideal as then the statistical results obtained would have been more significant and a subsection of the sample base (e.g., 10%) could have been used as a test group for more extensive cross-validation purposes. Thus, a study that afforded a longer sample collection period would be an improvement. Nevertheless, the present study is the largest and most comprehensive metabolomics study specifically on TBM in the literature to date.

Ideally, additional experimental groups, such as on viral and bacterial meningitis, would have been particularly beneficial in order to test the specificity of our putative urinary biomarkers. As TB is endemic in the environment of Tygerberg and a highly efficient paediatric group specialized in TBM is established at the hospital there, any nearby paediatric case suspected of harbouring TBM would have been directed to Tygerberg Hospital. Hence, sample collection for this PhD study was focused on TBM. On occasion, a case suspected of TBM was sent to the hospital but was later confirmed as viral or bacterial meningitis. The number of such cases was insufficient to provide statistically significant comparison. Also, the types of samples analysed in this PhD study involved only CSF and urine. The inclusion of blood samples would have provided a greater scope of the overall metabolome.

Hence, continued sample collection of sufficient quality, type, volume and description at Tygerberg Hospital over a prolonged period, following this PhD study, should prove especially useful in providing a larger sample base of TBM cases. Additionally, the collection of samples from viral or bacterial meningitis cases, at Tygerberg or at any other hospital, should also prove beneficial, especially with regard to follow-up studies on the putative biosignature and verification of the astrocyte–microglia lactate shuttle.

(2) **Unknown and non-annotated variables**

Over the course of this PhD study the focus was on known metabolites, mostly host response metabolites, annotated as such from spectral libraries. The detailed clinical information collected by the paediatricians upon the admission of patients to hospital and the corresponding creation of pure compound libraries of medications allowed the identification and isolation of spectral data, of both the NMR and GC-MS kind. There still existed, however, numerous unknown spectral peaks and, in the case of GC-MS, peaks of questionable origin and chemical reliability. These peaks were excluded from our final statistical analyses as no biological interpretation could be given to them. There undoubtedly exist numerous important variables in the urine and CSF metabolomes that have yet to be identified. The process of identifying such variables is time and labour intensive and is a separate research initiative all on its own, worthy of multiple PhD projects that analyse the complete metabolome.

In the process of investigating the GC-MS urine data reported in Chapter 8 there arose five important unknown variables during the preliminary statistical analyses. Tentative chemical structures to these unknowns could be postulated, but little to no information on these particular variables exists in the literature. Hence, it is not yet possible to identify their chemical origin with confidence, nor enough information to postulate a biological interpretation. Some of these variables hold strong statistical significance and are likely to be urinary biomarkers of Mtb. At least one of these variables shows high spectral intensities in TBM cases and is completely absent in cases not related to TB. Future in-depth chemical characterization of these entities is proposed in Chapter 8 (Mason *et al.* 2016c) and would provide a fruitful topic for a post-doctoral study. In fact, this is something that I personally would seriously consider undertaking.

(3) **Chiralty**

One of the clear and defining discoveries from the untargeted NMR metabolomics study was notably elevated lactic acid in the CSF of confirmed TBM cases. An interesting and

important question in this regard is: “What is the biological origin of this lactic acid?” This question can be answered by addressing the chirality of lactic acid, which occurs as two, optically active stereoisomers (enantiomers) — L-lactic acid and D-lactic acid. In humans, the L enantiomer is considered the normal, physiological form of lactic acid, whereas the D enantiomer typically originates from bacteria. NMR spectroscopy is incapable of differentiating the enantiomers of lactic acid. Thus, this question of chirality opens the door for future research.

Going beyond our initial aims and laying the foundation for important future studies, a follow-up study was conducted on the same CSF samples using targeted ultra-performance liquid chromatography–electrospray ionization–tandem mass spectrometry (UPLC-ESI-MS/MS). This follow-up study (end of Chapter 10 – Mason *et al.* 2016b) fittingly incorporates a unique metabolomics technique into this PhD study — completing the metabolomics journey from completely untargeted (NMR), through semi-targeted (GC-MS), and ending in the precise and targeted differentiation and quantification of a single metabolite (UPLC-ESI-MS/MS).

The results of this follow-up study that we pursued yielded pertinent information that not only supports our conceptual AMLS model but also opens the door for research beyond the aim of the present study, encapsulated in the final paragraph of Mason *et al.* (2016b):

‘*In summary, we have shown that the contribution from Mtb in the form of D-lactic acid in all our TBM-infected CSF cases was not demonstrable. The lactic acid consisted solely of the L-form, with the host being responsible for its high concentrations in the CSF. Beyond this study, the ramifications of determining the enantiomers — and by doing so the origin — of particular metabolites holds importance for future research. Knowledge of what is host produced and what is of microbial origin could revolutionize treatment regimes and allow us to understand disease pathogenesis and progress better.’*

10.1

CEREBROSPINAL FLUID IN TUBERCULOUS MENINGITIS EXHIBITS ONLY THE L-ENANTIOMER OF LACTIC ACID

Shayne Mason[1], Carolus J. Reinecke[1], Willem Kulik[2], Arno van Cruchten[2], Regan Solomons[3] and A. Marceline Tutu van Furth[4]

[1]Centre for Human Metabolomics, Faculty of Natural Sciences, Private Bag X6001, North-West University (Potchefstroom Campus), South Africa,
[2]Laboratory Genetic Metabolic Diseases, Department of Clinical Chemistry, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands,
[3]Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 19063, Tygerberg 7505, South Africa,
[4]Department of Paediatric Infectious Diseases–Immunology and Rheumatology, Vrije Universiteit Medical Centre, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands.

Keywords: cerebrospinal fluid (CSF), enantiomers, L and D-lactic acid, tuberculous meningitis (TBM), ultra-performance liquid chromatography–electrospray ionization–tandem mass spectrometry (UPLC-ESI-MS/MS).

Accepted: *BMC Infectious Diseases,* 2016 – in press.

10.1.1 Abstract

Background: The defining feature of the cerebrospinal fluid (CSF) collected from infants and children with tuberculous meningitis (TBM), derived from an earlier untargeted nuclear magnetic resonance (NMR) metabolomics study, was highly elevated lactic acid. Undetermined was the contribution from host response (L-lactic acid) or of microbial origin (D-lactic acid), which was set out to be determined in this study.

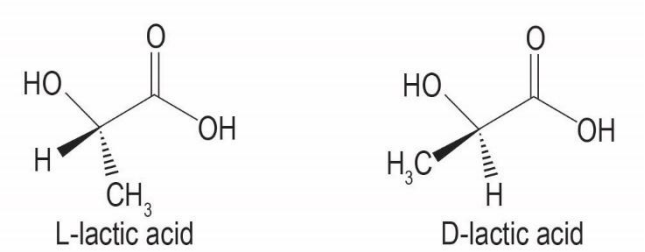
Methods: In this follow-up study, we used targeted ultra-performance liquid chromatography–electrospray ionization–tandem mass spectrometry (UPLC-ESI-MS/MS) to determine the ratio of the L and D enantiomers of lactic acid in these CSF samples.

Results: Here we report for the first time that the lactic acid observed in the CSF of confirmed TBM cases was in the L-form and solely a response from the host to the infection, with no contribution from any bacteria. The significance of elevated lactic acid in TBM appears to be that it is a crucial energy substrate, used preferentially over glucose by microglia, and exhibits neuroprotective capabilities.

Conclusion: These results provide experimental evidence to support our conceptual astrocyte–microglia lactate shuttle model formulated from our previous NMR-based metabolomics study — highlighting the fact that lactic acid plays an important role in neuroinflammatory diseases such as TBM. Furthermore, this study reinforces our belief that the determination of enantiomers of metabolites corresponding to infectious diseases is of critical importance in substantiating the clinical significance of disease markers.

10.1.2 Background

Lactic acid is a common metabolite and occurs as two, optically active stereoisomers (enantiomers) — L-lactic acid and D-lactic acid.



In humans, the L enantiomer is considered the normal, physiological form of lactic acid. The endogenous D form is also found in humans, but in nanomolar amounts due to methylglyoxal metabolism (Ewaschuk *et al.* 2005), which converts acetone derivatives to glutathione. The D enantiomer typically originates from gut microbiota, such that it can be detected in the blood and urine up to the millimolar range. Elevation of D-lactic acid typically occurs as a result of gut trauma or gastrointestinal disorder, for example, in jejuno-ileal bypass for obesity (Stolberg *et al.* 1982), appendicitis (Çaglayan *et al.* 2003; Duzgun *et al.* 2006), inflammatory bowel syndrome (Scheijen *et al.* 2012), and even in patients with chronic fatigue syndrome (Sheedy *et al.* 2009) and type 2 diabetes (Scheijen *et al.* 2012). The main source of elevated D-lactic acid in urine, however, is in patients with short bowel syndrome (Kowlgi & Chhabra 2015).

It is known that the gut possesses its own unique microbiotic environment; it is also generally believed that the brain is a fairly sterile environment without its own microbiota, with the microglia in the circulating cerebrospinal fluid (CSF) actively eliminating any microbial incursions beyond the blood–brain barrier (BBB). However, in pathological conditions where an invading pathogen successfully transgresses the BBB and an infection occurs in the central nervous system (CNS), any subsequent metabolic perturbations may be difficult to trace and interpret. In the case of bacterial meningitis, in which bacteria infect the meninges, with disastrous consequences due to neuroinflammation, there have been numerous reports of highly acidotic CSF owing to the presence of lactic acid (Abro *et al.* 2009; Genton & Berger 1990; Huy *et al.* 2010). A unique form of bacterial meningitis, caused by *Mycobacterium tuberculosis* (Mtb), is tuberculous meningitis (TBM). What makes this form of meningitis unique, and dangerous, is that it has a slow, insidious onset which often leads to accumulated damage due to neuroinflammation, resulting in irreparable neural damage, or even death. It is a disease that is difficult to diagnose in its early stages and fatal if recognized at a late stage. Of the various CSF diagnostic markers for TBM, lactic acid is one that has been reported in a limited number of studies (Thwaites *et al.* 2002; Thwaites *et al.* 2004; Torok *et al.* 2008). In a previous study we examined a cohort of infants and children with TBM by means of an untargeted ^{1}H nuclear magnetic resonance (NMR)-based metabolomics approach (Mason *et al.* 2015). The most noticeable feature that we detected was highly elevated lactic acid (7.36 ± 2.36 mM) in the CSF of these cases compared to the normal, age-related reference range (1.65 ± 0.63 mM) (Leen *et al.* 2012). The presence of highly elevated lactic acid in CSF is not unique to meningitis cases: in patients with cerebral malaria there was significantly higher CSF lactic acid recorded (9.0 ± 5.3 mM (White *et al.* 1985); 6.0 ± 1.0 mM (Medana *et al.* 2002)) in those who died compared to survivors. Yao *et al.* reported that CSF lactic acid levels reflect the severity of metabolic impairment of the

brain in patients with hepatic encephalopathy. Thus, knowledge of increased CSF lactic acid levels in response to neuropathology is not new; however, distinguishing the respective roles of the L and D forms has not been attempted hitherto.

In our NMR-based metabolomics study (Mason *et al.* 2015) the high levels of lactic acid, together with several other statistically significant metabolites, led to the formulation of the hypothesis: 'The host's response to neural infection results in an "astrocyte–microglia lactic acid shuttle" (AMLS) that operates in neuroinflammatory diseases, such as TBM'; thereafter, a conceptual model describing the AMLS was constructed. We are therefore seeing that lactic acid, previously considered an unimportant by-product of CNS metabolism, is now receiving more attention as it appears to play an essential role in normal neural homeostasis — in the astrocyte–neuron lactic acid shuttle (ANLS) (Pellerin & Magistretti 2011) — and possibly has an important role in neuropathological states. For example, lactic acid is neuroprotective in cerebral ischemia (Castillo *et al.* 2015), a condition often associated with TBM.

In the present investigation, we used the CSF samples collected in our previous study and employed a derivatization method adapted from Scheijen *et al.* to differentiate the enantiomers of lactic acid present. We analysed the samples using the highly sensitive, targeted method of ultra-performance liquid chromatography–electrospray ionization–tandem mass spectrometry (UPLC-ESI-MS/MS). We report here for the first time that highly lactic acidotic CSF from infants and children with confirmed TBM exhibits only the L-enantiomer – hence it is a response solely by the host to the infection — and then discuss the relevance of the phenomenon of elevated lactic acid in this example of a neuroinflammatory disease.

10.1.3 Methods

10.1.3.1 Sampling

The experimental group consisted of infants and children (<13 years of age) (n = 20) from the Western Cape region of South Africa, who were directed from local clinics to the paediatric unit of Tygerberg Hospital, Cape Town, on suspicion of meningitis, based on clinical symptoms. Upon admission to hospital, a lumbar CSF sample was taken for differential diagnosis, a portion of which was stored at –80°C for this study, which was used to confirm a diagnosis of TBM. A diagnosis of TBM was based on the procedure of the uniform research case definition of Marais *et al.* (2010), practiced in our clinical setting. Only

children with ‘definite’ and ‘probable’ TBM were included in the present patient group. TBM was classified as ‘definite’ when CSF demonstrated acid-fast bacilli on microscopy, a positive Mtb culture and/or passed a positive CSF Mtb commercial nucleic acid amplification test in a child with symptoms or signs suggestive of the disease. TBM was classified as ‘probable’ according to a scoring system based on clinical, cerebrospinal fluid and neuroimaging criteria, as well as evidence of extraneural TB. Clinical details of the patients are described in the supplementary information of Mason *et al.* (2015) — Annexure 1. For the purpose of the present study the stage of TBM, as well as the glucose concentration in 16 out of the 20 CSF samples, were made available, as shown in Figure 10-2. We also collected a urine sample from our subjects upon admission to hospital. Urine is often the biofluid of choice for investigating the various potential sources of lactic acid, as well as providing for a non-invasive mode of sample collection. A limitation in using urine remains the unpredictable fluctuation in the concentration of targeted metabolites linked to the disease state of the patients. Informed consent was obtained from all participants and this study was approved by the Human Research Ethics Committee of Stellenbosch University, South Africa (study no. N11/01/006).

10.1.3.2 Chemicals

The chemicals used as standards were: sodium L-lactic acid ≥99.0% (Sigma-Aldrich 71718, CAS: 867-56-1); sodium D-lactic acid ≥99.0% (Sigma-Aldrich 71716, CAS: 920-49-0); sodium L-lactic acid-3,3,3-d3 (CDN isotopes D-2646, CAS: 79-33-4).

The following chemicals were used in sample preparation and analyses: (+)-O,O’-Diacetyl-L-tartaric anhydride (DATAN) (Sigma-Aldrich 358924, CAS: 6283-74-5); acetonitrile HPLC supragrade (Biosolve 01203502, CAS: 75-05-8); MilliQ water (Millipore, CAS: 7732-18-5); acetic acid 100% (Merck 1.00063.1000, CAS: 64-19-7); ammonium formate (Sigma-Aldrich 25204, CAS: 540-69-2); dichloromethane (Lab-Scan AR1040A, CAS: 75-09-2); perchloric acid (Sigma-Aldrich 244252, CAS: 7601-90-3); formic acid 98-100% (Merck 1.00264.1000, CAS: 64-18-6).

10.1.3.3 Sample preparation and UPLC-ESI-MS/MS analysis

A dilution series of L and D-lactic acid and a 2.5 mM L-lactic acid-d_3 internal standard (IS) solution were prepared in advance for the calibration curve. A fresh diacetyl-L-tartaric anhydride (DATAN) derivatization solution was prepared by dissolving 250 mg of DATAN in 4 ml dichloromethane and 1 ml acetic acid. Samples were prepared by combining 50 µL CSF / 100 µL urine with 20 µL internal standard solution and 300 µL acetonitrile (ACN) in an

Eppendorf tube. Samples were vortexed and centrifuged for 10 minutes at 3000 RPM in order to remove proteins and other macromolecules. The supernatant was transferred to a clean glass vial and dried under nitrogen gas at 40 °C. To this, 100 µL DATAN solution was added, vortexed and incubated at 75 °C for 30 minutes. Once again the solution was dried under nitrogen gas at 40 °C and then reconstituted in 200 µL ACN / H_2O (1/2 v / v) for UPLC-ESI-MS/MS analysis. For the UPLC: eluent A (2.5 mM ammonium formate, pH = 3.6) and eluent B (100% acetonitrile) were used.

Samples were analysed on a Waters Acquity UPLC hyphenated to a Quattro premier XE triple quadrupole mass spectrometer using negative ion electrospray ionization (ESI) — details given in Annexure 4. After derivatization with DATAN, L- and D-lactic acid were separated from each other on the UPLC column and measured by MS/MS in MRM mode. Use of solvent delay time avoided the introduction of salts in the MS. Results were quantified in terms of the stable isotope-labelled analogue of L-lactic acid. The peak areas corresponding to component-specific transitions were converted via calibration lines to concentrations.

10.1.4 Results

Using a spiked sample containing both the L and D forms of lactic acid, and comparing the chromatogram qualitatively with the stable isotope (L-lactic acid-d_3), it was evident that the UPLC-ESI-MS/MS method used in this study was able to differentiate between the two enantiomers (Figures 10-1A and 10-2B). We were therefore able to identify and quantify the lactic acid content of the CSF samples from the 20 TBM cases, as described in the 'Methods' section. These results revealed that the lactic acid was exclusively in the L-form — no D-isomer was detected (Figure 10-3C) — with a mean concentration of 5.2 mM (SD = 2.6; range: 1.5–10.5 mM).

We also compared the total lactic acid concentrations in the corresponding samples with those derived from the UPLC-ESI-MS/MS method used in the current study and with those from the NMR method of Mason *et al.* (2015), and found a strong linear relationship between the two sets of results (r = 0.93), with a statistically significant positive correlation (0.86) and a statistically strong validation of the fit with an R^2 value of 0.73. These data therefore confirm that both methods are highly comparable, demonstrating high reproducibility.

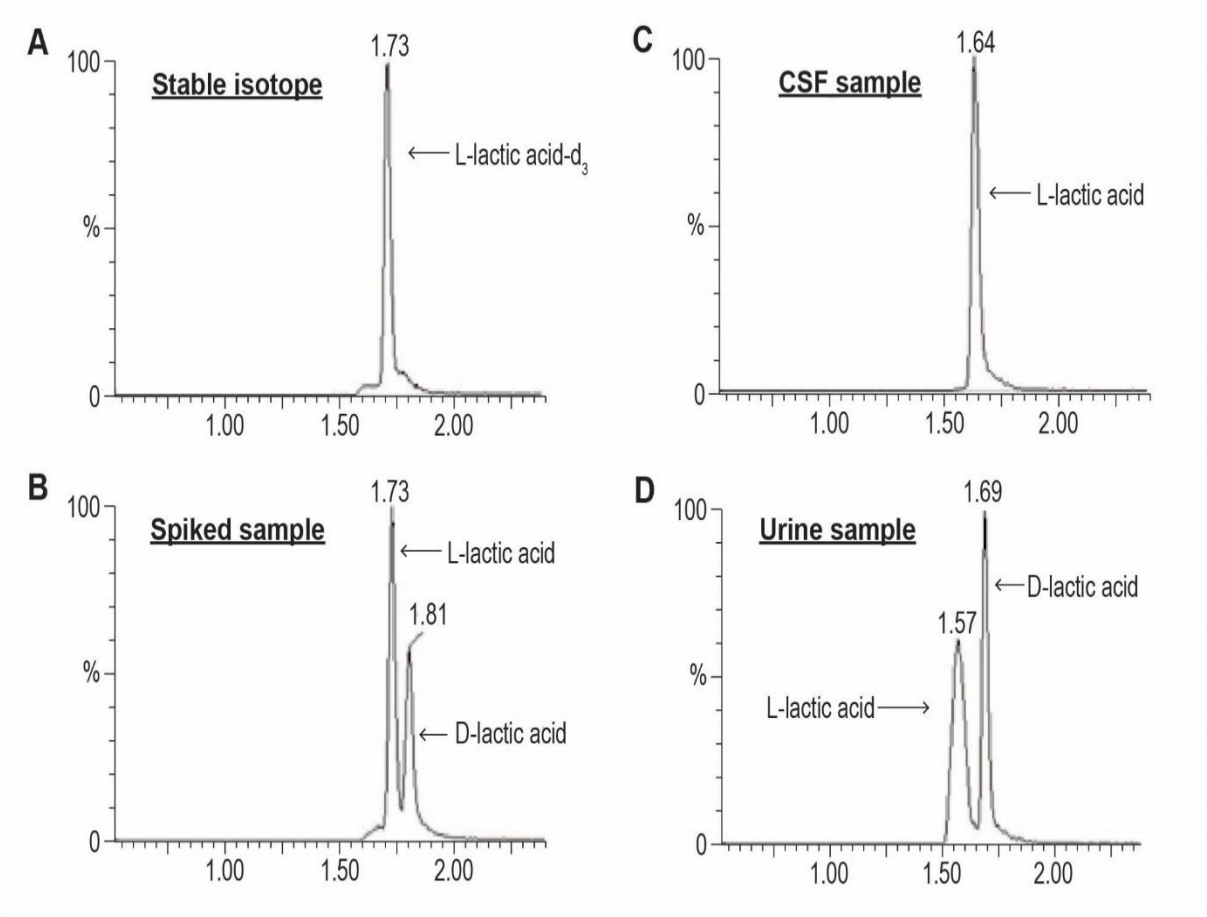


Figure 10-1: **Representative chromatograms depicting (A) definitive identification of L-lactic acid using the stable isotope (L-lactic acid-d_3); (B) clear differentiation of L and D forms of lactic acid in the spiked sample; (C) in CSF, complete lack of D form of lactic acid with only the L form present; and (D) presence of both L and D forms of lactic acid in urine.**

Of the 20 CSF samples examined, a clinician reported glucose values for 16 samples. Figure 10-2 is a scatterplot, showing both the lactic acid and glucose, together with the patient severity outcome after treatment. Additional statistics (shown in Annexure 4) support existing knowledge that glucose is significantly correlated with TBM stage and outcome severity. While L-lactic acid reached high concentrations in CSF samples in some patients from the present group (well above the reference ranges), the L-lactic acid did not correlate with statistical significance with the TBM stage or outcome severity – an observation that may, however, relate to the small sample size of the present group.

In addition to the CSF, the urine was examined of 7 patients upon admission to hospital. Both the L and D forms of lactic acid were detectable (illustrated in Figure 10-1D), but in small amounts. The mean urine lactic acid was determined to be 98.4 µM for the L form (SD = 37.4; range: <50–155 µM) and 47.4 µM for the D form (SD = 56.7; range: nd–148 µM). The contribution of D-lactic acid from the gut microbiota cannot be distinguished from any possible influence of the Mtb, and the small amount of urinary lactic acid was well within the normal range. We conclude that urine is not a suitable biofluid for monitoring and assessing lactic acid for diagnostic purposes in TBM.

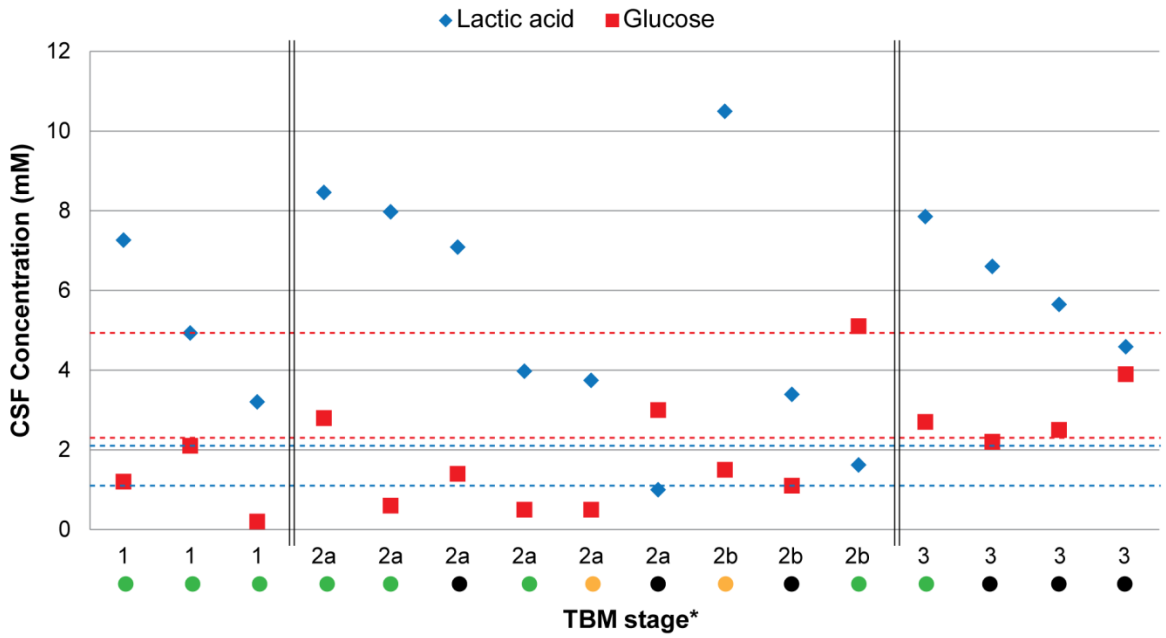


Figure 10-2: **Scatterplot showing the concentration of lactic acid in the CSF samples and the corresponding CSF glucose (values not reported in text) over different stages of TBM disease (dashed lines indicate respective reference ranges).**

Outcome severity: ● = normal; ● = mild neurological problems (e.g., learning difficulties, visual impairment); ● = severe neurological problems (e.g., partial paralysis, severe motor impairment, cranial nerve palsy).

***TBM stage: 1 = Glasgow coma score (GCS) 15 and no focal neurology; 2a = GCS 15 plus focal neurology; 2b = GCS 11–14 with focal neurology; 3 = GCS <11.**

10.1.5 Discussion

In the present investigation we identified and quantified the L and D enantiomers of lactic acid in the CSF of 20 cases of infants and children with confirmed TBM, using UPLC-ESI-MS/MS. This study was motivated by our NMR-based metabolomics study (Mason *et al.* 2015), in which the most prominent feature of the CSF of these TBM cases was highly elevated lactic acid. We therefore sought to determine the source and form of the lactic acid — was it a response from the host (which would generate the L-isomer), of microbial origin (which produces the D-isomer), or both? Because NMR is incapable of distinguishing the enantiomers of lactic acid, we conducted this follow-up study with the means to identify the separate isomers.

We have shown that both enantiomers were detectable in urine samples, but the elevated lactic acid in the CSF samples of the TBM patients consisted of only L-lactic acid, and hence was produced exclusively by the host with no contribution from the Mtb. Moreover, Figure 10–2 shows that as the TBM disease progressed into later stages there was a drop in patient prognosis (i.e., increased outcome severity). Generally, in more advanced TBM stages, fewer energy substrates are available for microglia activity and decreased protection of neurons from the lactic acid. During this period, irreversible neurological damage can begin to occur. Lactic acid boosts energy production in the brain and increases neuroprotection, thus potentially improving the prognosis of patients, especially those at high risk. TBM patients in late stages of the disease who exhibit CSF with low lactic acid and glucose may be beyond help as the accumulated neural damage and lack of energy substrate can be fatal.

The origin and rationale of high levels of lactic acid produced by the host in response to a neuroinflammatory disease such as TBM is unclear. Interconversion of lactic acid and pyruvate occurs via lactate dehydrogenase, with increased lactic acid typically being associated with anaerobic respiration. Thus, elevated levels of lactic acid in the CSF due to neuroinflammation could be due to hypoxia caused by ischemia, or by increased glucose levels and hence increased flow through the glycolysis pathway. However, in TBM cases there are typically low levels of glucose in the CSF (Thwaites *et al.* 2009), corresponding to a CSF to serum glucose ratio less than 0.5 or an absolute CSF glucose concentration of less than 2.2 mM. Furthermore, several studies have shown no correlation between CSF lactic acid levels and cerebral blood flow (i.e., they are unrelated to ischemia) (Brodersen & Jørgensen 1974; DeSalles *et al.* 1986; DeSalles *et al.* 1987). Thus, elevated lactic acid in CSF of TBM cases is unlikely to be due to anaerobic respiration but instead is possibly a product of temporarily increased flux in the glycolysis pathway.

The results reported here also provide experimental evidence which supports our proposed AMLS hypothesis (Mason *et al.* 2015). The hypothesis postulates that, when the brain is in crisis due to neuroinflammatory-inducing infection, energy flow in brain metabolism is shifted away from the neurons and shunted towards the microglia. Hence, in neuroinflammatory infectious diseases, such as TBM, lactic acid produced by glycolysis in astrocytes participates in the activated immune response and, in association with ketones and gluconeogenic amino acids, is collectively directed from the neurons preferentially into microglia where it enters the mitochondrial citric acid cycle. This process contributes to oxidative phosphorylation and hence produces high levels of ATP and forms of reactive oxygen species, such as hydrogen peroxide, required for degradation of the invading pathogen.

Further studies are needed to advance our understanding of the dynamics involved in this lactic acid phenomenon — for example: determining the rate at which microglia produce lactic acid under neuroinflammatory conditions such as TBM; an *in vitro* scenario testing whether a high level of lactic acid within the microglia reduces the uptake of interstitial lactic acid due to concentration gradients, accounting for the high concentrations of lactic acid in the CSF; more in-depth studies into transporters that shuttle lactate, namely, monocarboxylate transporters; and, in a clinical setting, determining the impact of adjunctive treatments involving direct infusion of sodium lactate into the CSF of advanced stage TBM patients.

While simple in design and execution, this study provides important information not reported before. The implications of these results are compelling in that the levels of lactic acid in the CSF, produced by the host, should be carefully considered by the clinician, especially regarding neuroinflammatory diseases. The reason for this is that lactic acid is used preferentially over glucose in such cases (Schurr *et al.* 1988; Schurr *et al.* 1997a; Schurr *et al.* 1997b; Schurr *et al.* 1997c; Schurr *et al.* 1999), provides a boost in neuroprotection (Cater *et al.* 2001; Cater *et al.* 2003), and aids microglia energy demands in their bactericidal actions. Thus lactic acid has a potential beneficial role in the clinical management of neurological disorders (Taher *et al.* 2015).

In summary, we have shown that the contribution from Mtb in the form of D-lactic acid in all our TBM-infected CSF cases was not demonstrable. The lactic acid consisted solely of the L-form, with the host being responsible for its high concentrations in the CSF. Beyond this study, the ramifications of determining the enantiomers — and by doing so the origin — of particular metabolites holds importance for future research. Knowledge of what is produced by the host and what is of microbial origin could revolutionize treatment regimes and allow

us to understand disease pathogenesis and progress better. Therefore adaptation of the method presented should be made routine for common, optically active markers of disease to determine their respective enantiomers, particularly in the field of metabolomics.

Competing interests: The authors confirm that there are no competing interests in this manuscript.

Authors' contributions: RS is the paediatrician who collected the samples and clinical information, and treated the patients. WK and AvC supervised and conducted the experimental work. CJR and AMvF are the supervisors of SM and directed and contributed to the paper. SM analysed the results and wrote the manuscript. All authors contributed in the revision of the manuscript and approved the final version.

Acknowledgements: S.M. and R.S. are recipients of a Desmond Tutu-NRF-VU doctoral fellowship for a joint PhD study between the Vrije Universiteit in Amsterdam, the Netherlands and North-West University (Potchefstroom), and the University of Stellenbosch, South Africa, respectively. Research funding for this project is provided by the Technological Innovation Agency of the Department of Science and Technology of South Africa.

Supplementary information: see Annexure 4.

REFERENCES

Abbott, N.J., Rönnbäck, L. & Hansson, E. (2006). Astrocyte–endothelial interactions at the blood–brain barrier. *Nature Reviews – Neuroscience*, 7:41–53.

Abro, A.H., Abdou, A.S., Ustadi, A.M., Saleh, A.A., Younis, J. & Doleh, W.F. (2009). CSF lactate level: a useful diagnostic tool to differentiate acute bacterial and viral meningitis. *Journal of Pakistan Medical Association*, 59:508–511.

Agren, H. & Niklasson, F. (1988). Creatinine and creatine in CSF: Indices of brain energy metabolism in depression. *Journal of Neural Transmission*, 74:55–59.

Alavinia, S.M., Khakshour, A., Habibi, G., Navabi, B., Mostafavi, S.A. & Moghadam, M.S. (2013). An overview and mapping of childhood tuberculosis: prevalence, scientific production and citation analysis. *Indian Journal of Tuberculosis,* 60:28–36.

Álvarez-Sánchez, B., Priego-Capote, F. & de Castro, M.D. (2010). Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation. *Trends in Analytical Chemistry,* 29(2):111–119.

Anderson, N.E., Somaratne, J., Mason, D.F., Holland, D. & Thomas, M.G. (2010). A review of tuberculous meningitis at Auckland City Hospital, New Zealand. *Journal of Clinical Neuroscience,* 17:1018–1022.

Andronikou, S., Smith, B., Hatherhill, M., Douis, H. & Wilmshurst, J. (2004). Definitive neuroradiological diagnostic features of tuberculous meningitis in children. *Pediatric Radiology*, 34:876–885.

Asano, K., Miyamoto, I., Matsushita, T., Murakami, Y., Minoura, S., Wagatsuma, T. & Oshima, M. (1988). Succinic acidemia: a new syndrome of organic acidemia associated with congenital lactic acidosis and decreased NADH-cytochrome c reductase activity. *Clinica Chimica Acta,* 173:305–312.

Asano, T., Ichiki, K., Koizumi, S., Kaizu, K., Hatori, T., Fujino, O., Mashiko, K., Sakamoto, Y., Miyasho, T. & Fukunaga, Y. (2010). IL-17 is elevated in cerebrospinal fluids in bacterial meningitis. *Cytokine,* 51:101–106.

Atkinson, A.J., Colburn, W.A., DeGruttola, V.G., DeMets, D.L., Downing, G.J., Hoth, D.F., Oates, J.A., Peck, C.C., Schooley, R.T., Spilker, B.A., Woodcock, J. & Zeger, S.L. (2001).

Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, 69(3):89–95.

Badenhorst, C.P.S., Erasmus, E., Van der Sluis, R., Nortje, C. & Van Dijk, A.A. (2014). A new perspective on the importance of glycine conjugation in the metabolism of aromatic acids. *Drug Metabolism Reviews,* 46(3):343–361.

Barker, M. & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, 17(3):166–173.

Baughn, A.D., Garforth, S.J., Vilchèze, C. & Jacobs Jr, W.R. (2009). An anaerobic-type α-ketoglutarate ferredoxin oxidoreductase completes the oxidative tricarboxylic acid cycle of Mycobacterium tuberculosis. *PLoS Pathogens*, 5(11):e1000662.

Bélanger, M., Allaman, I. & Magistretti, P.J. (2011). Brain energy metabolism: focus on astrocyte-neuron metabolic cooperation. *Cell metabolism*, 14(6):724–738.

Bender, A.S., Young, L.P. & Norenberg, M.D. (1997). Effect of lactic acid on L-glutamate uptake in cultured astrocytes: mechanistic considerations. *Brain research*, 750(1):59–66.

Benton, H.P., Want, E., Keun, H.C., Amberg, A., Plumb, R.S., Goldfain-Blanc, F., Walther, B., Reily, M.D., Lindon, J.C., Holmes, E., Nicholson, J.K. & Ebbels, T.M.D. (2012). Intra- and interlaboratory reproducibility of ultra performance liquid chromatography-time-of-flight mass spectrometry for urinary metabolic profiling. *Analytical Chemistry*, 84:2424–2432.

Beste, D.J.V., Hooper, T., Stewart, G., Bonde, B., Avignone-Rossa, C., Bushell, M.E., Wheeler, P., Klamt, S., Kierzek, A.M. & McFadden, J. (2007). GSMN-TB: a web-based genome-scale network model of mycobacterium tuberculosis metabolism. *Genome Biology,* 8(5):R89.

Beyers, N.B., Gie, R.P., Zietsman, H.L., Kunneke, M., Hauman, J., Tatley, M. & Donald, P.R. (1996). The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high-incidence community. *South African Medical Journal,* 86(1):40–44.

Bhigjee, A.I., Padayachee, R., Paruk, H., Hallwirth-Pillay, K.D., Marais, S. & Connoly, C. (2007). Diagnosis of tuberculous meningitis: clinical and laboratory parameters. *International Journal of Infectious Disease*, 11:248–254.

Bijlsma, S., Bobeldijk, I., Verheij, E.R., Ramaker, R., Kochnar, S., Macdonald, I.A., van Ommen, B. & Smilde, A.K. (2006). Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. *Analytical Chemistry*, 78:567–574.

Bittar, P.G., Charnay, Y., Pellerin, L., Bouras, C. & Magistretti, P.J. (1996). Selective distribution of lactate dehydrogenase isoenzymes in neurons and astrocytes of human brain. *Journal of Cerebral Blood Flow and Metabolism*, 16*:*1079–1089.

Blair, C., Raver, C.C., Granger, D., Mills-Koonce, R., Hibel, L. & The Family Life Project Key in Investigators (2011). Allostasis and allostatic load in the context of poverty in early childhood. *Development and Psychopathology,* 23:845–857.

Blau, N., Duran, M., Gibson, K.M., Dionisi-Vici, C. & Blaskovics, M.E. (2014). Physician's guide to the diagnosis, treatment, and follow-up of inherited metabolic diseases. Springer, 24–31.

Block, M.L. & Hong, J.S. (2005). Microglia and inflammation-mediated neurodegeneration: Multiple triggers with a common mechanism. *Progress in Neurobiology,* 76*:*77–98.

Blok, N., Visser, D.H., Solomons, R., van Elsland, S.L., Den Hartog, A.I. & van Furth, A.M. (2014). Lipoarabinomannan enzyme-linked immunosorbent assay for early diagnosis of childhood tuberculous meningitis. *The International Journal of Tuberculosis and Lung Disease*, 18:205–210.

Blumenthal, A., Isovski, F. & Rhee, K.Y. (2009). Tuberculosis and host metabolism: ancient associations, fresh insights. *Translational Research,* 154*:*7–14.

Blumenthal, A., Nagalingam, G., Huch, J.H., Walker, L., Guillemin, G.J. & Smythe, G.A. (2012). *M. tuberculosis* induces potent activation of IDO-1, but this is not essential for the immunological control of infection. *PLoS ONE*, 7:e37314.

Boenzi, S., Martinelli, D., Carrozzo, R., Piemonte, F., DiCiommo, V., Rizzo, C., Bertini, E., Dionisi-Vici, C. (2011). Plasma creatine is elevated in mitochondrial disorders: a new biomarker for the diagnosis. *Journal of Inherited Metabolic Disease*, 34:S160.

Boss, E.A., Moolenaar, S.H., Massuger, L.F.A.G., Boonstra, H., Engelke, U.F.H., de Jong, J.G.N. & Wevers, R.A. (2000). High-resolution proton nuclear magnetic resonance spectroscopy of ovarian cyst fluid. *NMR in Biomedicine,* 13*:*297–305.

Bouatra, S., Aziat, F., Mandal, R., Guo, A.C., Wilson, M.R., Knox, C., Bjorndahl, T.C., Krishnamurthy, R., Saleem, F., Liu, P. & Dame, Z.T. (2013). The human urine metabolome. *PloS ONE*, 8(9):e73076.

Bouzier-Sore, A.K., Merle, M., Magistretti, P.J. & Pellerin, L. (2002). Feeding active neurons:(re) emergence of a nursing role for astrocytes. *Journal of Physiology-Paris*, 96(3):273–282.

Brantley, P.J., Dietz, L.S., McKnight, G.T., Jones, G.N. & Tulley, R. (1988). Convergence between the Daily Stress Inventory and endocrine measures of stress. *Journal of Consulting and Clinical Psychology,* 4:549–551.

Brereton, R.G. (2003). Chemometrics – data analysis for the laboratory and chemical plant. John Wiley & Sons Ltd, West Sussex, England.

British Medical Research Council (1948). Streptomycin treatment of tuberculous meningitis. *British Medical Journal, 1(6503)*:582–596.

Brodersen, P. & Jørgensen, E.O. (1974). Cerebral blood flow and oxygen uptake, and cerebrospinal fluid biochemistry in severe coma. *Journal of Neurology, Neurosurgery and Psychiatry,* 37:384–391.

Brooks, G.A. (1986). The lactate shuttle during exercise and recovery. *Medicine and Science in Sports and Exercise*, 18:360–368.

Brooks, G.A. (2002). Lactate shuttles in nature. *Biochemical Society Transactions*, 30:258–264.

Brown, M., Dunn, W.B., Ellis, D.I., Goodacre, R., Handl, J., Knowles, J.D., O'Hagan, S., Spasić, I. & Kell, D.B. (2005). A metabolome pipeline: from concept to data to knowledge. *Metabolomics*, 1(1):39–50.

Bruce, S.J., Tavazzi, I., Parlsod, V., Rezzi, S., Kochhar, S. & Guy, P.A. (2009). Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. *Analytical Chemistry*, 81:3285–3296.

Bundy, J.G., Willey, T.L., Castell, R.S., Ellar, D.J. & Brindle, K.M. (2005). Discrimination of pathogenic clinical isolates and laboratory strains of Bacillus cereus by NMR-based metabolomics profiling. *FEMS Microbiology Letters*, 242(1):127–136.

Çaglayan, F., Çakmak, M., Çaglayan, O. & Çavusoglu, T. (2003). Plasma D-lactate levels in diagnosis of appendicitis. *Journal of Investigative Surgery,* 16:233–237.

Castillo, S., Gopalacharyulu, P., Yetukuri, L. & Orešič, M. (2011). Algorithms and tools for the preprocessing of LC-MS metabolomics data. *Chemometrics and Intelligent Laboratory Systems*, 108:23–32.

Castillo, X., Rosafio, K., Wyss, M.T., Drandarov, K., Buck, A., Pellerin, L., Weber, B. & Hirt, L. (2015). A probable dual mode of action for both L-and D-lactate neuroprotection in cerebral ischemia. *Journal of Cerebral Blood Flow & Metabolism*, 35(10):1561–1569.

Cater, H.L., Benham, C.D. & Sundstrom, L.E. (2001). Neuroprotective role of monocarboxylate transport during glucose deprivation in slice cultures of rat hippocampus. *The Journal of physiology*, 531(2):459–466.

Cater, H.L., Chandratheva, A., Benham, C.D., Morrison, B. & Sundstrom, L.E. (2003). Lactate and glucose as energy substrates during, and after, oxygen deprivation in rat hippocampal acute and cultured slices. *Journal of neurochemistry*, 87(6):1381–1390.

Cecchini, D., Ambrosioni, J., Brezzo, C., Corti, M., Rybko, A., Perez, M., Poggi, S. & Ambroggi, M. (2007). Tuberculous meningitis in HIV-infected patients: drug susceptibility and clinical outcome. *AIDS,* 21:373–374.

CDC (Centers for disease control and prevention) (2013). Reported tuberculosis in the United States 2012. (URL: http://www.cdc.gov/tb/statistics/reports/2012/pdf/report2012.pdf).

Chen, Y. & Swanson, R.A. (2003). Astrocytes and brain injury. *Journal of Cerebral Blood Flow & Metabolism*, 23(2):137–149.

Chen, J., Meng, C.K., Narayan, S.B., Luan, W. & Bennett, M.J. (2009). The use of Deconvolution Reporting Software© and backflush improves the speed and accuracy of data processing for urinary organic acid analysis. *Clinica Chimica Acta*, 405(1):53–59.

Çetin, K., Erol, S., Aksoy, H. & Taşyaran, H.A. (2002). The relation of cerebrospinal fluid nitric oxide levels to prognosis and differential diagnosis of meningitis. *Turkish Journal of Medical Sciences,* 32:385–390.

Cherian, A. & Thomas, S.V. (2011). Central nervous system tuberculosis. *African Health Sciences*, 11:116–127.

Chiang, S.S., Khan, F.A., Milstein, M.B., Tolman, A.W., Benedetti, A., Starke, J.R. & Becerra, M.C. (2014). Treatment outcomes of childhood tuberculous meningitis: a systematic review and meta-analysis. *The Lancet Infectious Diseases,* 14:947–957.

Chong, I.G. & Jun, C.H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems,* 78:103–112.

Christen, S., Schaper, M., Lykkesfeldt, J., Siegenthaler, C., Bifrare, Y.D., Banic, S., Leib, S.L. & Tauber, M.G. (2001). Oxidative stress in brain during experimental bacterial meningitis: differential effects of α–phenyl-*tert*-butyl nitrone and *N*-acetylcysteine treatment. *Free Radical Biology & Medicine,* 31(6):754–762.

Chung, Y., Rider, L., Bell, J., Summers, R.M., Zemel, L.S., Rennebohm, R.M., Passo, M.H., Hicks, J., Miller, F.W. & Scott, D.L. (2005). Muscle metabolites, detected in urine by proton spectroscopy, correlate with disease damage in juvenile idiopathic inflammatory myopathies. *Arthritis Care & Research,* 53(4):565–570.

Coen, M., O'Sullivan, M., Bubb, W.A., Kuchel, P.W. & Sorrell, T. (2005). Proton nuclear magnetic resonance-based metabonomics for rapid diagnosis of meningitis and ventriculitis. *Clinical Infectious Diseases*, 41:1582–1590.

Cole, S., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C. & Tekaia, F. (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393(6685):537–544.

Cordoba, J., Gottstein, J. & Blei, A.T. (1996). Glutamine, *myo*-inositol, and organic brain osmolytes after portacaval anastosmosis in the rat: Implications for ammonia-induced brain edema. *Hepatology*, 24(4):919–923.

Cori, C.F. & Cori, G.T. (1928). The carbohydrate metabolism of tumors: III. The rate of glycolysis of tumor tissue in the living animal. *The Journal of Cancer Research,* 12:301–313.

Croda, M.G., Vidal, J.E., Hernandez, A.V., Dal Molin, T., Gualberto, F.A. & Penalva de Oliveira, A.C. (2010). Tuberculous meningitis in HIV-infected patients in Brazil: clinical and laboratory characteristics and factors associated with mortality. *International Journal of Infectious Disease,* 14:e586–591.

Cunha, B.A. (2006). Distinguishing bacterial from viral meningitis: the critical importance of the CSF lactic acid levels. *Intensive Care Medicine*, 32(8):1272–1273.

Danese, A. & McEwen, B.S. (2012). Adverse childhood experiences, allostasis, allostatic load, and age-related disease. *Physiology & Behavior,* 106:29–39.

Das, M.K., Bishwal, S.C., Das, A., Dabral, D., Badireddy, V.K., Pandit, B., Varghese, G.M. & Nanda, R.K. (2015). Deregulated tyrosine–phenylalanine metabolism in pulmonary tuberculosis patients. *Journal of Proteome Research*, 4:1947–1956.

Davis, J.M. & Ramakrishnan, L. (2009). The role of the granuloma in expansion and dissemination of early tuberculous infection. *Cell*, 136:37–49.

de Carvalho, L.P.S., Fischer, S.M., Marrero, J., Nathan, C., Ehrt, S. & Rhee, K.Y. (2010). Metabolomics of Mycobacterium tuberculosis reveals compartmentalized co-catabolism of carbon substrates. *Chemistry & Biology*, 17(10):1122–1131.

de Souza, G.A., Fortuin, S., Aguilar, D., Pando, R.H., McEvoy, C.R., van Helden, P.D., Koehler, C.J., Thiede, B., Warren, R.M. & Wiker, H.G. (2010). Using a label-free proteomics method to identify differentially abundant proteins in closely related hypo-and hypervirulent clinical Mycobacterium tuberculosis Beijing isolates. *Molecular and Cellular Proteomics,* 9(11):2414–2423.

Deb, C., Lee, C.M., Dubey, V.S., Daniel, J., Abomoelak, B., Sirakova, T.D., Pawar, S., Rogers, L. & Kolattukudy, P.E. (2009). A novel in vitro multiple-stress dormancy model for Mycobacterium tuberculosis generates a lipid-loaded, drug-tolerant, dormant pathogen. *PLoS One*, 4(6):e6077.

Delage, G. & Dusseault, M. (1979). Tuberculous meningitis in children: a retrospective study of 79 patients, with an analysis of prognostic factors. *Canadian Medical Association Journal,* 120:305–309.

Denery, J.R., Nunes, A.A.K., Hixon, M.S., Dickerson, T.J. & Janda, K.D. (2010). Metabolomics-Based Discovery of Diagnostic Biomarkers for Onchocerciasis. *PLoS Neglected Tropical Diseases*, 4(10):e834.

Denoroy, L., Zimmer, L., Renaud, B. & Parrot, S. (2013). Ultra high performance liquid chromatography as a tool for the discovery and the analysis of biomarkers of disease: A review. *Journal of Chromatography B*, 927:37–53.

Deniz, O., Gumus, S., Yaman, H., Ciftci, F., Ors, F., Cakir, E., Tozkoparan, E., Bilgic, H. & Ekiz, K. (2007). Serum total cholesterol, HDL-C and LDL-C concentrations significantly

correlate with the radiological extent of disease and the degree of smear positivity in patients with pulmonary tuberculosis. *Clinical Biochemistry,* 40:162–166.

Dercksen, M., Duran, M., Ijlst, L., Mienie, L.J., Reinecke, C.J., Ruiter, J.P.N., Waterham, H.R. & Wanders, R.J.A. (2012). Clinical variability of isovaleric acidemia in a genetically homogeneous population. *Journal of Inherited Metabolic Disorders,* 35(6):1021–1029.

Dercksen, M., Koekemoer, G., Duran, M., Wanders, R.J.A., Mienie, L.J. & Reinecke, C.J. (2013). Organic acid profile of isovaleric acidemia: a comprehensive metabolomics approach. *Metabolomics*, 9:765–777.

DeSalles, A.A., Kontos, H.A., Becker, D.P., Yang, M.S., Ward, J.D., Moulton, R., Gruemer, H.D., Lutz, H., Maset, A.L., Jenkins, L. & Marmarou, A. (1986). Prognostic significance of ventricular CSF lactic acidosis in severe head injury. *Journal of Neurosurgery*, 65:615–624.

DeSalles, A.A., Muizelaar, P.J. & Young, H.F. (1987). Hyperglycemia, cerebrospinal fluid lactic acidosis, and cerebral blood flow in severely head-injured patients. *Neurosurgery*, 21:45–50.

Dienel, G. & McKenna, M.C. (2014). A dogma-breaking concept: Glutamate oxidation in astrocytes is the source of lactate during aerobic glycolysis in resting subjects. *Journal of Neurochemistry,* 10:1–4.

Doerr, C.A., Starke, J.R. & Ong, L.T. (1995). Clinical and public health aspects of tuberculous meningitis in children. *Journal of Pediatrics,* 127:27–33.

Donald, P.R., Schoeman, J.F. & van Schalkwyk, H.J.S. (1985). The “road–to health” card in tuberculous meningitis. *Journal of Tropical Pediatrics*, 31(2):117–120.

Donald, P.R., Schoeman, J.F., Beyers, N., Nel, E.D., Carlini, S.M., Olsen, K.D. & McCracken, G.H. (1995). Concentrations of interferon γ, tumor necrosis factor α, and interleukin-1β in the cerebrospinal fluid of children treated for tuberculous meningitis. *Clinical Infectious Diseases,* 21(4):924–929.

Donald, P.R., Cotton, M.F., Hendricks, M.K., Schaaf, H.S., de Villiers, J.N. & Willemse, T.E. (1996). Pediatric meningitis in the Western Cape Province of South Africa. *Journal of Tropical Pediatrics,* 42:256–261.

Donald, P.R., Schaaf, H.S. & Schoeman, J.F. (2005). Tuberculous meningitis and miliary tuberculosis: the Rich focus revisited. *Journal of Infection,* 50(3):193–195.

Drennan, M.B., Nicolle, D., Quesniaux, V.J., Jacobs, M., Allie, N., Mpagi, J., Frémond, C. & Ryffel, B. (2004). Toll-like receptor 2-deficient mice succumb to Mycobacterium tuberculosis infection. The American *Journal of Pathology,* 164(1):49–57.

Du, Y., Wu, X. & Li, L. (2006). Mechanisms of bacterial meningitis-related deafness. *Drug Discovery Today: Disease Mechanisms,* 3(1):115–118.

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

Duarte, I.F., Diaz, S.O. & Gil, A.M. (2014). NMR metabolomics of human blood and urine in disease research. *Journal of Pharmaceutical and Biomedical Analysis,* 93:17–26.

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

Dunne, V.G., Bhattachayya, S., Besser, M., Rae, C. & Griffin, J.L. (2005). Metabolites from cerebrospinal fluid in aneurismal subarachnoid haemorrhage correlate with vasospasm and clinical outcome: A pattern-recognition ^{1}H NMR study. *NMR in Biomedicine*, 18:24–33.

Duzgun, A.P., Bugdayci, G., Sayin, B., Ozmen, M.M., Ozer, M.V. & Coskun, F. (2006). Serum D-lactate: a useful diagnostic marker for acute appendicitis. *Hepato-gastroenterology,* 54:1483–1486.

Efron, B. & Tibshirani, R.J. (1993). *An Introduction to the Bootstrap (Monographs on Statistics and Applied Probability)*, Chapman and Hall, New York.

El-Kebir, M., van der Kuip, M., van Furth, A.M. & Kirschner, D.E. (2013). Computational modeling of tuberculous meningitis reveals an important role for tumor necrosis factor-α. *Journal of Theoretical Biology,* 328:43–53.

Ellinger, J.J., Chylla, R.A., Ulrich, E.L. & Markley, J.L. (2013). Databases and software for NMR-based metabolomics. *Current Metabolomics*, 1:28–40.

Ellis, S. & Steyn, H. (2003). Practical significance (effect sizes) versus or in combination with statistical significance (p-values). *Management Dynamics*, 12(4):51–53.

Elstner, M. & Turnbull, D.M. (2011). Transcriptome analysis in mitochondrial disease. *Brain Research Bulletin,* 88(4):285–293.

Engelke, U.F.H., Liebrand-van Sambeek, M.L.F., de Jong, J.G., Leroy, J.G., Morava, É., Smeitink, J.A. & Wevers, R.A. (2004). N-acetylated metabolites in urine: Proton nuclear magnetic resonance spectroscopic study on patients with inborn errors of metabolism. *Clinical Chemistry*, 50(1):58–66.

Engelke, U.F.H., Tangerman, A., Willemsen, M.A.A.P., Moskau, D., Loss, S., Mudd, S.H. & Wevers, R.A. (2005). Dimethyl sulfone in human cerebrospinal fluid and blood plasma confirmed by one-dimensional ^{1}H and two-dimensional ^{1}H-^{13}C NMR. *NMR in Biomedicine*, 18:331–336.

Engelke, U.F.H., Kremer, B., Kluijtmans, L.A.J., van der Graaf, M., Morava, E., Loupatty, F.J., Wanders, R.J.A., Moskau, D., Loss, S., van den Bergh, E. & Wevers, R.A. (2006). NMR spectroscopic studies on the late onset form of 3-methylglutaconic aciduria type I and other defects in leucine metabolism. *NMR in Biomedicine*, 19:271–278.

Engelke, U.F.H., Moolenaar, S.H., Hoenderop, S.M.G.C., Morava, E., Van der Graaf, M., Heerschap, A. & Wevers, R.A. (2007). Handbook of ^{1}H-NMR spectroscopy in inborn errors of metabolism: body fluid NMR spectrum and in vivo MR spectroscopy. 2nd Edition, Heilbronn, Germany: SHS Verlagsgesellshaft

Eoh, H. & Rhee, K.Y. (2013). Multifunctional essentiality of succinate metabolism in adaptation to hypoxia in Mycobacterium tuberculosis.*Proceedings of the National Academy of Sciences*, 110(16):6554–6559.

Ernst, J.D. (1998). Macrophage receptors for Mycobacterium tuberculosis. *Infection and Immunity,* 66:1277–1281.

Evans, D.J. (2008). The use of adjunctive corticosteroids in the treatment of pericardial, pleural and meningeal tuberculosis: do they improve outcome? *Respiratory Medicine,* 102:793–800.

Ewaschuk, J.B., Naylor, J.M. & Zello, G.A. (2005). D-lactate in human and ruminant metabolism. *Journal of Nutrition*, 135:1619–1625.

Faksri, K., Prammanannan, T., Leechewengwongs, M. & Chaiprasert, A. (2012). Molecular epidemiology and drug resistance of tuberculous meningitis, *Meningitis*:85-112.

Farinha, N.J., Razali, K.A., Holzel, H., Morgan, G. & Novelli, V.M. (2000). Tuberculosis on the central nervous system in children: a 20-year survey. *Journal of Infection,* 41:61–68.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27:861–874.

Ferrara, C.T., Wang, P., Neto, E.C., Stevens, R.D., Bain, J.R., Wenner, B.R., Ilkayeva, O.R., Keller, M.P., Blasiole, D.A., Kendziorski, C. & Yandell, B.S. (2008). Genetic networks of liver metabolism revealed by integration of metabolic and transcriptional profiling. *PLoS Genetics*, 4(3):e1000034.

Fiehn, O. (2002). Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology*, 48:155–171.

Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., Summer, L.W., Goodacre, R., Hardy, N.W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., van Ommen, B., Lindon, J.C. & Sansone, S.A. (2007). The metabolomics standards initiative (MSI). *Metabolomics*, 3(3):175–178.

Fluss, R., Faraggi, D. & Reiser, B. (2005). Estimation of the Youden index and its associated cutoff point. *Biometrical Journal,* 47(4):458–472.

Frankenhaeuser, M., Lundberg, U., Von Wright, M.R., Von Wright, J. & Sedvall, G. (1986). Urinary monoamine metabolites as indices of mental stress in healthy males and females. *Pharmacology Biochemistry and Behavior*, 24(6):1521–1525.

Fritsch, F.N. & Carlson, R.E. (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17:238–246.

Fukusaki, E. & Kobayashi, A. (2005). Plant metabolomics: potential for practical operation. *Journal of Bioscience and Bioengineering*, 100(4):347–354.

Geiszler, P.C., Auer, D.P. & Daykin, C.A. (2013). The journey from metabolic profiling to biomarkers: The potential of NMR spectroscopy based metabolomics in neurodegenerative disease research. *Current Metabolomics*, 1:160–179.

Genton, B. & Berger, J.P. (1990). Cerebrospinal fluid lactate in 78 cases of adult meningitis. *Intensive Care Medicine*, 16(3):196–200.

Ghielmetti, M., Ren, H., Leib, S.L., Tauber, M.G. & Christen, S. (2003). Impaired cortical energy metabolism but not major antioxidant defenses in experimental bacterial meningitis. *Brain Research,* 976:139–148.

Giaume, C. (2010). Astroglial wiring is adding complexity to neuroglial networking. *Frontiers in Neuroenergetics*, 2:129.

Giaume, C., Kirchhoff, F., Matute, C., Reichenbach, A. & Verkhratsky, A. (2007). Glia: the fulcrum of brain diseases. *Cell Death & Differentiation*, 14(7):1324–1335.

Gika, H.G., Theodoridis, G.A., Wingate, J.E. & Wilson, I.D. (2007). Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: application to human urine. *Journal of Proteome Research,* 6:3291–3303.

Gimeno-Bayón, J., López-López, A., Rodríguez, M.J. & Mahy, N. (2014). Glucose pathways adaptation supports acquisition of activated microglia phenotype. *Journal of Neuroscience Research*, 92:723–731.

Girgis, N.L., Sultan, Y., Farid, Z., Mansour, M.M., Erian, M.W., Hanna, L.S. & Mateczun, A.J. (1998). Tuberculous meningitis, Abbassia fever hospital – naval medical research unit no. 3 – Cairo, Egypt, from 1976 to 1996. *American Journal of Tropical Medicine and Hygiene,* 58(1):28–34.

Gladden, L.B. (2004). Lactate metabolism: a new paradigm for the third millennium. *The Journal of Physiology*, 558(1):5–30.

Godreuil, S., Tazi, L. & Bañuls, A.L. (2007). Chapter 1. Pulmonary tuberculosis and mycobacterium tuberculosis: modern molecular epidemiology and perspectives. *Encyclopedia of infectious diseases: modern methodologies.*

Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G. & Kell, D.B. (2004). Metabolomics by numbers: Acquiring and understanding global metabolite data. *Trends in Biotechnology*, 22:245–252.

Goodacre, R., Broadhurst, D., Smilde, A.K., Kristal, B.S., Baker, J.D., Beger, R., Bessant, C., Connor, S., Capuani, G., Craign, A., Ebbels, T., Kell, D.B., Manetti, C., Newton, J., Paternostro, G., Somorlai, R., Sjöström, M., Trygg, J. & Wifert, F. (2007). Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics,* 3(3):231–241.

Goodacre, R. (2010). An overflow of… what else but metabolism! *Metabolomics*, 6:1–2.

Graham, S.M., Sismanidis, C., Menzies, H.J., Marais, B.J., Detjen, A.K. & Black, R.E. (2014). Importance of tuberculosis control to address child survival. *Lancet*, 383:1605–1607.

Green, J.A., Elkington, P.T., Pennington, C.J., Roncaroli, F., Dholakia, S., Moores, R.C., Bullen, A., Porter, J.C., Agranoff, D., Edwards, D.R. & Friedland, J.S. (2010). Mycobacterium tuberculosis Upregulates Microglial Matrix Metalloproteinase-1 and-3 Expression and Secretion via NF-κB–and Activator Protein-1–Dependent Monocyte Networks. *The Journal of Immunology*, *184*(11):6492–6503.

Griffin, J.E., Gawronski, J.D., DeJesus, M.A., Ioerger, T.R., Akerley, B.J. & Sassetti, C.M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathogens*, 7(9):e1002251.

Guo, K., Bamforth, F. & Li, L. (2011). Qualitative metabolome analysis of human cerebrospinal fluid by ^{13}C-/^{12}C-isotope dansylation labeling combined with liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *Journal of American Society for Mass Spectrometry*, 22:339–347.

Gupta, N., Goyal, N., Singha, U.K., Bhakuni, V., Roy, R. & Rastogi, A.K. (1999). Characterization of intracellular metabolites of axenic amastigotes of Leishmania donovani by ^{1}H NMR spectroscopy. *Acta Tropica,* 73(2):121–133.

Guy, P.A., Tavazzi, I., Bruce, S.J., Ramadan, Z. & Kochhar, S. (2008). Global metabolic profiling analysis on human urine by UPLC-TOFMS: Issues and method validation in nutritional metabolomics. *Journal of Chromatography B*, 871:253–260.

Haas, R.H., Parikh, S., Falk, M.J., Saneto, R.P., Wolf, N.I., Darin, N., Wong, L.J., Cohen, B.H. & Naviaux, R.K. (2008). The in-depth evaluation of suspected mitochondrial disease. *Molecular Genetics and Metabolism,* 94(1):16–37.

Hakim, J.G., Gangidzo, I.T., Heyderman, R.S., Mielke, J., Mushangi, E., Taziwa, A., Robertson, V.J., Musvaire, P. & Mason, P.R. (2000). Impact of HIV infection on meningitis in Harare, Zimbabwe: a prospective study of 406 predominantly adult patients. *AIDS,* 14:1401–1407.

Hall, P. & Selinger, B. (1989). A statistical justification to relating interlaboratory coefficients of variation with concentration levels. *Analytical Chemistry*, 61(13):1465–1466.

Hall, E.D., Oostveen, J.A. & Gurney, M.E. (1998). Relationship of microglial and astrocytic activation to disease onset and progression in a transgenic model of familial ALS. *Glia*, 23(3):249–256.

Hallows, K.R. & Frank, R.S. (1992). Changes in mechanical properties with DMSO-induced differentiation of HL-60 cells. *Biorheology*, 29:295–309.

Halouska, S., Chacon, O., Fenton, R.J., Zinniel, D.K., Barletta, R.G. & Powers, R. (2007). Use of NMR metabolomics to analyze the targets of D-cycloserine in mycobacteria: role of D-alanine racemase. *Journal of Proteome Research,* 6:4608–4614.

Harris, J.E., Nuttal, R.K., Elkington, P.T., Green, J.A., Horncastle, D.E., Graeber, M.B., Edwards, D.R. & Friedland, J.S. (2007). Monocyte-astrocyte networks regulate matrix metalloproteinase gene expression and secretion in central nervous system tuberculosis in vitro and in vivo. *Journal of Immunology,* 178:1199–1207.

Hashimoto, T., Hussien, R., Oommen, S., Gohil, K. & Brooks, G.A. (2007). Lactate sensitive transcription factor network in L6 cells: Activation of MCT1 and mitochondrial biogenesis. *The FASEB Journal*, 21:2602–2612.

Hashimoto, T., Hussien, R., Cho, H.S., Kaufer, D. & Brooks, G.A. (2008). Evidence for the mitochondrial lactate oxidation complex in rat neurons: Demonstration of an essential component of brain lactate shuttles. *PLoS ONE,* 3:e2915.

Hawkins, B.T. & Davis, T.P. (2005). The blood-brain barrier/neurovascular unit in health and disease. *Pharmacological Reviews*, 57:173–185.

Haydon, P.G. & Carmignoto, G. (2006). Astrocyte control of synaptic transmission and neurovascular coupling. *Physiological Reviews*, 86(3):1009–1031.

Hektoen, L. (1896). The vascular changes of tuberculous meningitis, especially tuberculous endarteritis. *The Journal of Experimental Medicine,* 1:112–163.

Hendriks, M.M.W.B., van Eeuwijk, F.A., Jellema, R.H., Westerhuis, J.A., Reijmers, T.H., Hoefsloot, H.C.J. & Smilde, A.K. (2011). Data-processing strategies for metabolomics studies. *Trends in Analytical Chemistry*, 30(10):1685–1698.

Hertz, L., Dringen, R., Schousboe, A. & Robinson, S.R. (1999). Astrocytes: glutamate producers for neurons. *Journal of Neuroscience Research*, 57(4):417–428.

Hewer, R., Vorster, J., Steffens, F.E. & Meyer, D. (2006). Applying biofluid ^{1}H NMR-based metabonomic techniques to distinguish between HIV-1 positive/AIDS patients on antiretroviral treatment and HIV-1 negative individuals. *Journal of Pharmaceutical and Biomedical Analysis,* 41(4):1442–1446.

Heyes, M.P., Ellis, R.J., Childers, L.R., Grant, I., Wolfson, T., Archibald, S. & Jernigan, T.L. (2001). Elevated cerebrospinal fluid quinolinic acid levels are associated with region-specific cerebral volume loss in HIV infection. *Brain*, 124:1033–1042.

Himmelreich, U., Malik, R., Kühn, T., Daniel, H.M., Somorjai, R.L., Dolenko, B. & Sorrell, T.C. (2009). Rapid etiological classification of meningitis by NMR spectroscopy based on metabolite profiles and host response. *PLoS ONE*, 4(4):e5328–e5335.

Hirose, Y., Mokuno, K., Wakai, M., Takahashi, A., Hashizume, Y., Yanagi, T. & Kato, K. (1995). Elevated cerebrospinal fluid levels of manganese superoxide dismutase in bacterial meningitis. *Journal of the Neurological Sciences,* 131:51–57.

Ho, J., Marais, B.J., Gilbert, G.L. & Ralph, A.P. (2013). Diagnosing tuberculous meningitis – have we made any progress? *Tropical Medicine and International Health*, 18:783–793.

Hoffmann, G.F., Meier-Augenstein, W., Stöckler, S., Surtees, R. & Nyhan, W.L. (1993). Physiology and pathophysiology of organic acids in cerebrospinal fluid. *Journal of Inherited Metabolic Disease*, 16(4):648–669.

Horsburgh Jr., C.R. & Rubin, E.J. (2011). Clinical practice. Latent tuberculosis infection in the United States. *New England Journal of Medicine*; 364(15):1441–1448.

Hosoglu, S., Ayaz, C., Geyik, M.F., Kökoglu, O.F. & Ceviz, A. (1998). Tuberculous meningitis in adults: an eleven-year review. *International Journal of Tuberculosis and Lung Disease,* 2(7):553–557.

Huy, N.T., Thao, N.T., Diep, D.T., Kikuchi, M., Zamora, J. & Hirayama, K. (2010). Cerebrospinal fluid lactate concentration to distinguish bacterial from aseptic meningitis: a systemic review and meta-analysis. *Critical Care,* 14(6):R240.

Ioannidis, J.P.A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8):696–701.

Jacobs, D.M., Deltimple, N., Van Velzen, E., Van Dorsten, F.A., Bingham, M., Vaughan, E.E. & Van Duynhoven, J. (2008). ^{1}H NMR metabolite profiling of feces as a tool to assess the impact of nutrition on the human microbiome. *NMR Biomedicine*, 21:615–626.

Jacobsen, M., Mattow, J., Repsilber, D. & Kaufmann, S.H.E. (2008). Novel strategies to identify biomarkers in tuberculosis. *Journal of Biological Chemistry,* 389(5):487–495.

Janse van Rensburg, P., Andronikou, S., van Toorn, R. & Pienaar, M. (2008). Magnetic resonance imaging of miliary tuberculosis of the central nervous system in children with tuberculous meningitis. *Pediatric Radiology,* 38(12):1306–1313.

Jeong, Y.J. & Lee, K.S. (2008). Pulmonary tuberculosis: up-to-date imaging and management. *American Journal of Roentgenology,* 191:834–844.

Jiménez, C., Ventura, R. & Segura, J. (2002). Validation of qualitative chromatographic methods: strategy in antidoping control laboratories. *Journal of Chromatography B*, 767:341–351.

Johnson, R.A. & Wichern, D.W. (1998). Applied multivariate statistical analysis, 4th ed Prentice-Hall Inc.

Kaddurah-Daouk, R. & Krishnan, K.R.R. (2009). Metabolomics: A global biochemical approach to the study of central nervous system diseases. *Neuropsychopharmacology*, 34:173–186.

Kamerling, J.P., Gerwig, G.J., Duran, M., Ketting, D. & Wadman, S.K. (1978). The absolute configuration of urinary 2-hydroxybutyric acid in patients with ketosis and lactic acidosis. *Clinica Chimica Acta*, 88:183–188.

Kanani, H., Chrysanthopoulos, P.K. & Klapa, M.I. (2008). Standardizing GC-MS metabolomics. *Journal of Chromatography B,* 871:191–201.

Kandanearatchi, A., & Brew, B.J. (2012). The kynurenine pathway and quinolinic acid: pivotal roles in HIV associated neurocognitive disorders. *FEBS Journal*, 279:1366–1374.

Karstaedt, A.S., Valtchanova, S., Barriere, R. & Crewe-Brown, H.H. (1998). Tuberculous meningitis in South African urban adults. *Quarterly Journal of Medicine,* 91:743–747.

Kashyap, R.S., Kainthla, R.P., Mudaliar, A.V., Purohit, H.J., Taori, G.M. & Daginawala, H.F. (2006). Cerebrospinal fluid adenosine deaminase activity: a complimentary tool in the early diagnosis of tuberculous meningitis. *Cerebrospinal Fluid Research,* 3(5):2–6.

Kashyap, R.S., Deshpande, P.S., Ramteke, S.R., Panchbhai, M.S., Purohit, H.J., Taori, G.M. & Hatim, F. (2010). Changes in cerebrospinal fluid cytokine expression in tuberculous meningitis patients with treatment. *Neuroimmunomodulation,* 17:333–339.

Katajamaa, M. & Orešič, M. (2007). Data processing for mass spectrometry-based metabolomics. *Journal of Chromatography A*, 1158:318–328.

Katrak, S.M., Shembalkar, P.L., Bijwe, S.R. & Bhandarkar, L.D. (2000). The clinical, radiological and pathological profile of tuberculous meningitis in patients with and without human immunodeficiency virus infection. *Journal of the Neurological Sciences,* 181:118–126.

Kaufmann, S.H.E. (2005). Recent findings in immunology give tuberculosis vaccines a new boost. *TRENDS in Immunology,* 26:660–667.

Kaufmann, S.H.E. & Parida, S.K. (2008). Tuberculosis in Africa: learning from pathogenesis for biomarker identification. *Cell Host & Microbe,* 4:219–228.

Kell, D.B. & Oliver, S.G. (2003). Here is the evidence, now where is the hypothesis? The complementary roles of inductive hypothesis-driven science in the post genomics-era. *BioEssays,* 66:99–105.

Kell, D.B. (2004). Metabolomics and systems biology: making sense of the soup. *Current Opinions in Microbiology,* 7(3):296–307.

Kell, D.B. (2006). Systems biology, metabolic modelling in drug discovery and development. *Drug Discovery Today*, 11(23/24).

Kilpatrick, M.E., Girgis, N.I., Yassin, M.W. & Abu El Ella, A.A. (1986). Tuberculous meningitis – clinical and laboratory review of 100 patients. *Journal of Hygiene,* 96:231–238.

Koene, S. & Smeitink, J. (2011). Mitochondrial medicine. *Journal of Inherited Metabolic Disease,* 34:247–248.

Kolb, S.A., Lahrtz, F., Paul, R., Leppert, D., Nadal, D., Pfister, H.W. & Fontana, A. (1998). Matrix metalloproteinases and tissue inhibitors of metalloproteinases in viral meningitis: upregulation of MMP-9 and TIMP-1 in cerebrospinal fluid. *Journal of Neuroimmunology,* 84:143–150.

Kowlgi, N.G. & Chhabra, L. (2015). D-Lactic acidosis: an underrecognized complication of short bowel syndrome. *Gastroenterology Research and Practice*, 2015.

Krishnan, N., Robertson, B.D. & Thwaites, G. (2010). The mechanisms and consequences of the extra-pulmonary dissemination of *Mycobacterium tuberculosis. Tuberculosis*, 90:361–366.

Kruman, I.I., Kostenko, M.A., Gordon, R.Y., Popov, V.I. & Umansky, S.R. (1993). Differentiation and apoptosis of murine neuroblastoma cells N1E115. *Biochemical and Biophysical Research Communications*, 191(3):1309–1318.

Kumar, R., Singh, S.N. & Kohli, N. (1999). A diagnostic rule for tuberculous meningitis. *Archives of Disease in Childhood,* 81:221–224.

Kumar, P., Bharti, S.K. & Varshney, U. (2011). Uracil excision repair in Mycobacterium tuberculosis cell-free extracts. *Tuberculosis*, 91:212–218.

Lal, S., Singhal, S.N., Burley, D.M. & Crossley, G. (1972). Effect of rifampicin and isoniazid on liver function. *BMJ*, 1:148–150.

Landaas, S. (1975). Accumulation of 3-hydroxyisobutyric acid, 2-methyl-3-hydroxybutyric acid and 3-hydroxyisovaleric acid in ketoacidosis. *Clinica Chimica Acta,* 64:143–154.

Lay Jr, J.O., Borgmann, S., Liyanage, R. & Wilkins, C.L. (2006). Problems with the “omics”. *Trends in Analytical Chemistry*, 25(11):1046–1056.

Lee, L.V. (2000). Neurotuberculosis among Filipino children: an 11 years experience at the Philippine Children’s Medical Center. *Brain & Development,* 22:469–474.

Lee, K.Y., Kim, E.H., Yang, W.S., Ryu, H., Cho, S.N., Lee, B.I. & Heo, J.H. (2004). Persistent increase of matrix metalloproteinases in cerebrospinal fluid of tuberculous meningitis. *Journal of the Neurological Sciences,* 220:73–78.

Lee, H., Sun, E., Ham, D. & Weissleder, R. (2008). Chip-NMR biosensor for detection and molecular analysis of cells. *Nature Medicine,* 14:869–874.

Lee, A., Toffaletti, D.L., Tenor, J., Soderblom, E.J., Thompson, J.W., Moseley, M.A., Price, M. & Perfect, J.R. (2010a). Survival defects of Cryptococcus neoformans mutants exposed to human cerebrospinal fluid result in attenuated virulence in an experimental model of meningitis. *Infection and Immunity,* 78(10):4213–4225.

Lee, Y.K., Menezes, J.S., Umesaki, Y. & Mazmanian, S.K. (2010b). Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proceedings of the National Academy of Sciences of the United States of America,* 108(Supplement 1):4615–4622.

Leen, W.G., Willemsen, M.A., Wevers, R.A. & Verbeek, M.M. (2012). Cerebrospinal fluid glucose and lactate: Age-specific reference values and implications for clinical practice. *PLoS ONE*, 7(8):e42745.

Leib, S.L., Bosacci, R., Gratzl, O. & Zimmerli, W. (1999). Predictive value of cerebrospinal fluid (CSF) lactate level versus CSF/blood glucose ratio for the diagnosis of bacterial meningitis following neurosurgery. *Clinical Infectious Diseases*, 29:69–74.

Leonard, J.M. & Des Prez, R.M. (1990). Tuberculous meningitis. *Infectious Disease Clinics of North America*, 4:769–787.

Levine, R.J. & Conn, H.O. (1967). Tyrosine metabolism in patients with liver disease. *Journal of Clinical Investigation,* 46(12):2012.

Levine, J., Panchalingam, K., Rapoport, A., Gershon, S., McClure, R.J. & Pettegrew, J.W. (2000). Increased cerebrospinal fluid glutamine levels in depressed patients. *Biological Psychiatry*, 47:586–593.

Li, J.V., Wang, Y., Saric, J., Nicholson, J.K., Dirnhofer, S., Singer, B.H., Tanner, M., Wittlin, S., Holmes, E. & Utzinger, J. (2008a). Global metabolic responses of NMRI mice to an experimental Plasmodium berghei infection. *Journal of Proteome Research,* 7(9):3948–3956.

Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., Zhou, H., Shen, J., Pang, X., Zhang, M. & Wei, H. (2008b). Symbiotic gut microbes modulate human metabolic phenotypes. *Proceedings of the National Academy of Sciences of the United States of America,* 105:2117–2122.

Liland, K.H. (2011). Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. *Trends in Analytical Chemistry*, 30(6):827–841.

Lindon, J.C., Nicholson, J.K. & Holmes, E. (2011). The handbook of metabonomics and metabolomics. London: Elsevier Science.

Liu, B., Wang, K., Hui-Ming Gao, H.M., Mandavilli, B., Wang, J.Y. & Hong, J.S. (2001). Molecular consequences of activated microglia in the brain: Overactivation induces apoptosis. *Journal of Neurochemistry*, 77:182–189.

Logan, J.G. & Barksdale, D.J. (2008). Allostasis and allostatic load: expanding the discourse on stress and cardiovascular disease. *Journal of Nursing and Healthcare of Chronic Illness* in association with *Journal of Clinical Nursing*, 17:201–208.

Lu, H., Dunn, W.B., Shen, H., Kell, D.B. & Liang, Y. (2008). Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS. *Trends in Analytical Chemistry*, 27(3):215–227.

Madsen, R., Lundstedt, T. & Trygg, J. (2010). Chemometrics in metabolomics – a review in human disease diagnosis. *Analytica Chimica Acta*, 659:23–33.

Magistretti, P.J. & Pellerin, L. (1996). Cellular bases of brain energy metabolism and their relevance to functional brain imaging: evidence for a prominent role of astrocytes. *Cerebral Cortex*, 6(1):50–61.

Magistretti, P.J. & Pellerin, L. (1999a). Astrocytes couple synaptic activity to glucose utilization in the brain. *Physiology*, 14(5):177–182.

Magistretti, P.J. & Pellerin, L. (1999b). Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. *Philosophical Transactions of the Royal Society B*, 354:1155–1163.

Magistretti, P.J. (2000). Cellular bases of functional brain imaging: insights from neuron-glia metabolic coupling. *Brain Research*, 886(1):108–112.

Magistretti, P.J. (2006). Neuron–glia metabolic coupling and plasticity. *Journal of Experimental Biology*, 209(12):2304–2311.

Magistretti, P.J. (2009). Role of glutamate in neuron-glia metabolic coupling. *The American Journal of Clinical Nutrition*, 90(3):875S–880S.

Maizels, N. & Scharff, M.D. (2004). Molecular mechanisms of hypermutation. In Molecular Biology of B cells. T. Honjo *et al.*, eds. (Amsterdam: Elsevier Academic Press), 327–338.

Mamas, M., Dunn, W.B., Neyses, L. & Goodacre, R. (2011). The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Archives of Toxicology*, 85:5–17.

Mancuso, M., Orsucci, D., Coppedè, F., Nesti, C., Choub, A., Siciliano, G. (2009). Diagnostic approach to mitochondrial disorders: The need for a reliable biomarker. *Current Molecular Medicine,* 9(9):1095–1107.

Mandal, R., Guo, A.C., Chaudhary, K.K., Liu, P., Yallou, F.S., Edison, D., Aziat, F. & Wishart, D.S. (2012). Multi-platform characterization of the human cerebrospinal fluid metabolome: A comprehensive and quantitative update. *Genome Medicine*, 4:38–49.

Marais, S., Thwaites, G., Schoeman, J.F., Torok, M.E., Misra, U.K., Prasad, K., Donald, P.R., Wilkinson, R.J. & Marais, B.J. (2010). Tuberculous meningitis: a uniform case definition for use in clinical research. *The Lancet Infectious Diseases,* 10(11):803–812.

Marcos, M.A., Martinez, E., Almela, M., Mensa, J. & Jiménez de Anta, M.T. (2001). New rapid antigen test for diagnosis of pneumococcal meningitis. *Lancet,* 357:1499–1500.

Martín-Hernández, E., García-Silva, M.T., Vara, J., Campos, Y., Cabello, A., Muley, R., del Hoyo, P., Martín, M.A. & Arenas, J. (2005). Renal pathology in children with mitochondrial diseases. *Pediatric Nephrology,* 20(9):1299–1305.

Martineau, A.R., Honecker, F.U., Wilkinson, R.J. & Griffiths, C.J. (2007). Vitamin D in the treatment of pulmonary tuberculosis. *Journal of Steroid Biochemistry & Molecular Biology,* 103:793–798.

Marthyn, P., Beuscart, A., Coll, J., Moreau-Gachelin, F. & Righi, M. (1998). DMSO reduces CSF-1 receptor levels and causes apoptosis in v-*myc* immortalized mouse macrophages. *Experimental Cell Research*, 243:94–100.

Mason, S., Moutloatse, G.P., van Furth, A.M., Solomons, R., van Reenen, M., Reinecke, C.J. & Koekemoer, G. (2014). KEMREP: A new qualitative method for the assessment of an analyst's ability to generate a metabolomics data matrix by gas chromatography – mass spectrometry. *Current Metabolomics,* 2(1):15–26.

Mason, S., van Furth, A.M., Mienie, L.J., Engelke, U.F.H., Wevers, R.A., Solomons, R. & Reinecke, C.J. (2015). A hypothetical astrocyte–microglia lactate shuttle derived from a ^{1}H NMR metabolomics analysis of cerebrospinal fluid from a cohort of South African children with tuberculous meningitis. *Metabolomics*, 11:822–837.

Mason, S., Reinecke, C.J., Solomons, R. & van Furth, A.M. (2016a). Tuberculous meningitis in infants and children: Insights from nuclear magnetic resonance metabolomics. *South African Journal of Science,* 112(3/4), (http://dx.doi.org/10.17159/sajs.2016/2015008).

Mason, S., Reinecke, C.J., Kulik, W., van Cruchten, A., Solomons, R. & Tutu van Furth, A.M. (2016b). Cerebrospinal fluid in tuberculous meningitis exhibits only the L-enantiomer of lactic acid. *BMC Infectious Diseases* – in press.

Mason, S., Tutu van Furth, A.M., Solomons, R., Wevers, R.A., van Reenen, M. & Reinecke, C.J. (2016c). A putative urinary biosignature for diagnosis and follow-up of tuberculous meningitis in children: Outcome of a metabolomics study disclosing host–pathogen responses. *Metabolomics* – submitted.

McEwen, B.S. (2006). Sleep deprivation as a neurobiologic and physiologic stressor: allostasis and allostatic load. *Metabolism Clinical and Experimental,* 55 (Suppl 2):S20–S23.

McEwen, B.S. (2008). Central effects of stress hormones in health and disease: understanding the protective and damaging effects of stress and stress mediators. *European Journal of Pharmacology,* 583:174–185.

McEwen, B.S. & Gianaros, P.J. (2010). Stress- and allostasis induced brain plasticity. *Annual Review of Medicine,* 62:431–445.

McKinney, J.D., zu Bentrup, K.H., Muñoz-Elías, E.J., Miczak, A., Chen, B., Chan, W.T., Swenson, D., Sacchettini, J.C., Jacobs, W.R. & Russell, D.G. (2000). Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature*, 406(6797):735–738.

Medana, I.M., Hien, T.T., Day, N.P., Nguyen, H.P., Nguyen, T.H.M., Van Chu'ong, L., Tran, T.H., Taylor, A., Salahifar, H., Stocker, R. & Smythe, G. (2002). The clinical significance of cerebrospinal fluid levels of kynurenine pathway metabolites and lactate in severe malaria. *Journal of Infectious Diseases*, 185:650–656.

Medana, I.M., Day, N.P., Salahifar-Sabet, H., Stocker, R., Smythe, G., Bwanaisa, L., Njobvu, A., Kayira, K., Turner, G.D., Taylor, T.E. & Hunt, N.H. (2003). Metabolites of the kynurenine pathway of tryptophan metabolism in the cerebrospinal fluid of Malawian children with malaria. *Journal of Infectious Diseases*, 188(6):844–849.

Meissner-Roloff, R.J., Koekemoer, G. & Warren, R.M. (2012). A metabolomics investigation of a hyper-and hypo-virulent phenotype of Beijing lineage M. tuberculosis. *Metabolomics*, 8(6):1194–1203.

Mels, C.M.C., van Rensburg, P.J., van der Westhuizen, F.H., Pretorius, P.J. & Erasmus, E. (2011). Increased excretion of C4-carnitine species after a therapeutic acetylsalicylic acid dose: Evidence for an inhibitory effect on short-chain fatty acid metabolism. *ISRN Pharmacology,* doi:10.540/2011/851870.

Meyer, M.R., Weber, A.A. & Maurer, H.M. (2011). A validated GC-MS procedure for fast, simple, and cost-effective quantification of glycols and GHB in human plasma and their identification in urine and plasma for emergency toxicology. *Analytical and Bioanalytical Chemistry*, 400:411–414.

Milman, B.L. (2005). Identification of chemical compounds. *Trends in Analytical Chemistry*, 24(6):493–508.

Miric, D., Katanic, R., Kisic, B., Zoric, L., Miric, B., Mitic, R. & Dragojevic, I. (2010). Oxidative stress and myeloperoxidase activity during bacterial meningitis: effects of febrile episodes and the BBB permeability. *Clinical Biochemistry,* 43:246–252.

Misra, U.K., Kalita, J., Srivastava, M. & Mandal, S.K. (1996). Prognosis of tuberculous meningitis: a multivariate analysis. *Journal of the Neurological Sciences,* 137:57–61.

Mok, D.K.W. & Chau, F.T. (2006). Chemical information of Chinese medicines: a challenge to chemist. *Chemometrics and Intelligent Laboratory Systems*, 82:210–217.

Möller, M. & Hoal, E.G. (2010). Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. *Tuberculosis,* 90:71–83.

Moore, K. & Roberts, L.J. (1998). Measurement of lipid peroxidation *Free Radical Research*, 28(6):659–671.

Morath, M.A., Okun, J.G., Müller, I.B., Sauer, S.W., Hörster, F., Hoffmann, G.F. & Kölker, S. (2008). Neurodegeneration and chronic renal failure in methylmalonic aciduria — A pathophysiological approach. *Journal of Inherited Metabolic Disease*, 31(1):35–43.

Morgado, T.C., Kinsky, M., Carrara, H., Rothemeyer, S. & Semple, P. (2013). Prognostic value of computed tomography – evident cerebral infarcts in adult patients with tuberculous meningitis and hydrocephalus treated with an external ventricular drain. *World Neurosurgery,* 80(6):e255–e260.

Muñoz-Elías, E.J., Upton, A.M., Cherian, J. & McKinney, J.D. (2006). Role of the methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and virulence. *Molecular Microbiology*, 60(5):1109–1122.

Nagesh Babu, G., Kumar, A., Kalita, J. & Misra, U.K. (2008). Proinflammatory cytokine levels in the serum and cerebrospinal fluid of tuberculous meningitis patients. *Neuroscience Letters,* 436:48–51.

Nareika, A., He, L., Game, B.A., Slate, E.H., Sanders, J.J., London, S.D., Lopes-Virella, M.F. & Huang, Y. (2005). Sodium lactate increases LPS-stimulated MMP and cytokine expression in U937 histiocytes by enhancing AP-1 and NF-kappaB transcriptional activities. *American Journal of Physiology, Endocrinology and Metabolism,* 289:E534–E542.

Nesbitt, N.M., Yang, X., Fontán, P., Kolesnikova, I., Smith, I., Sampson, N.S. & Dubnau, E. (2010). A thiolase of Mycobacterium tuberculosis is required for virulence and production of androstenedione and androstadienedione from cholesterol. *Infection and Immunity*, 78(1):275–282.

Newington, J.T., Harris, R.A. & Cumming, R.C. (2013). Reevaluating metabolism in Alzheimer's disease from the perspective of the astrocyte-neuron lactate shuttle model. *Journal of Neurodegenerative Diseases,*1–13.

Nishizuka, Y., Ichiyama, A. & Hayaishi, O. (1970). Metabolism of the benzene ring of tryptophan (Mammals). In: Methods of Enzymology, Metabolism of amino acids and amines, Ed. H. Tabor and C.W. Tabor, Academic Press, New York and London, 17A:463–491.

Norenberg, M.D., Mozes, L.W., Gregorios, J.B. & Norenberg, L.O.B. (1987). Effects of lactic acid on astrocytes in primary culture. *Journal of Neuropathology & Experimental Neurology*, 46(2):154–166.

Odièvre, M., Lombes, A., Dessemme, P., Santer, R., Brivet, M., Chevallier, B., Lagardère, B. & Odièvre, M. (2002). A secondary respiratory chain defect in a patient with Fanconi–Bickel syndrome. *Journal of Inherited Metabolic Disease,* 25(5):379–384.

Oliver, S.G., Winson, M.K., Kell, D.B. & Baganz, F. (1998). Systematic functional analysis of the yeast genome. *Trends in Biotechnology*, 16:373–378.

Orešič, M., Hänninen, V.A. & Vidal-Puig, A. (2008). Lipidomics: a new window to biomedical frontiers. *Trends in Biotechnology,* 26(12):647–652.

Parida, S.K. & Kaufmann, S.H.E. (2010). The quest for biomarkers in tuberculosis. *Drug Discovery Today,* 15(3/4):148–157.

Pardasani, V., Shukla, G., Singh, S., Goyal, V. & Behari, M. (2008). Abnormal sleep-wake cycles in patients with tuberculous meningitis: a case-control study. *Journal of the Neurological Sciences,* 269:126–132.

Pasikanti, K.K., Ho, P.C. & Chan, E.C.Y. (2008). Development and validation of a gas chromatography/mass spectrometry metabonomic platform for the global profiling of urinary metabolites. *Rapid Communication in Mass Spectrometry,* 22:2984–2992.

Patel, V.B., Padayatchi, N., Bhigjee, A.I., Allen, J., Bhagwan, B., Moodley, A.A. & Mthiyane, T. (2004). Multidrug-resistant tuberculous meningitis in KwaZulu-Natal, South Africa. *Clinical Infectious Diseases,* 38:851–856.

Patel, V.B., Singh, R., Connolly, C., Kasprowicz, V., Ndung'u, T. & Dheda, K. (2011). Comparative utility of cytokine levels and quantitative RD-1-specific T cell responses for rapid immunodiagnosis of tuberculous meningitis. *Journal of Clinical Microbiology,* 49(11):3971–3976.

Patel, A.B., Lai, J.C., Chowdhury, G.M., Hyder, F., Rothman, D.L., Shulman, R.G. & Behar, K.L. (2014). Direct evidence for activity-dependent glucose phosphorylation in neurons with implications for the astrocyte-to-neuron lactate shuttle. *Proceedings of the National Academy of Sciences*, 111:5385–5390.

Pears, M.R., Cooper, J.D., Mitchison, H.M., Mortishire-Smith, R.J., Pearce, D.A. & Griffin, J.L. (2005). High resolution ^{1}H NMR-based metabolomics indicates a neurotransmitter cycling deficit in cerebral tissue from a mouse model of batten disease. *The Journal of Biological Chemistry*, 280(52):42508–42514.

Pellerin, L., Pellegri, G., Bittar, P.G., Charnay, Y., Bouras, C., Martin, J.L., Stella, N. & Magistretti, P.J. (1998). Evidence supporting the existence of an activity-dependent astrocyte-neuron lactate shuttle. *Developmental Neuroscience,* 20:291–299.

Pellerin, L. (2003). Lactate as a pivotal element in neuron–glia metabolic cooperation. *Neurochemistry international*, 43(4):331–338.

Pellerin, L. & Magistretti, P.J. (2003). Food for thought: Challenging the dogmas. *Journal of Cerebral Blood Flow and Metabolism,* 23:1282–1286.

Pellerin, L. & Magistretti, P.J. (2004). Neuroenergetics: Calling upon astrocytes to satisfy hungry neurons. *Neuroscientist*, 10:53–62.

Pellerin, L. (2010). Food for thought: the importance of glucose and other energy substrates for sustaining brain function under varying levels of activity. *Diabetes & Metabolism*, 36:S59–S63.

Pellerin, L. & Magistretti, P.J. (2011). Sweet sixteen for ANLS. *Journal of Cerebral Blood Flow & Metabolism*, 32(7):1152–1166.

Pérez-Guzmán, C. & Vargas, M.H. (2006). Hypocholesterolemia: a major risk factor for developing pulmonary tuberculosis. *Medical Hypotheses,* 66:1227–1230.

Perez-Velez, C.M. & Marais, B.J. (2012). Tuberculosis in children. *The New England Journal of Medicine,* 367(4):348–361.

Peterson, P.K., Gekker, G., Hu, S., Sheng, W.S., Anderson, W.R., Ulevitch, R.J., Tobias, P.S., Gustafson, K.W., Molitor, T.W. & Chao, C.C. (1995). CD14 receptor-mediated uptake of nonopsonized *Mycobacterium tuberculosis* by human microglia. *Infection and Immunity*, 63:1598–1602.

Philp, A., Macdonald, A.L. & Watt, P.W. (2005). Lactate–a signal coordinating cell and systemic function. *Journal of Experimental Biology*, 208:4561–4575.

Pierre, K., & Pellerin, L. (2005). Monocarboxylate transporters in the central nervous system: distribution, regulation and function. *Journal of Neurochemistry,* 94:1–14.

Principi, N. & Esposito, S. (2012). Diagnosis and therapy of tuberculous meningitis in children. *Tuberculosis*, 92:377–383.

Qureshi, G.A., Baig, S.M., Bednar, I., Halawa, A. & Parvez, S.H. (1998). The neurochemical markers in cerebrospinal fluid to differentiate between aseptic and tuberculous meningitis. *Neurochemistry International,* 32:197–203.

Rake, A.J., Srinivasan, V., Nadkarni, V., Kaptan, R. & Newth, C.J.L. (2010). Glucose variability and survival in critically ill children: Allostasis or harm? *Pediatric Critical Care Medicine,* 11(6):707–712.

Ranjbar, M. & Amiri, M. (2015). On the role of astrocyte analog circuit in neural frequency adaptation. *Neural Computing and Applications*, 1–13.

Rauste-von Wright, M. & Frankenhaeuser, M. (1989). Females' emotionality as reflected in the excretion of the dopamine metabolite HVA during mental stress. *Psychological Reports*, 64:856–858.

Reddy, Y.N., Murthy, S.V., Krishna, D.R. & Prabhakar, M.C. (2009). Oxidative metabolic changes in pleural fluid of tuberculosis patients. *Bangladesh Journal of Pharmacology,* 4:69–72.

Reinecke, F., Smeitink, J.A.M. & van der Westhuizen, F.H. (2009). OXPHOS gene expression and control in mitochondrial disorders. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease,* 1792(12):1113–1121.

Reinecke, C.J., Koekemoer, G., van der Westhuizen, F.H., Louw, R., Lindeque, J.Z., Mienie, L.J. & Smuts, I. (2012). Metabolomics of urinary organic acids in respiratory chain deficiencies in children. *Metabolomics,* 8(2):264–283.

Reither, K., Saathoff, E., Jung, J., Minja, L.T., Kroidl, I., Saad, E., Huggett, J.F., Ntinginya, E.N., Maganga, L., Maboko, L. & Hoelscher, M. (2009). Low sensitivity of a urine LAM-ELISA in the diagnosis of pulmonary tuberculosis. *BMC Infectious Diseases*, 9(1):141–150.

Rich, A. & McCordock, H. (1933). The pathogenesis of tuberculous meningitis. *Bulletin of Johns Hopkins Hospital,* 52:2–37.

Rieder, H.L., Kelly, G.D., Bloch, A.B., Cauthen, G.M. & Snider, D.E. (1991). Tuberculosis diagnosed at death in the United States. *Chest,* 100:678–681.

Rock, R.B., Hu, S., Gekker, G., Sheng, W.S., May, B., Kapur, V. & Peterson, P.K. (2005). *Mycobacterium tuberculosis*-induced cytokine and chemokine expression by human microglia and astrocytes: effects of dexamethasone. *The Journal of Infectious Diseases*, 192:2054–205.

Rocke, D.M. & Lorenzato, S. (1995). A two-component model for measurement error in analytical chemistry. *Technometrics*, 37:176–184.

Rosenblum, W. (1983). Dimethyl sulfoxide effects on platelet aggregation and vascular reactivity in pial microcirculation. *Annals of the New York Academy of Sciences*, 411:110–119.

Ruslami, R., Ganiem, A.R., Dian, S., Apriana, L., Achmad, T.H., van der Ven, A.J., Borm, G., Aarnoutse, R.E. & van Crevel, R. (2013). Intensified regimen containing rifampicin and moxifloxacin for tuberculous meningitis: An open-label, randomised controlled phase 2 trial. *Lancet Infectious Diseases,* 13:27–35.

Russell, D.G., Barry, C.E. & Flynn, J.L. (2010). Tuberculosis: what we don't know can, and does, hurt us. *Science*, 328(5980):852–856.

Savvi, S., Warner, D.F., Kana, B.D., McKinney, J.D., Mizrahi, V. & Dawes, S.S. (2008). Functional characterization of a vitamin B12-dependent methylmalonyl pathway in

mycobacterium tuberculosis: Implications for propionate metabolism during growth on fatty acids. *Journal of Bacteriology,* 190(1):3886–3895.

Scheijen, J.L., Hanssen, N.M., van de Waarenburg, M.P., Jonkers, D.M., Stehouwer, C.D. & Schalkwijk, C.G. (2012). L (+) and D (-) lactate are increased in plasma and urine samples of type 2 diabetes as measured by a simultaneous quantification of L (+) and D (-) lactate by reversed-phase liquid chromatography tandem mass spectrometry. *Experimental Diabetes Research*, 2012:234812.

Schnell, R. & Schneider, G. (2010). Structural enzymology of sulphur metabolism in mycobacterium tuberculosis. *Biochemical and Biophysical Research Communications,* 396:33–38.

Schoeman, J.F., Wait, J., Burger, M., Van Zyl, F., Fertig, G., Janse van Rensburg, A., Springer, P. & Donald, P. (2002). Long-term follow-up of childhood tuberculous meningitis. *Developmental Medicine and Child Neurology,* 44(8):522–526.

Schoeman, J.C. & du Preez, I. (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.

Schurr, A., West, C.A. & Rigor, B.M. (1988). Lactate-supported synaptic function in the rat hippocampal slice preparation. *Science,* 240:1326–1328.

Schurr, A., Payne, R.S., Miller, J.J. & Rigor, B.M. (1997a). Glia are the main source of lactate utilized by neurons for recovery of function posthypoxia. *Brain Research*, 774:221–224.

Schurr, A., Payne, R.S., Miller, J.J. & Rigor, B.M. (1997b). Brain lactate is an obligatory aerobic energy substrate for functional recovery after hypoxia: further in vitro validation. *Journal of Neurochemistry*, 69(1):423–426.

Schurr, A., Payne, R.S., Miller, J.J. & Rigor, B.M. (1997c). Brain lactate, not glucose, fuels the recovery of synaptic function from hypoxia upon reoxygenation: an in vitro study. *Brain Research*, 744(1):105–111.

Schurr, A., Miller, J.J., Payne, R.S. & Rigor, B.M. (1999). An increase in lactate output by brain tissue serves to meet the energy needs of glutamate-activated neurons. *The Journal of Neuroscience*, 19(1):34–39.

Schurr, A. (2005). Lactate: the ultimate cerebral oxidative energy substrate? *Journal of Cerebral Blood Flow & Metabolism*, 26(1):142–152.

Schutte, C.M. & van der Meyden, C.H. (1998). A prospective study of Glasgow Coma Scale (GCS), age, CSF-neutrophil, and CSF-protein and glucose levels as prognostic indicators in 100 adult patients with meningitis. *Journal of Infection*, 37:112–115.

Scriver, C.R. (2012). Physician's guide to the laboratory diagnosis of metabolic diseases. Blau N, Duran M, Blaskovics ME, Gibson KM, (Eds.). Springer Science & Business Media.

Seddon, J.A., Visser, D.H., Bartens, M., Jordaan, A.M., Victor, T.C., van Furth, A.M., Schoeman, J.F. & Schaaf, H.S. (2012). Impact of drug resistance on clinical outcome in children with tuberculous meningitis. *The Pediatric Infectious Disease Journal,* 31(7):711–716.

Seehusen, D.A., Reeves, M.M. & Fomin, D.A. (2003). Cerebrospinal fluid analysis. *American Family Physician*, 68:1103–1108.

Shaham, O., Slate, N.G., Goldberger, O., Xu, Q., Ramanathan, A., Souza, A.L., Clish, C.B., Sims, K.B. & Mootha, V.K. (2010). A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent media from cultured muscle cells. *Proceedings of the National Academy of Sciences of the USA*, 107(4):1571–1575.

Shannon, M., King, T.L. & Kennedy, H.P. (2007). Allostasis: A theoretical framework for understanding and evaluating perinatal health outcomes. *Journal of Obstetric, Gynecologic, & Neonatal Nursing,* 36:125–134.

Shapiro, S., Miller, A., Lahat, N., Sobel, E. & Lerner, A. (2003). Expression of matrix metalloproteinases, sICAM-1 and IL-8 in CSF from children with meningitis. *Journal of the Neurological Sciences,* 206:43–48.

Sharma, S.K. & Mohan, A. (2004). Extrapulmonary tuberculosis. *Indian Journal of Medical Research,* 120:316–353.

Shaw, P.J., Forrest, V., Ince, P.G., Richardson, J.P. & Wastell, H.J. (1995). CSF and plasma amino acid levels in motor neuron disease: Elevation of CSF glutamate in a subset of patients. *Neurodegeneration*, 4:209–216.

Sheather, S.J. & Jones, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B*, 53(3):683–690.

Sheedy, J.R., Wettenhall, R.E., Scanlon, D., Gooley, P.R., Lewis, D.P., McGregor, N., Stapleton, D.I., Butt, H.L. & De Meirleir, K.L. (2009). Increased D-lactic acid intestinal bacteria in patients with chronic fatigue syndrome. *In Vivo*, 23:621–628.

Shendurnikar, N. & Shastri, N. (1994). Biochemical basis of inflammation with special reference to acute bacterial meningitis. *Indian Pediatrics,* 3:487–490.

Shepherd, T., Dobson, G., Verrall, S.R., Conner, S., Griffiths, D.W., McNicol, J.W., Davies, H.V. & Stewart, D. (2007). Potato metabolomics by GC-MS: what are the limiting factors? *Metabolomics*, 3(4):475–488.

Shin, J.H., Yang, J.Y., Jeon, B.Y., Yoon, Y.J., Cho, S.N., Kang, Y.H., Ryu, D.H. & Hwang, G.S. (2011). ^{1}H NMR metabolomic profiling in mice infected with *Mycobacterium tuberculosis. Journal of Proteome Research*, 10:2238–2247.

Shurubor, Y.I., Matson, W.R., Willett, W.C., Hankinson, S.E. & Kristal, B.S. (2007). Biological variability dominates and influences analytical variance in HPLC-ECD studies of the human plasma metabolome. *BMC Clinical Pathology,* 7:9*.*

Siesjo, B.K. (1988). Acidosis and ischemic brain damage. *Neurochemical Pathology*, 9:31–88.

Silber, E., Sonnenberg, P., Ho, K.C., Koornhof, H.J., Eintracht, S., Morris, L. & Saffer, D. (1999). Meningitis in a community with high prevalence of tuberculosis and HIV infection. *Journal of Neurological Sciences,* 162:20–26.

Sinclair, A.J., Viant, M.R., Ball, A.K., Burdon, M.A., Walker, E.A., Stewart, P.M., Rauz, S. & Young, S.P. (2009). NMR-based metabolomic analysis of cerebrospinal fluid and serum in neurological diseases – a diagnostic tool? *NMR in Biomedicine*, 23:123–132.

Singh, B.S., Patwari, A.K. & Deb, M. (1994). Serum sodium and osmolal changes in tuberculous meningitis. *Indian Pediatrics,* 31:1345–1350.

Sinha, M.K., Garg, R.K., Anuradha, H.K., Agarwal, A., Singh, M.K., Verma, R. & Shukla, R. (2010). Vision impairment in tuberculous meningitis: predictors and prognosis. *Journal of the Neurological Sciences,* 290:27–32.

Slupsky, C.M., Rankin, K.N., Fu, H., Chang, D., Rowe, B.H., Charles, P.G., McGeer, A., Low, D., Long, R., Kunimoto, D., Sawyer, M.B., Fedorak, R.N., Adamko, D.J., Saude, E.J.,

Shah, S.L. & Marrie, T.J. (2009). Pneumococcal pneumonia: potential for diagnosis through a urinary metabolic profile. *Journal of Proteome Research,* 8(12):5550–5558.

Smilde, A.K., van der Wef, M.J., Schaller, J.P. & Kistemaker, C. (2009). Characterizing the precision of mass-spectrometry-based metabolic profiling platforms. *Analyst*, 134:2281–2285.

Smuts, I., Louw, R., Du Toit, H., Klopper, B., Mienie, L.J. & Van Der Westhuizen, F.H. (2010). An overview of a cohort of South African patients with mitochondrial disorders. *Journal of Inherited Metabolic Disease,* 33(3):95–104.

Smuts, I., van der Westhuizen, F.H., Louw, R., Mienie, L.J., Engelke, U.F., Wevers, R.A., Mason, S., Koekemoer, G. & Reinecke, C.J. (2013). Disclosure of a putative biosignature for respiratory chain disorders through a metabolomics approach. *Metabolomics*, 9(2):379–391.

Sokoloff, L. (1989). Circulation and energy metabolism of the brain. *Basic Neurochemistry*, 2:338–413.

Solari, L., Soto, A., Agapito, J.C., Acurio, V., Vargas, D., Battaglioli, T., Accinelli, R.A., Gotuzzo, E. & van der Stuyft, P. (2013). The validity of cerebrospinal fluid parameters for the diagnosis of tuberculous meningitis. *International Journal of Infectious Diseases,* 17:e1111–e1115.

Somashekar, B.S., Amin, A.G., Rithner, C.D., Troudt, J., Basaraba, R., Izzo, A., Crick, D.C. & Chatterjee, D. (2011). Metabolic profiling of lung granuloma in *Mycobacterium tuberculosis* infected guinea pigs: ex vivo ^{1}H magic angle spinning NMR studies. *Journal of Proteome Research*, 10:4186–4195.

Spranger, M., Krempien, S., Schwab, S., Maiwald, M., Bruno, K. & Hacke, W. (1996). Excess Glutamate in the cerebrospinal fluid in bacterial meningitis. *Journal of the Neurological Sciences,* 143:126–131.

Starke, J.R. (2003). Pediatric tuberculosis: Time for a new approach. *Tuberculosis,* 83:208–212.

Steele, M.L. & Robinson, S.R. (2012). Reactive astrocytes give neurons less support: implications for Alzheimer's disease. *Neurobiology of Aging*, 33(2):423e1–423e13.

Sterling, P. & Eyer, J. (1988). Allostasis: A new paradigm to explain arousal pathology. Handbook of life stress, cognition and health. New York: John Wiley & Sons, 629–649.

Stewart, J.A. (2006). The detrimental effects of allostasis: allostatic load as a measure of cumulative stress. *Journal of Physiological Anthropology,* 25(1):133–145.

Stolberg, L., Rolfe, R., Gitlin, N., Merritt, J., Mann Jr, L., Linder, J. & Finegold, S. (1982). D-Lactic acidosis due to abnormal gut flora: diagnosis and treatment of two cases. *New England Journal of Medicine*, 306:1344–1348.

Stoop, M.P., Coulier, L., Rosenling, T., Shi, S., Smolinska, A.M., Buydens, L., Ampt, K., Sting, C., Dane, A., Muilwijk, B., Luitwieler, R.L., Smitt, P.A.E.S., Hintzen, R.Q., Bischoff, R., Wijmenga, S.S., Hankemeier, T., van Gool, A.J. & Luider, T.M. (2010). Quantitative proteomics and metabolomics analysis of normal human cerebrospinal fluid samples. *Molecular & Cellular Proteomics*, 9:2063–2075.

Stover, J.F., Pleines, U.E., Morganti-Kossmann, M.C., Kossmann, T., Lowitzsch, K. & Kempski, O.S. (1997). Neurotransmitters in cerebrospinal fluid reflect pathological activity. *European Journal of Clinical Investigation*, 27:1038–1043.

Stumvoll, M., Tataranni, P.A., Stefan, N., Vozarova, B. & Bogardus, C. (2003). Perspective in diabetes: Glucose allostasis. *Diabetes,* 52:903–909.

Stumvoll, M., Tataranni, P.A. & Bogardus C (2004). The role of glucose allostasis in type 2 diabetes. *Reviews in Endocrine & Metabolic Disorders,* 5:99–103.

Subramanian, A., Gupta, A., Saxena, S., Gupta, A., Kumar, R., Nigam, A., Kumar, R., Mandal, S.K. & Roy, R. (2005). Proton MR CSF analysis and a new software as predictors of the differentiation of meningitis in children. *NMR in Biomedicine*, 18:213–225.

Suomalainen, A. (2011). Biomarkers for mitochondrial respiratory chain disorders. *Journal of Inherited Metabolic Disease,* 34(2):277–282.

Suomalainen, A., Elo, J.M., Pietiläinen, K.H., Hakonen, A.H., Sevastianova, K., Korpela, M., Isohanni, P., Marjavaara, S.K., Tyni, T., Kiuru-Enari, S. & Pihko, H. (2011). FGF-21 as a biomarker for muscle-manifesting mitochondrial respiratory chain deficiencies: A diagnostic study. *The Lancet Neurology,* 10(9):806–818.

Sürken, M., Keller, C., Röhker, C., Ehlers, S. & Bange, F.C. (2008). Anaerobic arginine metabolism of mycobacterium tuberculosis is mediated by arginine deiminase (arcA), but is not essential for chronic persistence in an aerogenic mouse model of infection. *International Journal of Medical Microbiology,* 298:657–661.

Swahn, C.G. & Sedvall, G. (1988). CSF creatinine in schizophrenia. *Biological Psychiatry*, 23(6):586–594.

Swaminathan, S. & Rekha, B. (2010). Pediatric tuberculosis: Global overview and challenges. *Clinical Infectious Diseases*, 50(S3):S184–S194.

Sweatt, A.J., Wood, M., Suryawan, A., Wallin, R., Willingham, M.C. & Hutson, S.M. (2003). Branched-chain amino acid catabolism: Unique segregation of pathway enzymes in organ systems and peripheral nerves. *American Journal of Physiology-Endocrinology and Metabolism*, 286:E64–E76.

Sysi-Aho, M., Katajamaa, M., Yetukuri, L. & Orešič, M. (2007). Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinformatics*, 8:93–110.

Sztajnkrycer, M.D. (2002). Valproic acid toxicity: Overview and management. *Clinical Toxicology*, 40(6):789–801.

t'Kindt, R., Morreel, K., Deforce, D., Boerjan, W. & Van Bocxlaer, J. (2009). Joint GC-MS and LC-MS platforms for comprehensive plant metabolomics: Repeatability and sample pre-treatment. *Journal of Chromatography B*, 877:3752–3580.

Taher, M., Leen, W.G., Wevers, R.A. & Willemsen, M.A. (2015). Lactate and its many faces. *European Journal of Paediatric Neurology,* 20(1):3–10.

Tan, E.K., Chee, M.W.L., Chan, L.L. & Lee, Y.L. (1999). Culture positive tuberculous meningitis: clinical indicators of poor prognosis. *Clinical Neurology and Neurosurgery,* 101:157–160.

Tanaka, Y., Higashi, T., Rakwal, R., Wakida, S. & Iwahashi, H. (2007). Quantitative analysis of sulfur-related metabolites during cadmium stress response in yeast by capillary electrophoresis-mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis*, 44:608–613.

Tannenbaum, B., Tannenbaum, G.S. & Anisman, S.H. (2002). Neurochemical and behavioural alterations elicited by a chronic intermittent stressor regimen: Implications for allostatic load. *Brain Research,* 963:82–92.

Tautenhahn, R., Patti, G.J., Kalisiak, E., Miyamoto, T., Schmidt, M., Lo, F.Y., McBee, J., Baliga, N.S. & Siuzdak, G. (2010). metaXCMS: Second-order analysis of untargeted metabolomics data. *Analytical Chemistry*, 83:696–700.

Teahan, O., Gamble, S., Holmes, E., Waxman, J., Nicholson, J.K., Bevan, C. & Keun, H.C. (2006). Impact of analytical bias in metabonomic studies of human blood serum and plasma. *Analytical Chemistry,* 78(13)*:*4307–4318.

Thi, E.P., Lambertz, U. & Reiner, N.E. (2012). Sleeping with the enemy: How intracellular pathogens cope with a macrophage lifestyle. *PLoS Pathogens*, 8(3).

Thomas, S.T., VanderVen, B.C., Sherman, D.R., Russell, D.G. & Sampson, N.S. (2011). Pathway profiling in Mycobacterium tuberculosis elucidation of cholesterol-derived catabolite and enzymes that catalyze its metabolism. *Journal of Biological Chemistry*, *286*(51):43668–43678.

Thwaites, G.E., Chau, T.T.H., Stepniewska, K., Phu, N.H., Chuong, L.V., Sinh, D.X., White, N.J., Parry, C.M. & Farrar, J.J. (2002). Diagnosis of adult tuberculous meningitis by use of clinical and laboratory features. *Lancet,* 360:1287–1292.

Thwaites, G.E., Simmons, C.P., Than Ha Quyen, N., Thi Hong Chau, N., Phuong Mai, P., Thi Dung, N., Hoan Phu, N., White, N.P., Tinh Hien, T. & Farrar, J.J. (2003). Pathophysiology and prognosis in Vietnamese adults with tuberculous meningitis. *Journal of Infectious Diseases*,188:1105–1115.

Thwaites, G.E., Caws, M., Chau, T.T.H., Dung, N.T., Campbell, J.I., Phu, N.H., Hien, T.T., White, N.J. & Farrar, J.J. (2004). Comparison of conventional bacteriology with nucleic acid amplification (amplified mycobacterium direct test) for diagnosis of tuberculous meningitis before and after inception of antituberculosis chemotherapy. *Journal of Clinical Microbiology,* 42(3):996–1002.

Thwaites, G.E. & Hien, T.T. (2005). Tuberculous meningitis: many question, too few answers. The *Lancet Neurology,* 4:160–170.

Thwaites, G.E., Bang, N.D., Dung, N.H., Quy, H.T., Oanh, D.T.T., Thoa, N.T.C., Hien, N.Q., Thuc, N.T., Hai, N.N., Lan, N.T.N., Lan, N.N., Duc, N.H., Tuan, V.N., Hiep, C.H., Chau, T.T.H., Mai, P.P., Dung, N.T., Stepniewska, K., Simmons, C.P., White, N.J., Hien, T.T. & Farrar, J.J. (2005). The influence of HIV infection on clinical presentation, response to treatment, and outcome in adults with tuberculosis meningitis. *Journal of Infectious Diseases,* 195(12):2134–2141.

Thwaites, G.E. & Schoeman, J.F. (2009). Update on tuberculosis of the central nervous system: pathogenesis, diagnosis, and treatment. *Clinics in Chest Medicine,* 30(4):745–754.

Thwaites, G., Fisher, M., Hemingway, C., Scott, G., Solomon, T. & Innes, J. (2009). British Infection Society guidelines for the diagnosis of tuberculosis of the central nervous system in adults and children. *Journal of Infection,* 59:167–187.

Thwaites, G.E., van Toorn, R. & Schoeman, J. (2013). Tuberculous meningitis: more questions, still too few answers. *The Lancet Neurology,* 12*:* 999–1010.

Tian, J., Bryk, R., Itoh, M., Suematsu, M. & Nathan, C. (2005). Variant tricarboxylic acid cycle in Mycobacterium tuberculosis: identification of α-ketoglutarate decarboxylase. *Proceedings of the National Academy of Sciences of the United States of America,* 102:10670–10675.

Tinsa, F.T., Essaddam, L., Fitouri, Z., Boussetta, K., Becher, S.B. & Bousnina, S. (2010). Central system nervous tuberculosis in infants. *Journal of Child Neurology,* 25(1):102–106.

Tomiyama, A.J., O'Donovan, A., Lin, J., Puterman, E., Lazaro, A., Chan, J., Dhabhar, F.S., Wolkowitz, O., Kirschbaum, C., Blackburn, E. & Epel, E. (2012). Does cellular aging relate to patterns of allostasis? An examination of basal and stress reactive HPA axis activity and telomere length. *Physiology & Behavior,* 106:40–45.

Torok, M.E., Chau, T.T.H., Mai, P.P., Phong, N.D., Dung, N.T., Chuong, L.V., Lee, S.J., Caws, M., de Jong, M.D., Hien, T.T. & Farrar, J.J. (2008). Clinical and microbiological features of HIV-associated tuberculous meningitis in Vietnamese adults. *PLoS ONE,* 3(3):e1772.

Toshniwald, P.K. & Zarling, E.J. (1992) Evidence for increased lipid peroxidation in multiple sclerosis. *Neurochemical Research*, 17:205–207.

Tredwell, G.D., Behrends, V., Geier, F.M., Liebeke, M. & Bundy, J.G. (2011). Between-person comparison of metabolite fitting for NMR-based quantitative metabolomics. *Analytical Chemistry*, 83:8683–8687.

Tsukahara, H., Haruta, T., Todoroki, Y., Hiraoka, M., Noiri, E., Maeda, M. & Mayumi, M. (2002). Oxidant and antioxidant activities in childhood meningitis. *Life Science*, 71:2797–2806.

Tung, Y.R., Lai, M.C., Lui, C.C., Tsai, K.L., Huang, L.T., Chang, Y.C., Huang, S.C., Yang, S.N. & Hung, P.L. (2002). Tuberculous meningitis in infancy. *Pediatric Neurology,* 27:262–266.

Turnbull, D. (2011). A new biomarker for mitochondrial disease. *The Lancet Neurology*, 10(9):777–778.

Urban, M., Enot, D.P., Dallmann, G., Körner, L., Forcher, V., Enoh, P., Koal, T., Keller, M. & Deigner, H.P. (2010). Complexity and pitfalls of mass spectrometry-based targeted metabolomics in brain research. *Analytical Biochemistry*, 406:124–131.

van Batenburg, M.F., Coulier, L., van Eeuwijk, F., Smilde, A.K. & Westerhuis, J.A. (2011). New figures of merit for comprehensive functional genomics data: the metabolomics case. *Analytical Chemistry,* 83(9):3267–3274.

van de Kamp, J.L. & Smolen, A. (1995). Response of kynurenine pathway enzymes to pregnancy and dietary level of vitamin B-6. *Pharmacology Biochemistry and Behavior,* 51:753–758.

van den Berg, R.A., Hoefsloot, H.C.J., Westerhus, J.A., Smilde, A.K. & van der Werf, M.J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7*:*142–157.

van den Bos, F., Terken, M., Ypma, L., Kimpen, J.L.L., Nel, E.D., Schaaf, H.S. & Schoeman, J. (2004). Tuberculous meningitis and miliary tuberculosis in young children. *Tropical Medicine and International Health,* 9(2):309–313.

van der Geize, R., Yam, K., Heuser, T., Wilbrink, M.H., Hara, H., Anderton, M.C., Sim, E., Dijkhuizen, L., Davies, J.E., Mohn, W.W. & Eltis, L.D. (2007). A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into Mycobacterium tuberculosis survival in macrophages. *Proceedings of the National Academy of Sciences*, 104(6):1947–1952.

van der Greef, J., Martin, S., Juhasz, P., Adourian, A., Plasterer, T., Verheij, E.R. & McBurney, R.N. (2007). The art and practice of systems biology in medicine: mapping patterns of relationships. *Journal of Proteome Research,* 6:1540–1559.

van der Voet, H., de Boer, W.J., de Ruig, W.G. & van Rhijn, J.A. (1998). Detection of residues using multivariate modelling of low-level GC-MS measurements. *Journal of Chemometrics,* 12:279–294.

van Hall, G., Strømstad, M., Rasmussen, P., Jans, Ø., Zaar, M., Gam, C., Quistorff, B., Secher, N.H. & Nielsen, H.B. (2009). Blood lactate is an important energy source for the human brain. *Journal of Cerebral Blood Flow & Metabolism*, 29(6):1121–1129.

van Liempd, S., Cabrera, D., Mato, J.M. & Falcon-Perez, J.M. (2013). A fast method for the quantitation of key metabolites of the methionine pathway in liver tissue by high-resolution mass spectrometry and hydrophilic interaction ultra-performance liquid chromatography. *Analytical and Bioanalytical Chemistry*, 405:5301–5310.

van Reenen, M., Reinecke, C.J., Westerhuis, J.A. & Venter, J.H. (2016). Variable selection for binary classification using error rate *p*-values applied to metabolomics data. *BMC Bioinformatics,* 17(1):33.

van Rie, A., Beyers, N., Gie, R.P., Kunneke, M., Zietsman, L. & Donald, P.R. (1999). Childhood tuberculosis in an urban population in South Africa: burden and risk factor. *Archives of Disease in Children,* 80:433–437.

van Toorn, R. & Solomons, R. (2014). Update on the diagnosis and management of tuberculous meningitis in children. *Seminars in Pediatric Neurology,* 21:12–18.

van Toorn, R., Schaaf, H.S., Laubscher, J.A., van Elsland, S.L., Donald, P.R. & Schoeman, J.F. (2014). Short intensified treatment in children with drug-susceptible tuberculous meningitis. *The Pediatric Infectious Disease Journal,* 33(3):248–252.

van Well, G.T.J., Wieland, C.W., Florquin, S., Roord, J.J., van der Poll, T. & van Furth, A.M. (2007). A new murine model to study the pathogenesis of tuberculous meningitis. *Journal of Infectious Diseases,* 195:694–697.

van Well, G.T.J., Paes, B.F., Terwee, C.B., Springer, P., Roord, J.J., Donald, P.R., van Furth, A.M. & Schoeman, J.F. (2009). Twenty years of pediatric tuberculous meningitis: A retrospective cohort study in the Western Cape of South Africa. *Pediatrics*, 123:e1–e8.

Venta, R. (2001). Year-long validation study and reference values for urinary amino acids using a reversed-phase HPLC method. *Clinical Chemistry*, 47(3):575–583.

Venugopal, A., Bryk, R., Shi, S., Rhee, K., Rath, P., Schnappinger, D., Ehrt, S. & Nathan, C. (2011). Virulence of Mycobacterium tuberculosis depends on lipoamide dehydrogenase, a member of three multienzyme complexes. *Cell Host & Microbe*, 9(1):21–31.

Verdon, R., Chevret, S., Laissy, J-P. & Wolff, M. (1996). Tuberculous meningitis in adults: review of 48 cases. *Clinical Infectious Diseases,* 22:982–988.

Visser, D.H., Solomons, R.S., Ronacher, K., van Well, G.T., Heymans, M.W., Walzl, G., Chegou, N.N., Schoeman, J.F. & van Furth, A.M. (2014). Host immune response to tuberculous meningitis. *Clinical Infectious Diseases,* 60(2):177–187.

Volobоueva, L.A., Emery, J.F., Sun, X. & Giffard, R.G. (2013). Inflammatory response of microglial BV-2 cells includes a glycolytic shift and is modulated by mitochondrial glucose-regulated protein 75/mortalin. *FEBS Letters,* 587:756–762.

Wallis, R.S., Kim, P., Cole, S., Hanna, D., Andrade, B.B., Mauurr, M., Schito, M. & Zumla, A. (2013). Tuberculosis biomarkers discovery developments, needs, and challenges. *The Lancet Infectious Diseases*, 13(4):362–372.

Wand, M.P. & Jones, M.C. (1995). *Kernel Smoothing*, Chapman and Hall, London.

Wang, Y., Holmes, E., Nicholson, J.K., Cloarec, O., Chollet, J., Tanner, M., Singer, B.H. & Utzinger, J. (2004). Metabonomic investigations in mice infected with Schistosoma mansoni: an approach for biomarker identification. *Proceedings of the National Academy of Sciences of the United States of America,* 101(34):12676–12681.

Wang, Y., Utzinger, J., Saric, J., Li, J.V., Burckhardt, J., Dirnhofer, S., Nicholson, J.K., Singer, B.H., Brun, R. & Holmes, E. (2008). Global metabolic responses of mice to Trypanosoma brucei brucei infection. *Proceedings of the National Academy of Sciences of the United States of America* 105(16):6127–6132.

Want, E.J., Wilson, I.D., Gika, H., Theodoridis, G., Plumb, R.S., Shockcor, J., Holmes, E. & Nicholson, J.K. (2010). Global metabolic profiling procedures for urine using UPLC-MS. *Nature Protocols,* 5(6):1005–1018.

Wasim, M. & Brereton, R.G. (2004). Determination of the number of significant components in liquid chromatography nuclear magnetic resonance spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, 72:133–151.

Watson, M.A. & Scott, M.G. (1995). Clinical utility of biochemical analysis of cerebrospinal fluid. *Clinical Chemistry*, 41(3):343–360.

Watson, R.W.G., Rotstein, O.D., Parodo, J., Bitar, R., Hackam, D. & Marshall, J.C. (1997). Granulocytic differentiation of HL-60 cells results in spontaneous apoptosis mediated by increased caspase expression. *FEBS Letters*, 412:603–609.

Weckwerth, W. & Fiehn, O. (2002). Can we discover novel pathways using metabolomic analysis? *Current Opinion in Biotechnology*, 13:156–160.

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

Westerhuis, J.A., Hoefsloot, H.C.J., Smit, S., Vis, D.J., Smilde, A.K., van Velzen, E.J.J., van Duijnhoven, J.P.M. & van Dorsten, F.A. (2008). Assessment of PLSDA cross validation. *Metabolomics*, 4:81–89.

Wevers, R.A., Engelke, U., Wendel, U., de Jong, J.G.N., Gabreëls, F.J.M. & Heerschap, A. (1995). Standardized method for high-resolution ^{1}H-NMR of cerebrospinal fluid. *Clinical Chemistry*, 41(5):744–751.

White, N., Looareesuwan, S., Phillips, R., Warrell, D., Chanthavanich, P. & Pongpaew, P. (1985). Pathophysiological and prognostic significance of cerebrospinal-fluid lactate in cerebral malaria. *The Lancet*, 325:776–778.

Wikoff, W.R., Gangoiti, J.A., Barshop, B.A. & Siuzdak, G. (2007). Metabolomics identifies perturbations in human disorders of propionate metabolism. *Clinical Chemistry,* 53:2169–2176.

Wikoff, W.R., Pendyala, G., Siuzdak, G. & Fox, H.S. (2008). Metabolomic analysis of the cerebrospinal fluid reveals changes in phospholipase expression in the CNS of SIV-infected macaques. *The Journal of Clinical Investigation*, 118(7):2661.

Wilcken, B., Wiley, V., Hammond, J. & Carpenter, K. (2003). Screening newborns for inborn errors of metabolism by tandem mass spectrometry. *New England Journal of Medicine*, 348:2304–2312.

Williams, A., Koekmoer, G., Lindeque, Z., Reinecke, C. & Meyer, D. (2012). Qualitative serum organic acid profiles of HIV-infected individuals not on antiretroviral treatment. *Metabolomics*, 8:804–818.

Wishart, D.S., Lewis, M.J., Morrissey, J.A., Flegel, M.D., Jeroncic, K., Xiong, Y., Cheng, D., Eisner, R., Guatam, B., Tzur, D., Sawhney, S., Bamforth, F., Greiner, R. & Li, L. (2008). The human cerebrospinal fluid metabolome. *Journal of Chromatography B*, 871:164–173.

Wishart, D.S., Jewison, T., Guo, A.C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E. & Bouatra, S. (2012). HMDB 3.0 – the human metabolome database in 2013. *Nucleic Acids Research*, 1(41):1–7.

WHO (World Health Organization) (2006). Treatment of tuberculosis; guidelines for national tuberculosis programmes on the management of tuberculosis in children. Geneva, Switzerland. (URL: http://http://www.who.int/tb/publications/2006/en/).

WHO (World Health Organization) Global Tuberculosis Report 2014 (URL: http://www.who.int/tb/publications/global_report/).

Wolf, N.I. & Smeitink, J.A.M. (2002). Mitochondrial disorders. *Neurology,* 59(9):1402–1405.

Wolzak, N.K., Cooke, M.L., Orth, H. & van Toorn, R. (2012). The changing profile of pediatric meningitis at a referral centre in Cape Town, South Africa. *Journal of Tropical Pediatrics*, 58(6):491–495.

Wong, L.J.C., Scaglia, F., Graham, B.H. & Craigen, W.J. (2010). Current molecular diagnostic algorithm for mitochondrial disorders. *Molecular Genetics and Metabolism*, 100(2):111–117.

Wood, M. & Anderson, M. (1998). Cerebrospinal fluid and infections of the central nervous system. In: Walton Sir J, ed. Neurological infections monograph 16: major problems in neurology. London:W. B. Saunders, 1–48.

Worley, B., Halouska, S. & Powers, R. (2013). Utilities for quantifying separation in PCA/PLS-DA scores plots. *Analytical Biochemistry*, 433:102–104.

Xia, J., Psychogios, N., Young, N. & Wishart, D.S. (2009). MetaboAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Research,* 37(Web Server issue):W652–660.

Xu, F., Zou, L. & Ong, C.N. (2010). Experiment-originated variations, and multi-peak and multi-origination phenomena in derivatized-based GC-MS metabolomics. *Trends in Analytical Chemistry,* 29(3):269–280.

Yao, H., Sadoshima, S., Fujii, K., Kusuda, K., Ishitsuka, T., Tamaki, K. & Fujishima, M. (1987). Cerebrospinal fluid lactate in patients with hepatic encephalopathy. *European Neurology*, 27:182–187.

Yang, M., Lu, R., Guja, K.E., Wipperman, M.F., St. Clair, J.R., Bonds, A.C., Garcia-Diaz, M. & Sampson, N.S. (2015). Unraveling cholesterol catabolism in Mycobacterium tuberculosis: ChsE4-ChsE5 α2β2 acyl-CoA dehydrogenase initiates β-oxidation of 3-oxo-cholest-4-en-26-oyl CoA. *ACS Infectious Diseases*, *1*(2):110–125.

Yin, P., Peter, A., Franken, H., Zhao, X., Neukamm, S.S., Rosenbaum, L., Lucio, M., Zell, A., Häring, H-U., Xu, G. & Lehman, R. (2013). Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. *Clinical Chemistry*, 59:833–845.

Young, D., Stark, J. & Kirschner, D. (2008). Systems biology of persistent infection: tuberculosis as a case study. *Nature,* 6:520–528.

Youssef, F.G., El-Sakka, H., Azab, A., Eloun, S., Chapman, G.D., Ismail, T., Mansour, H., Hallaj, Z. & Mahoney, F. (2004). Etiology, antimicrobial susceptibility profiles, and mortality associated with bacterial meningitis among children in Egypt. *Annals of Epidemiology,* 14:44–48.

Youssef, F.G., Afifi, S.A., Azab, A.M., Wasfy, M.M., Abdel-Aziz, K.M., Parker, T.M., Oun, S.A., Jobanputra, N.N. & Hajjeh, R.A. (2006). Differentiation of tuberculous meningitis from acute bacterial meningitis using simple clinical and laboratory parameters. *Diagnostic Microbiology and Infectious Disease,* 55:275–78.

Yu, K., Sheng, G., Sheng, J., Chen, Y., Xu, W., Liu, X., Cao, H., Qu, H., Cheng, Y. & Li, L. (2007). A metabonomic investigation on the biochemical perturbation in liver failure patients caused by Hepatitis B virus. *Journal of Proteome Research*, 6:2413–2419.

Zelena, E., Dunn, W.B., Broadhurst, D., Francis-McIntyre, S., Carroll, K.M., Begley, P., O'Hagan, S., Knowles, J.D. & Halsall, A. (2009). Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Analytical Chemistry,* 81:1357–1364.

Zhang, A., Sun, H., Wang, P., Han, Z. & Wang, X. (2012). Modern analytical techniques in metabolomics analysis. *Analyst*, 137:293–300.

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*, 12:4642–4649.

Zhou, A., Ni, J., Xu, Z., Wang, Y., Zhang, W., Wu, W., Lu, S., Karakousis, P.C. & Yao, Y.F. (2015). Metebolomics specificity of tuberculosis plasma revealed by ^{1}H NMR spectroscopy. *Tuberculosis*, 95:294–302.

Zimmerman, G., Njunting, M., Ivens, S., Tolner, E., Behrens, C.J., Gross, M., Soreq, H., Heinemann, U. & Friedman, A. (2008). Acetylcholine-induced seizure-like activity and modified cholinergic gene expression in chronically epileptic rats. *European Journal of Neuroscience*, 27:965–975.

Zysset-Burri, D.C., Bellac, C.L., Leibm, S.L. & Wittwerm, M. (2013). Vitamin B6 reduces hippocampal apoptosis in experimental pneumococcal meningitis. *BMC Infectious Diseases*, 13:393–407.

“The woods are lovely, dark and deep,

But I have promises to keep,

And miles to go before I sleep,

And miles to go before I sleep.”

~ Robert Frost

ANNEXURE 1:

Supplementary information:

A hypothetical astrocyte–microglia lactate shuttle derived from a ^{1}H NMR metabolomics analysis of cerebrospinal fluid from a cohort of South African children with tuberculous meningitis

A1.1 Sampling and clinical information on patients and controls

The three experimental groups used in this study were: 1) South African patients with confirmed TB meningitis, 2) non-meningitis South African controls (SA_Controls), and 3) neurological controls from the Netherlands (NL_Controls). The first two groups comprised children between the ages of 6 months and 12 years, all of whom were originally suspected meningitis cases and referred from local clinics to the paediatrics unit at Tygerberg Hospital in the Western Cape province of South Africa. In most of these cases, a broad range of non-specific treatments (e.g., broad-spectrum antibiotics, analgesics and anti-inflammatories) were initiated prior to admission. These individuals were given a thorough assessment by a paediatric neurologist at Tygerberg Hospital, which involved an extensive description of the clinical background, including analysis of a CSF sample obtained through a lumbar puncture. The main physical presentations and clinical symptoms of these cases, as testified by parents and observed by the respective clinicians (see Table A1-1), were compatible with non-CNS indications of TB outside the CNS, typically observed in small TB-infected children from the region. As part of the diagnostic process a detailed inspection of the clinical chemistry of the CSF was conducted (see Table A1-2) and yielded a report describing the count and type of cells present, particularly immune response-related cells, as well as a measure of CSF protein (mostly high: >1 g/L) and glucose (mostly low: <2.2 mmol/L or CSF:blood glucose ratio <50%) levels. Most, but not all, of the children received a wide range of medication uniquely based upon their presentation of clinical symptoms, which included: TB-specific drugs, antibiotics and analgesics, anti-inflammatory drugs, sedatives and drugs directed to specific symptoms like seizures or for directed treatment (e.g. diuretics, laxatives, muscle relaxants, anti-psychotics), as well as multivitamins and other supplements summarized in Table A1-3.

Approximately 1 mL of the CSF samples was stored at –80^{o}C for this metabolomics study, for which informed consent from the respective parents was obtained. Following polymerase

chain reaction (PCR), clinical chemistry and culture analysis of the CSF, an informed diagnosis was made where those in the patient group were confirmed positive for TB meningitis. The South African controls were confirmed negative for any form of meningitis, despite being ill with symptoms indicative of meningitis. Ethical and practical considerations limit the availability of obtaining healthy controls, which could partially be overcome through a comparative analysis of a second control group involving CSF collected from untreated individuals from the Radboud University Medical Centre in Nijmegen, the Netherlands. The Netherlands non-TBM control group (the third experimental group) consists of CSF samples from age-matched patients who were suspected to suffer from a neurometabolic disease. After appropriate and in depth investigations no clinical or biochemical evidence was found for such diagnosis in any of these patients. The only exclusion criterion applied to all cases was HIV co-infection.

Table A1-1: General clinical symptoms of TBM cases in duration (days) prior to admission to hospital, evidence of TB outside the CNS and stage of TBM.

Ref	TBM stage on admission	Chest X-ray evidence of PTB	Other evidence of TB outside CNS	Clinical symptoms (duration in days prior to admission to hospital)			
				Fever	Vomiting	Seizures	Poor feeding
138	**1**	yes	no	3	3		3
140	**1**	not done	no				
193	**1**	no	Positive gastric	1	1		2
146	**2a**	yes	not done	2	4	2	10
148	**2a**	no	negative	14			14
161	**2a**	yes	negative	6			
162	**2a**	not done	negative	10	1	1	10
191	**2a**	yes	no		14		7
229	**2a**	yes	Positive gastric	8		1	
139	**2b**	no	no	1		1	1
208	**2b**	no	no	1	1		5
216	**2b**	yes	Positive gastric	3	2		
227	**2b**	yes	no	12	1	1	14
228	**2b**	no	no	1		1	
144	**3**	no	Positive gastric	7			
195	**3**	no	no	7	2		7
211	**3**	yes	Pos sputum	3	1	5	3
221	**3**	no	no		1		
226	**3**	no	no	21	21		
209	**3**	no	Trachea aspirate Mtb	14			14
239	**3**	no	Mandible pus swab Mtb	7	7		

Table A1-2: Clinical chemistry of CSF describing elevated immuno-cells, decreased glucose and elevated proteins in TBM cases compared to SA_Control cases.

	Ref.	Erythrocytes (cells/µL)	Leucocytes (cells/µL)	PMN's (cells/µL)	Lymphocytes	Protein (g/L)	Glucose (mmol/L)
TBM	138	1320	344	96	238	1.49	1.2
	139	39		7	17	1.07	2.1
	140	55	133	28	105	3	2.2
	144	1	40	0	40	0.76	2.8
	146	14	440	30	410	2.1	0.6
	148	28	199	6	193	2	0.5
	161	0	29	0	29	1.24	0.5
	162	4	172	15	157	0.72	1.4
	191	0	27	0	27	3	0.2
	193		92	12	80	0.94	1.6
	195	1875	349	2	347	1.09	2
	208	0	11	0	11	0.96	1.5
	209	155	30	0	30		
	211	0	0	0	0	1.2	3.9
	216	83	270	33	237	0.94	1.1
	221	5	32	6	26	3	2.5
	226	220	7	2	5	0.21	
	227	462	9	0	9		
	228	0	0	0	0	0.62	5.1
	229	100	0	0	21	0.87	2.7
	239	0	1	1	0	1.57	3
SA_Control	C143	0	0	0	0	0.15	3.8
	C149	0	0	0	0	0.24	4.2
	C151	0	0	0	0	0.14	3.7
	C159	1	2	0	2	0.14	4.0
	C171	0	2	0	2	0.3	3.7
	C173	0	0	0	0	0.11	3.9
	C184	0	0	0	0	0.32	4.6
	C190	0	0	0	0	0.11	3.9
	C192	0	0	0	0	0.11	3.3
	C199	0	0	0	0		4.1
	C203	0	2	0	2	0.25	4.0
	C204	0	0	0	0	0.09	4.6
	C206	0	0	0	0	0.27	5.2
	C212	0	1	0	1	0.18	3.1
	C220	0	0	0	0	0.18	3.7
	C222	0	0	0	0	0.25	3.6
	C232	1	1	0	1	0.34	4.0
	C236	0	34	0	34	0.32	4.6
	C240	0	0	0	0	0.23	4.3

Table A1-3: Summary of total, and type of, medication administered to the TBM and SA_Control cases, analysed in this study, prior to admission to hospital. Medications classified as: (i) **TB-specific** (Pyrazinamide, Rifampicin, Ethionamide, Ethambutol, Isoniazid, Rimcure); (ii) **Antibiotic** (Cephtriaxone, Acyclovir, Amoxicillin, Vancomycin, Ciprofloxacin, Cloxacillin, Coamoxiclav, Ampicillin, Metronidazole, Penicillin, Flucloxacillin); (iii) **Analgesic** (Paracetamol); (iv) **Anti-inflammatory** (Prednisone, Ibuprofen, Dexamethasone, Solu-Medrol); (v) **Anti-seizure** (Phenobarbital); (vi) **Sedative** (Midazolam, Ketamine, chloral hydrate, Diazepam); and (vii) **Other** (Lasix (diuretic), mannitol (diuretic), Sorbitol (laxative), Allergex, Diamox (carbonic anhydrase inhibitor), Albendazole (Anti-Worm), Valeron (muscle relaxant), Haloperidol (anti-psychotic), Nystatin (anti-fungal), Illidian Drops (decongestant), multivitamin and other supplements (pyridoxine, zinc, potassium chloride, vitamin A, glycerine, lactulose)).

	Ref.	Total	TB-	Antibioti	Analgesi	Anti-	Anti-	Sedativ	Othe
TBM	138	0							
	140	0							
	144	0							
	195	0							
	229	0							
	146	1		1					
	161	5	4			1			
	191	5	4			1			
	211	6	4	1		1			
	227	6	3	1		1		1	
	228	6	4	2					
	148	7	4	1	1	1			
	209	7	4	1	1	1			
	162	8	4	2		1			1
	221	8	4	1					3
	193	9	4	2	1	1		1	1
	139	10	4	2		1	1		2
	208	10	4	2	1	2			1
	216	10	4	1	1	1			3
	226	11	4	1	1	2		1	2
	239	14	3	4	1	3			2
SA_Control	c14	0							
	c14	2	1	1					
	c15	2			1				1
	c20	3		2					1
	c15	4		2	1			1	
	c19	4		1				1	2
	c21	4		2	1				1
	c23	4		1	1		1		1
	c24	4		2	1			1	
	c17	5		3	1				1
	c19	6		2	1			2	1
	c20	6	3	1	1	1			
	c18	7	4	1	1	1			
	c17	9	4	3		1			1
	c19	9	2	1					6
	c20	9		2	1	1		1	4
	c22	10	3	3		1	1		2
	c22	10	4	2	1			1	2
	c23	10		5	1	1			3

A1.2 Experimental aspects of the NMR protocol

All CSF samples were stored at –80°C prior to analysis and transported to the NMR facility at the Laboratory for Genetic, Endocrine and Metabolic Diseases at the Radboud University Medical Centre in Nijmegen, the Netherlands, where the NMR analyses were done. Sample preparation followed the standard operating procedure (SOP) as set up by the Nijmegen laboratory (Engelke *et al.* 2005; Engelke *et al.* 2006; Wevers *et al.* 1995). The first step involved cleaning the Sartorius Centrisart®1 10-kDa centrifugal unit by pre-centrifugation twice (washing involved centrifugation with distilled water for 10 min at 3000 rpm in order to remove excess glycerol in the filter membrane), followed by centrifugation of approximately 1 mL CSF sample for 15 min at 3000 rpm; 700 µL ultrafiltrate was collected (in those cases where there was insufficient volume the ultrafiltrate was diluted to 700 µL final volume with distilled water and the dilution factor noted) and 20 µL internal standard (20.2 mM trimethylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP, sodium salt; Aldrich) in deuterium oxide) added, after which the pH was adjusted to 2.50 ± 0.05 using concentrated HCl. Finally, 650 µL of the prepared sample was analyzed in a 5-mm NMR tube.

Each CSF sample was measured at 500 MHz on a Bruker DRX Avance spectrometer equipped with a triple-resonance inverse (TXI) $^{1}H\{^{15}N, ^{13}C\}$ probe head and equipped with x, y, z gradient coils. ^{1}H spectra were acquired as 128 transients in 32k data points with a spectral width of 6002 Hz. The sample temperature was 298 K and the H_2O resonance was pre-saturated by single-frequency irradiation during a relaxation delay of 10 s, and a 90° excitation was used. Shimming of the sample was performed automatically on the deuterium signal. The resonance line widths for TSP and metabolites were <1 Hz. A π/2-shifted sine-bell window function was applied to the free induction decay. Fourier transformation was performed after zero-filling to 64k data points. The phase and the baseline were corrected manually. Software used was: Bruker Topspin(V3.1) for data pre-processing and Bruker AMIX(V3.9.12) for binning and quantification (Ellinger *et al.* 2013). A final bucket data matrix was created scaled relative to the total spectrum intensity and final metabolite concentrations quantified as µmol/L, using the TSP peak as a reference. A Bruker spectral library database and an in-house pure compound database were used to identify NMR peaks.

A1.3 Data pre-processing

Some of the variation that can occur in NMR spectra includes: spectral noise, changes in chemical shifts and peak widths due to fluctuations in pH, instrument and/or temperature. This variation can be minimized by spectral binning, which is a common approach in NMR-based metabolomics (Wishart 2008) and involves dividing the NMR spectra into regions, also known as frequency windows, or 'buckets', of defined width. The norm with NMR-based metabolomics studies involves segmenting the NMR spectrum into set width rectangular buckets (e.g., 0.005 ppm bin width (Viant *et al.* 2005; Lin *et al.* 2007; Sinclair *et al.* 2009), 0.04 ppm bin width (Pears *et al.* 2005; Waters *et al.* 2001), and 0.02 ppm bin width (Pan *et al.* 2007)). The total cumulative peak intensity of these buckets is summed and each bucket is then represented as a fraction relative to this total, effectively smoothing (normalizing) the spectrum. While this smoothing serves to stabilize and reduce the variation between spectra, accounting for random variation, dilution and bulk mass differences, it also reduces resolution and can introduce other errors through the rigid and inaccurate definition of the bin boundaries that can produce artefacts (Craig *et al.* 2006; Ellinger *et al.* 2013; Powers 2009). To overcome the possibility of inducing such error, variable-sized bucketing was used in this study which, based upon prior spectral knowledge of typical CSF profiles and an understanding of the clinical information associated with the experimental cases, utilized user-defined bin widths that are unique and specific to the data set examined here. Based upon this customized spectral binning, each spectrum was divided into 110 buckets, each associated with one specific chemical substance (variable), unless overlap with more than one substance occurs within a specific bucket. These data were transposed into an n x m matrix, where n is the number of cases (rows) and m is the number of variables (columns) – in our case m =110 – and subsequently log-transformed. Log transformation of the binned data causes the variance of the spectral intensity in each bin to become more constant, allowing for more efficient multivariate analyses (Purohit *et al.* 2004; Lin *et al.* 2007; Viant *et al.* 2005). In the case of untransformed data, the bins with the highest spectral intensity (largest variance) will dominate the analysis. The NMR region representing the suppressed water signal (4.67–4.96 ppm) was excluded from the data matrix.

Table A1-4: PCA power values of SA_Controls vs TBM and NL_Controls vs TBM cases for all 109 variables, with a cutoff value of 0.5, illustrating quantitatively that lactate and glucose were the two dominant discriminating metabolites.

Variables (SA Controls vs	Power		Variables (NL Controls vs	Power
^{13}C Lactate (1.27)	0.931390		lactate (1.41)	0.918325
glucose (3.86)	0.878166		glucose (3.24)	0.886187
lactate (4.36)	0.872698		glucose (3.86)	0.885595
glucose (3.71)	0.867672		glucose (3.48)	0.874681
lactate (1.41)	0.866338		lactate (4.36)	0.872905
glucose (3.24)	0.861590		glucose (4.63)	0.871041
^{13}C Lactate (1.53)	0.859754		glucose (3.38)	0.868400
^{13}C Lactate (1.283)	0.857551		glucose (3.71)	0.867712
glucose (3.48)	0.849267		glucose (3.21)	0.861035
glucose (3.38)	0.847711		glucose (5.22)	0.847384
glucose (5.22)	0.844244		glucose/glycerol (3.43)	0.800225
glucose (3.21)	0.838305		Pyrazinamide (8.81)	0.763011
glucose (4.63)	0.824405		Pyrazinamide (8.75)	0.731714
glucose/glycerol (3.43)	0.822220		Pyrazinamide (9.18)	0.713093
alanine (1.51)	0.781326		alanine (1.51)	0.687848
acetate (2.08)	0.759398		3OHisovalerate/threonine (1.33)	0.672409
valine 1.04)	0.758630		^{13}C Lactate (1.283)	0.635886
lysine (1.73)	0.748802		^{13}C Lactate (1.27)	0.621801
1.0255	0.723804		creatinine (3.13)	0.618343
creatine (3.05)	0.685460		valine/isoleucine (1.01)	0.613439
choline (3.19)	0.678086		1.93715	0.600289
3.03299999	0.668447		8.70874977	0.600242
1.93715	0.668090		phenylalanine (7.42)	0.578138
myo-inositol (4.05)	0.662570		8.77799988	0.573148
7.96000004	0.662132		myo-inositol (3.285)	0.563730
3.56700003	0.658366		^{13}C Lactate (1.53)	0.548962
1.29099995	0.652387		3.14499998	0.546544
Pyrazinamide (9.18)	0.631420		valine 1.04)	0.546472
1.27685004	0.628092		myo-inositol (4.05)	0.535355
valine/isoleucine (1.01)	0.624230		tyrosine (7.19)	0.532997
7.99500012	0.620387		9.19474983	0.526925
8.66499996	0.613486		3.59799993	0.524536
Pyrazinamide (8.81)	0.612643		Isoniazid (8.94)	0.522664
formate (8.25)	0.608682		7.96000004	0.521607
Pyrazinamide (8.75)	0.597625		8.09674978	0.512562
creatine (4.10)	0.586217			
2.00975001	0.582802			
7.77749991	0.579755			
3OHbutyrate (2.53)	0.574481			
myo-inositol (3.285)	0.573663			
creatinine (3.13)	0.564981			
tyrosine (7.19)	0.532237			
Acyclovir (5.55)	0.527150			
8.09674978	0.505257			
3.14499998	0.505217			

Table A1-5: Quantitative data for all identified variables, indicating the most important metabolites in discriminating between TBM and non-TBM for both SA_Controls vs TBM and NL_Controls vs TBM cases; which includes: PLS-DA VIP values, fold change d-values and p-values for the t-test and from the cross-validation of the PLS model. [the chemical shift, in ppm, of each identified metabolite given in brackets].

SA Controls vs	PLS	Fold	t-	Valid		NL Controls	PL	Fold	t-	Valid
	VIP	d-	p-	p-			VI	d-	p-	p-
lactate (1.41)	1.63	3.172	<0.	<0.0		lactate (1.41)	1.7	2.83	<0.	<0.0
lactate (4.36)	1.63	3.032	<0.	<0.0		lactate (4.36)	1.7	2.58	<0.	<0.0
^{13}C lactate (1.27)	1.53	2.931	<0.	<0.0		glucose	1.6	0.30	<0.	<0.0
^{13}C lactate (1.53)	1.47	2.536	<0.	<0.0		glucose	1.5	0.49	<0.	<0.0
^{13}C lactate (1.283)	1.47	2.817	<0.	<0.0		glucose	1.5	0.61	<0.	<0.0
glucose (3.86)	1.42	0.571	<0.	<0.0		glucose	1.5	0.54	<0.	<0.0
glucose/glycerol	1.41	0.582	<0.	<0.0		glucose	1.5	0.50	<0.	<0.0
glucose (3.71)	1.41	0.526	<0.	<0.0		glucose	1.5	0.50	<0.	<0.0
glucose (3.24)	1.40	0.467	<0.	<0.0		glucose	1.5	0.49	<0.	<0.0
glucose (3.48)	1.39	0.464	<0.	<0.0		glucose	1.5	0.48	<0.	<0.0
choline (3.19)	1.39	4.736	<0.	<0.0		^{13}C lactate	1.4	3.05	<0.	<0.0
alanine (1.51)	1.38	3.351	<0.	<0.0		glucose/glyc	1.4	0.65	<0.	<0.0
glucose (5.22)	1.38	0.440	<0.	<0.0		alanine	1.4	3.44	<0.	<0.0
glucose (3.38)	1.37	0.455	<0.	<0.0		valine (1.04)	1.3	3.45	<0.	<0.0
glucose (3.21)	1.37	0.482	<0.	<0.0		^{13}C lactate	1.3	2.88	<0.	<0.0
glucose (4.63)	1.33	0.440	<0.	<0.0		creatinine	1.3	0.57	<0.	<0.0
lysine (1.73)	1.29	2.053	<0.	<0.0		$DMSO_2$	1.2	0.32	<0.	<0.0
valine/isoleucine	1.24	2.940	<0.	<0.0		^{13}C lactate	1.2	2.45	<0.	<0.0
valine (1.04)	1.17	2.867	<0.	<0.0		lysine (1.73)	1.1	1.86	<0.	<0.0
glucose/glycerol	1.15	0.704	<0.	<0.0		valine/isoleuc	1.1	2.38	<0.	0.00
creatine (3.05)	1.11	1.502	<0.	<0.0		creatinine	1.0	0.60	<0.	<0.0
isoleucine/leucine	1.07	1.784	<0.	<0.0		myoinositol	1.0	0.61	<0.	0.00
pyruvate (1.56)	1.02	1.915	<0.	0.00		tyrosine/medi	1.0	1.67	<0.	0.00
pyruvate (2.37)	1.00	1.777	<0.	0.00		ethanol	1.0	4.48	<0.	0.00
formate (8.25)	0.95	1.948	<0.	<0.0		myo-inositol	0.9	0.60	0.0	0.00
creatinine (3.13)	0.94	0.707	<0.	<0.0		tyrosine	0.9	1.53	0.0	0.00
threonine (1.34)	0.93	1.707	<0.	<0.0		phenylalanin	0.8	1.89	0.0	0.00
tyrosine (7.19)	0.91	1.805	0.0	<0.0		isoleucine/le	0.7	1.40	0.0	0.00
carnitine (3.22)	0.87	1.379	0.0	0.00		phenylalanin	0.7	1.46	0.0	0.00
2-oxoglutarate	0.82	2.190	0.0	0.00		glucose/glyc	0.7	0.82	0.0	0.00
citrate (3.00)	0.81	1.234	0.0	0.00		choline	0.7	1.50	0.0	0.00
phenylalanine	0.78	1.739	0.0	0.00		creatine	0.6	1.22	0.0	0.01
betaine/myoinosit	0.76	1.309	0.0	<0.0		acetoacetate	0.6	0.91	0.0	0.01
acetate (2.08)	0.75	0.744	0.0	0.04		pyruvate	0.5	1.21	0.0	0.04
SA Controls vs	PLS	Fold	t-	Valid		NL Controls	PL	Fold	t-	Valid
	VIP	d-	p-	p-			VI	d-	p-	p-
myo-inositol	0.75	0.793	0.0	0.00		pyruvate	0.5	1.20	0.0	0.06
phenylalanine	0.71	1.775	0.0	0.00		creatine	0.3	0.87	0.2	0.44
succinate (2.66)	0.63	1.830	0.0	0.00		glutamine	0.3	0.90	0.2	0.11
$DMSO_2$ (3.14)	0.62	0.637	0.0	0.01		citrate (2.84)	0.3	0.85	0.2	0.33
tyrosine/medicatio	0.60	1.327	0.0	0.05		3-OHbutyrate	0.3	0.97	0.2	0.26
myo-inositol	0.55	0.819	0.0	0.07		mannose	0.3	0.21	0.2	0.08
3-	0.51	1.321	0.0	0.00		citrate (2.97)	0.3	1.06	0.2	0.24
citrate (2.97)	0.50	1.149	0.0	0.09		citrate (3.00)	0.3	1.05	0.2	0.15
acetoacetate	0.45	1.227	0.1	0.06		carnitine	0.2	1.10	0.3	0.33
glutamine/medicat	0.43	1.103	0.1	0.27		acetate	0.2	0.89	0.3	0.53
creatinine (4.29)	0.38	0.971	0.1	0.00		citrate (2.81)	0.2	1.02	0.3	0.31
ethanol (1.18)	0.38	1.522	0.1	0.17		formate	0.2	1.13	0.4	0.22
3-OHbutyrate	0.35	1.080	0.2	0.10		glutamine/me	0.2	1.04	0.4	0.83
acetone (2.22)	0.19	0.569	0.5	0.43		succinate	0.1	0.96	0.6	0.20

glutamine (2.47)	0.15	1.059	0.5	0.92		3-OHbutyrate	0.1	0.73	0.6	0.44
3-OHbutyrate	0.13	0.935	0.6	0.70		3-	0.0	0.93	0.7	0.93
mannose (5.17)	0.12	0.776	0.6	0.51		2-	0.0	0.93	0.7	0.84
ethanol (3.64)	0.07	1.128	0.7	0.59		acetone	0.0	0.57	0.8	0.84
citrate (2.81)	0.05	0.996	0.8	0.95		betaine/myoi	0.0	0.96	0.9	0.86
creatine (4.10)	0.04	0.982	0.8	0.44		threonine	0.0	1.01	0.9	0.81
citrate (2.84)	0.00	1.009	0.9	0.74		ethanol	0.0	1.08	0.9	0.95

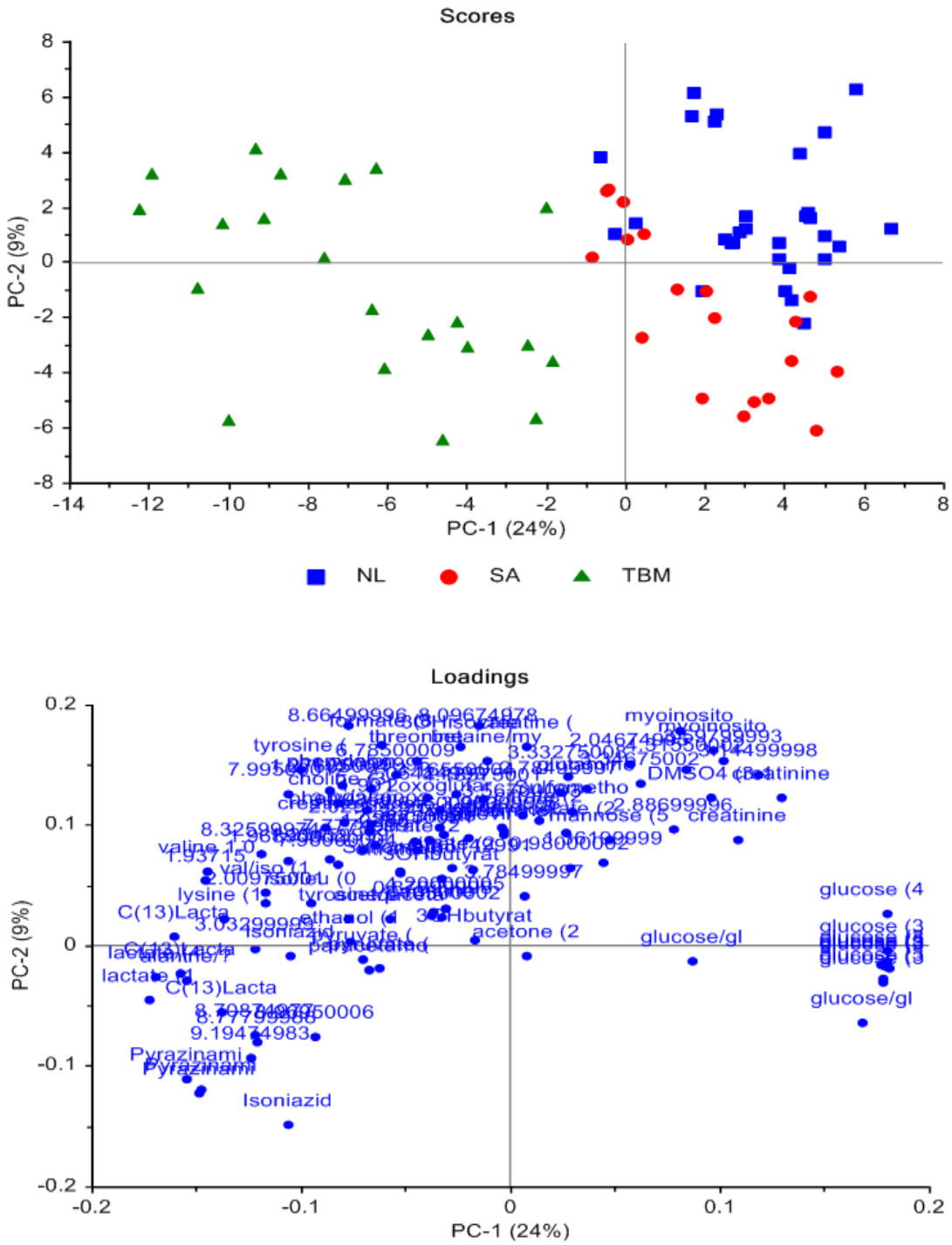


Figure A1-1: PCA scores plot (above) showing natural separation between TBM and non-TBM cases, with overlapping between control groups, and associated loadings plot (below) for all 109 variables.

A1.4 A conceptual model of the AMLS hypothesis

A1.4.1 Conceptual modelling

Model development provides a way of integrating biological insights, and of combining the interpretation of experimental results and existing views in the literature into functional contexts, resulting in a wide array of possible models [e.g., conceptual, mathematical, computational, and informational], depending on the purpose of the modelling initiative (Goel *et al.* 2006). In their review, Goel *et al.* emphasize that modelling often requires that purely descriptive approaches to biology be accompanied by the ability to make reliable, quantitative predictions of the responses of cells or organisms to experimentally untested situations. In this SI we present a conceptual model on Mtb for discussion purposes. Conceptual models differ from experiment- and data-based models as they are formed by a mental process, resulting in visual representations which are constructed to facilitate understanding. They incorporate diverse inputs that can relate to each other through a common framework and have predictive value to promote new views or hypotheses. Our conceptual model of the AMLS hypothesis, formulated for TBM, is based on the interpretation of the clinical profile of the TBM and control cases, the outcomes of the NMR metabolomics investigation, as well as on an interpretation of CNS metabolism, which we derived from contemporary knowledge and paradigms prevailing in the scholarly literature. Central nervous system (CNS) metabolism and regulation cover a vast field, and for the present purpose we only focused on: (1) astrocyte and microglia activation in response to TB infection of the meninges; (2) cell–cell communication; (3) metabolic coupling, and (4) effector isoforms.

A1.4.2 Astrocyte and microglia activation

Astrocytes and microglia act as immune surveillant cells of the nervous system to protect the integrity of the CNS. Under normal physiological conditions, astrocytes buffer ions and biogenic amines provide trophic and survival factors for neurons and oligodendrocytes, stabilizing and maintaining CNS homeostasis. Likewise, microglia produce neuro-specific factors that are important in neuronal survival and thus concomitantly contribute to homeostasis and neuroplasticity in the brain (Merrill and Jonakait 1995). During endogenous or exogenous stress conditions in the CNS – induced, for example, by brain lesions (stroke, head injury), chronic diseases (multiple sclerosis, Alzheimer's or Parkinson's diseases) or infectious states (viral, bacterial or, specifically, tuberculous meningitis) – microglia and astrocytes become activated and de-regulated relative to the homeostatic state (Chen and Swanson 2003). The reaction of astrocytes to brain injury is similar in some respects to the

inflammatory response of peripheral tissues. Inflammatory activation results in the multistep process of leukocyte invasion in the brain, with leukocytes, in particular the presence of lymphocytes in the CSF, being the diagnostic hallmark of meningitis, as seen in TBM patients in the present study (Table A1-2). Simultaneously, microglial cells function as the resident macrophages of the brain parenchyma; they also share many, if not all, of the properties of macrophages in other tissues (Nareika *et al.* 2005), such as production of a variety of cytokines and chemokines on activation. In addition, it has become progressively clear that astrocytes and microglia, in conjunction with neurons and endothelial cells of microvessels, are organized into well-structured neurovascular units (Abbott *et al.* 2006), which are involved in the regulation of cerebral blood flow and CNS homeostasis. As homeostatic signalling within the neurovascular unit underlies normal brain function, we regard it as conceptually sound to postulate that perturbed signalling and reordered metabolic functioning might become operational under conditions where stressors, such as pathogenic organisms, invade the CNS. This led us to the formulation of the AMLS hypothesis as outlined in the main text.

The model shown in Figure A1-2 aims to address some conceptual aspects in relation to this hypothesis. We recognize that the complex CNS metabolism and distinct regulatory metabolic pathways exclude incorporation of important mechanistic features in the model. We nonetheless attempted to integrate several conceptual paradigms and our own metabolomics information into the conceptual framework. We believe that the diverse inputs sufficiently relate to each other to stimulate new views on TBM, critical assessment of the AMLS hypothesis and, we hope, pave the way for its experimental validation or improvement.

A1.4.3 Cell–cell communication

Non-neuronal glial cells (e.g., astrocytes, microglia and oligodendrons) have unique cellular structures and phenotypic characteristics to sense their surroundings and respond dynamically to changes in their microenvironment (reviewed by Allaman *et al.* 2011). They constitute the macrophage complement of the CNS; the microglial uptake of Mtb is facilitated by receptor mediation (Peterson *et al.* 1995). Their subsequent transformation to a reactive state with a high energy demand, known as microglial activation, includes morphological changes and appears to produce diverse and dynamic transcriptional and non-transcriptional responses, which have clear functional consequences for the communication network between glial cells and neurons (Allaman *et al.* 2011) through direct membrane contact, soluble signalling mediators or exosomes.

A1.4.4 Metabolic coupling

Increased physiological activation of neural functions is accompanied by additional energy demand, provided through glucose utilization in metabolic coupling between astrocytes and neurons (Pellerin and Magistretti 1994). This coupling became expressed in a hypothetical "astrocyte–neuron lactate shuttle" (Pellerin *et al.* 1998), where the astrocytes perform glycolysis under aerobic conditions, export the metabolic end-products (pyruvate and lactate) for uptake by neurons, and stimulate the neuronal TCA cycle and oxidative phosphorylation to generate large amounts of ATP. Conceptually, the "Reverse Warburg Effect" (Pavlides *et al.* 2009), proposed as operating in tumour cells, has been interpreted to be analogous to the astrocyte–neuron metabolic coupling (Pavlides *et al.* 2010), where the astrocytes resemble cancer-associated fibroblasts and the neuron the epithelial tumour cells, resulting in their higher proliferative capacity.

A1.4.5 Effector isoforms

A wide array of isoenzymes and transporters (chaperones) participate in and affect neural metabolism. They have a crucial role as catalysts and gatekeepers of the metabolic bidirectional cell–cell interactions. Their respective catalytic capacity (isoenzymes), or their up- or down-regulation (transcriptional or otherwise), have an important role in homeostasis or under conditions of increased physiological activation of specific neural functions. (1) The pyruvate dehydrogenase complex (PDC) functions as the rate-limiting point of entry for pyruvate into the TCA; the phosphorylation status of the pyruvate dehydrogenase alpha (PDHα) subunit of the PDC has been shown to distinguish the metabolic phenotypes between astrocytes and neurons (Halim *et al.* 2010; Xing *et al.* 2012). (2) Five lactate dehydrogenase (LDH) isoenzymes catalyze the interconversion of lactate and pyruvate – with LDH-I (neuron located) being inhibited by pyruvate – and preferentially drive the reaction towards the production of pyruvate. The LDH-5 subunit (astrocyte located) is present in glycolytic tissues, favouring the formation of lactate from pyruvate (Bittar *et al.* 1996). (3) Glucose is transported across cellular membranes by specific glucose transporters (GLUTs), which are members of the SLC2 family; GLUT1, GLUT3 and GLUT4 are the most abundantly expressed in the brain (Duelli and Kuschinsky 2001). The solute carriers (SLCs) are the largest group of transporters encoded by the human genome (Fredriksson *et al.* 2008), ranging from transporters of inorganic ions, metabolites such as organic acids, amino acids, fatty acids, sugars and purines, and regulatory substances such as neurotransmitters. (4) Monocarboxylate transporters (MCTs, reviewed by Pierre and Pellerin 2005), are composed of 14 members of the SLC16 gene family and are proton-linked membrane carriers involved in the transport of monocarboxylates such as lactate and

pyruvate, as well as ketone bodies. In the brain, MCT-1 is expressed by endothelial cells of microvessels as well as by astrocytes. MCT-4 expression appears to be specific for astrocytes, whereas the predominant neuronal isoform is MCT-2. (5) Glutamate is the main excitatory amino acid in the brain, and an ordered coupling between glutamatergic neurons and surrounding glia cells is fundamental for excitatory transmission. Different subtypes of high-affinity glutamate transporter systems participate in this coupling; the EAAT1/GLAST and GLT-1 participate in neuronal protection against excitotoxicity *in vivo* (reviewed by López-Bayghen and Ortega 2011). (6) The brain is unique in that it expresses two separate branched-chain aminotransferase (BCAT) isoenzymes: a peripheral form [mitochondrial (BCATm)] and a form unique to cerebral tissue, placenta and ovaries [cytosolic (BCATc)]. In culture, BCATm functions predominantly in astrocytes and BCATc in neuronal cultures. Metabolic studies indicated that BCAAs promote the efflux of glutamine from astrocytes (Hutson *et al.* 1998).

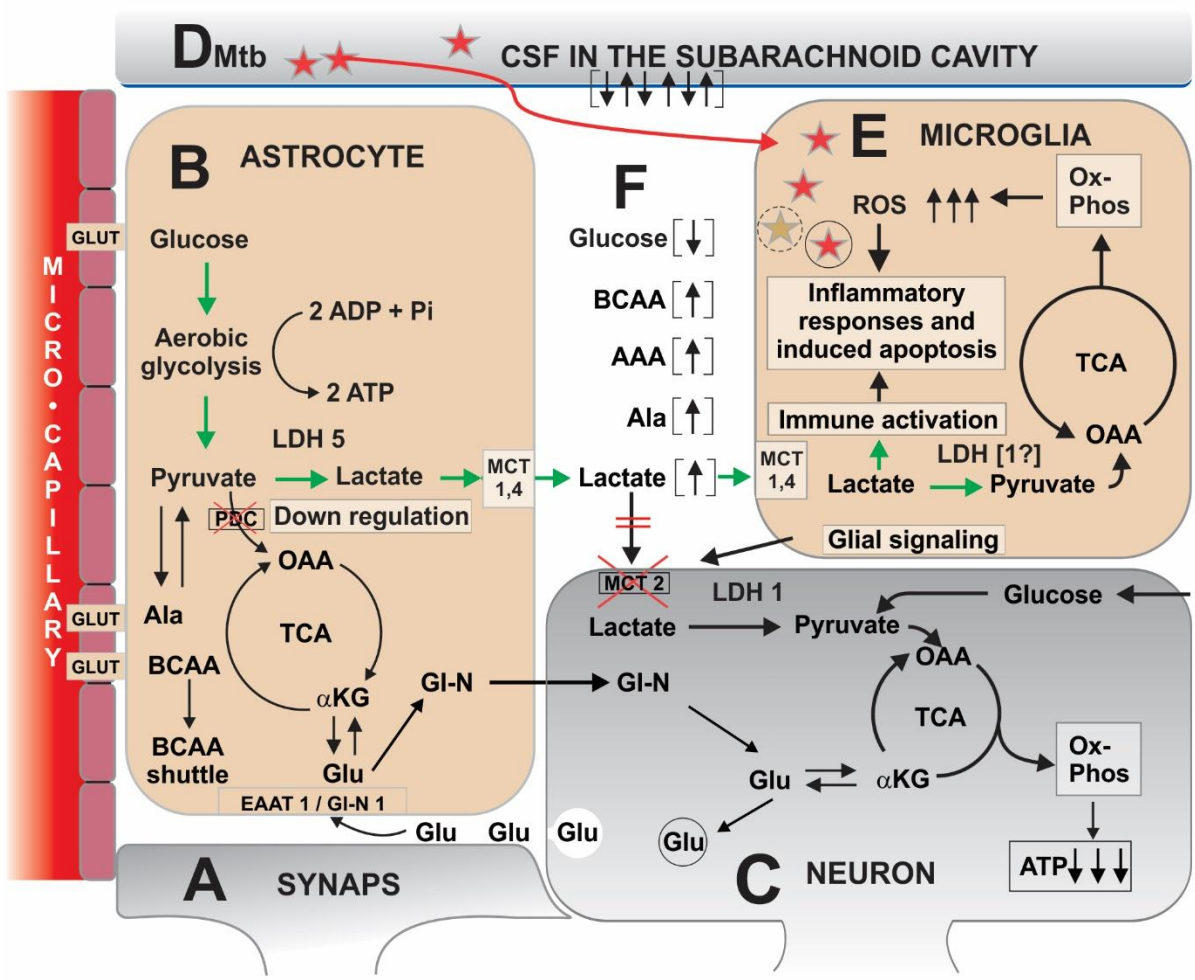


Figure A1-2: Conceptual model of the hypothetical “astrocyte–microglia lactate shuttle”.

Legend:

1. Energy profile during homeostasis and neuron activation**.** Lactate shuttles provide models describing the movement of intra- and intercellular lactate between diverse cells in several tissues under both anaerobic and aerobic conditions. An “astrocyte–neuron lactate shuttle” in the brain is the model for energy cooperation between astrocytes and neurons during high energy demand in the neurons. During homeostasis and neuron activation, an important energy load of the neurons is required for the operation of glutamatergic synapses [**A**], which is tightly regulated by intercellular links and dynamic interactions between astrocytes and neurons. The glutamate-induced glucose uptake in the astrocytes by endothelial GLUT1 transporters [**B**], followed by glycolysis, functions as the cellular mechanism coupling the astrocytes’ energy production to neuronal energy-dependent activity [**B** and **C**]**.** A putative branched-chain amino acid (BCAA) shuttle, involving isoforms of branched-chain aminotransferases where BCAAs provide the amino group for glutamate synthesis from α-ketoglutarate, promote glutamine transfer to neurons. In their turn, the neurons re-aminate the amino acids for another cycle of the shuttle. Decreased aerobic oxidation (TCA and oxidative phosphorylation) in the astrocytes during neuron activation results from down regulation of the PDC

through phosphorylation of its PDHα subunit. The LDH of astrocytes is primarily of the type-5 isoform (LDH-5), which catalyzes pyruvate hydrogenation to lactate, the final catabolic end product of astrocytes' aerobic glycolysis. Neurons primarily express LDH-1, which catalyzes lactate to pyruvate. Furthermore, astrocytes, as well as neurons, express MCT proteins participating in the lactate shuttle. MCT-4 in astrocytes is a low affinity transporter for lactate (high K_m: facilitates export of lactate produced by glycolysis in the astrocytes). Neurons express MCT-2, a high affinity transporter for lactate (low K_m: facilitates the uptake of lactate produced by the astrocytes). Consistent with their higher energy requirements, activated neurons sustain a high rate of oxidative metabolism, emanating from lactate, which enters the neuronal TCA cycle via pyruvate, neuronal oxidative phosphorylation and ATP production.

2. Energy profile during TBM neuroinflammation**.** In TBM, the causative Mtb apparently enters the subarachnoid space [**D**], followed by its distinctive capacity to enter and replicate within the microglial cell [**E**], which, however, also provides the first line of defence against invading pathogens. Brain cells are surrounded by interstitial fluid [**F**], which is contiguous with CSF in the subarachnoid space [**D**]. The profile of energy-associated metabolites in CSF from TBM patients differs significantly from that of these metabolites associated with the astrocyte–neuron–lactate shuttle, leading to the present hypothesis of an "astrocyte–microglia lactate shuttle" as operating in neuroinflammation, triggered by Mtb infection. It is proposed that an analogue of the reverse Warburg process underpins this shuttle. Bidirectional communication networks between glial cells and neurons may produce signalling substances directed to downregulation of neuronal MCT-2 and of the BCAA shuttle for high lactate and keto acids provision to neurons. Stimulated by proinflammatory signals, including from lactate, the microglia soon undergo an array of transformations, resulting in an oxidative burst that produces superoxide from an orchestrated mechanism involving nicotinamide adenine dinucleotide phosphate (NADPH) oxidase, leading to the formation of reactive oxygen species (ROS) required for bactericidal activity and eradication of pathogens. However, activation of microglia may likewise compromise their viability, and some of this vulnerability possibly comes from the oxidative stress of the oxidative burst and a perturbed ROS balance. Activated microglia finally secrete a diverse array of proinflammatory molecules, suggesting that oxidative stress emanating from activated microglia affects the astrocytes, causes loss of neuron function and eventually precipitates the severe clinical symptoms of TBM.

 Mtb;  destroyed Mtb;  protected Mtb; 

neuronal glutamate vesicle; GLUT, microvascular transporters for glucose, organic acids, fatty acids, amino acids and ketones; PDC, pyruvate dehydrogenase complex; MCT, monocarboxylate transporter; EAAT 1/Gl-N 1, transporters for glutamate; Ala, alanine; BCAA, branched-chain amino acids; AAA, aromatic amino acids; Glu, glutamate; Gl-N, glutamine; OAA, oxalic acetic acid; αKG, α-ketoglutaric acid; Ox-Phos, oxidative phosphorylation; ROS, reactive oxygen species; CSF, cerebrospinal fluid.

A1.4.6 Directive

One purpose of models of biological phenomena is to provide a bridge between disciplines and to facilitate the inter-disciplinary research necessary to understand complex phenomena, of which CNS infection by Mtb is a prime example. In this paper, data from an explorative metabolomics investigation on TBM are used to evaluate the relationship between the intricate spatial network of cell–cell interaction and other characteristics of the CNS with empirical metabolomics data. For this purpose we accepted the physiological role of lactate in homeostasis and in stress as a key player in the paradigms on bioenergetics. Where the conceptual and experimental models agree and disagree will reveal issues for further research. We present the model of the AMLS hypothesis in the hope that it will stimulate directed analytical and systems research into its details, paving the way for improved insights and potential applications in understanding and management of TBM.

ANNEXURE 2:

Supplementary information:

KEMREP: A new qualitative method for the assessment of an analyst's ability to generate a metabolomics data matrix by gas chromatography–mass spectrometry

A2.1 Abstract

In the supplementary information (SI) we present the laboratory standard operating protocol (SOP) used for the analysis of the cerebrospinal fluid (CSF), of commercially obtained lyophilized plasma sample spiked with known concentrations of amino acids and a normal urine sample obtained from a healthy volunteer as discussed in the manuscript. This is followed by the raw data for these respective samples, made available online as part of the SI, the additional statistical data of samples analysed not included in the manuscript and a detailed biological/chemical description of the 95 common CSF features in the final data matrix. Finally, we also include instructions on how to use the R script in order to execute KEMREP, as well as the coded program, made available online as part of the SI.

A2.2 Laboratory SOP

A2.2.1 Overview

This section provides the extended laboratory standard operating protocol (SOP) that we use in metabolomics analyses of urinary organic acids in disease states (Reinecke *et al.* 2012) and in dietary treatment (Dercksen *et al.* 2013). Here it is applied for analysis of cerebrospinal fluid (CSF), commercially obtained lyophilized plasma sample spiked with known concentrations of amino acids, and a normal urine sample obtained from a healthy volunteer. This section also includes the methods for GC-MS analysis and the GC separation results.

The three principal steps for the procedures used here are:

1. Isolation of the organic acids from physiological fluids
2. Formation of volatile derivatives
3. GC-MS analysis.

Organic acids are isolated from physiological fluids with ethyl acetate and diethyl ether extractions. Standard volumes are used when extracting organic acids from CSF and plasma, given the well-regulated nature of these biofluids and thus the low variances in component concentrations, whereas urine volumes used are dependent upon the creatinine level to account for dilution. The organic acid extract is evaporated to dryness under nitrogen; volatile trimethylsilyl (TMS) derivatives of the extracted organic acids are formed by heating with N,O-bis-(trimethylsilyl)trifluoraceteamine (BSTFA). The TMS derivatives are less than ideal products for some classes of compounds such as acylglycine, which form mono- and di-TMS derivatives, yet they are the most useful and versatile compounds for the wide range of functional groups in organic acids. The derivatives are analysed by GC-MS.

GC-MS is able to separate the highly volatile organic acids using gas chromatography, followed by detection of individual components by means of mass spectroscopy. This procedure permits rapid identification and quantification of constituent organic acids with a high degree of sensitivity and chromatographic resolution

A2.2.2 Reagents

NOTE: refer to chemical information sheet below for descriptions and precautions/hazards of chemicals used.

Internal standard (3-phenylbutyric acid solution) prepared by measuring precisely 26.25 mg 3-phenylbutyric acid, adding 3 drops of 1M sodium hydroxide (NaOH) to dissolve and adding 50 ml distilled H_2O (dH_2O). Other reagents include: 5M hydrochloric acid (HCl); ethyl acetate (HPLC grade), distilled once to purify further; diethyl ether (HPLC grade), distilled once to purify further; anhydrous sodium sulphate (Na_2SO_4); bis(trimethylsilyl)-trifluoracetamide (BSTFA); trimethylchlorosilane (TMCS); pyridine and hexane.

A2.2.3 Instrument settings

Gas chromatography (GC)

	GC-1	GC-2
GC	Agilent 7890A	Agilent 7890A
Autosampler	Agilent 7683B	Agilent 7693
Oven program	60°C for 2 min; then 4°C/min to 120°C for 0 min; then 6°C/min to 285°C for 2 min	50°C for 1 min; then 20°C/min to 60°C for 0 min; then 5°C/min to 120°C for 0 min; then 7°C/min to 280°C for 4 min
Run time	46.5 min	40.35 min
Post run	5 min at 295°C	1 min at 300°C
Injection volume	1 µl	1 µl
Pre-injection washes	Solvent A: 0 Solvent B: 0	Solvent A: 2 x 4 µl Solvent B: 0
Post-injection washes	Solvent A: 3 x 8 µl Solvent B: 3 x 8 µl	Solvent A: 1 x 4 µl Solvent B: 2 x 4 µl
Front inlet	Heater: 280°C Carrier gas: Helium Total flow: 14 ml/min Split ratio: 10:1 Split flow: 10 ml/min	Heater: 280°C Carrier gas: Helium Total flow: 16.68 ml/min Split ratio: 12:1 Split flow: 12.63ml/min
Column	DB-1MS 450°C: 30 m x 250 µm x 0.25 µm	DB-1MS 340°C: 30 m x 250 µm x 0.25 µm

Mass spectrometry (MS)

	MS-1	MS-2
MS	Agilent 5975B XL MSD	Agilent 5975C VL MSD
Solvent delay	6.5 min	6 min
Acquisition mode	Scan	scan
Scan parameters	Low mass: 50.0 High mass: 550.0 Threshold: 150	Low mass: 50.0 High mass: 600.0 Threshold: 150
MS source	230°C (max 250°C)	230°C (max 250°C)
MS quad	150°C (max 200°C)	150°C (max 200°C)

A2.2.4 Organic acid extraction

- For urine:

Aliquot volume of urine to large glass test tube according to creatinine value:

Creatinine	< 100 mg%	use 1 ml urine
Creatinine	> 100 mg%	use 0.5 ml urine
Creatinine	< 5 mg%	use 2 ml urine
Creatinine	< 2 mg%	use 3 ml urine

Add 6 drops 5M HCl to adjust to pH 1 (using glass pipette), internal standard (IS) 5x creatinine mg% = volume in µl.

- For plasma/CSF:

Add 1 ml sample to large glass test tube, 6 drops 5M HCl to adjust to pH 1 (using glass pipette) and 100 µl internal standard.

Procedure continues for all samples as follows:

1. Add 6 ml ethyl acetate to large glass test tube
2. Cap test tubes and check there is no leakage by inverting test tube (quality control step to ensure no sample is lost during next step)
3. Mix for 30 min in Roto-torque
4. Centrifuge at 3000 rpm for 3 min
5. Aspirate the organic phase into clean, large glass test tube (using glass pipette) and set aside
6. Add 3 ml diethyl ether to the aqueous (lower) phase
7. Cap test tubes and check there is no leakage by inverting test tube
8. Mix for 10 min in Roto-torque
9. Centrifuge at 3000 rpm for 3 min
10. Aspirate the organic phase and add to the ethyl acetate phase (using glass pipette)
11. Discard lower aqueous phase into appropriate organic waste container
12. Add two level spatula scoops of anhydrous Na_2SO_4 to organic phase
13. Cap and invert test tube several times (or vortex for 5 seconds) to ensure good mixing (proper dispersion of Na_2SO_4 ensures all water molecules removed from organic phase as water reverses chemical process of silylation, thereby reducing the efficiency of derivatization)
14. Centrifuge at 3000 rpm for 1 min
15. Pour/decant the organic phase into a clean, small glass test tube

16. Evaporate to dryness in heating block at 37°C under nitrogen gas (~1 hour)
17. Using Hamilton glass syringe:

add BSTFA	A: (2x creat mg% = volume in µl) for urine.
	B: 40 µl for plasma/CSF.
add TMCS	A: (0.4x creat mg% = volume in µl) for urine.
	B: 8 µl for plasma/CSF.
add pyridine	A: (0.4x creat mg% = volume in µl) for urine.
	B: 8 µl for plasma/CSF.

NOTE: Hamilton glass syringe is kept clean with pyridine and approximately 100 µl hexane is withdrawn into syringe and discarded (five times) between the addition of each reagent (quality control step to ensure syringe is clean and avoid cross-contamination).

18. Cap test tubes and incubate at 60°C for 1 hour (45 min at 70°C)
19. Set up and label GC-MS vials, with insert and cap
20. Transfer approximately 100 µl sample to GC/MS vial

NOTE: clean glass syringe with hexane (five times) after each transfer

21. Cap GC-MS vial and place in autosampler and process via GC-MS.

NOTE: Amino acid analysis conducted using EZ:faast™ Kit from Phenomenex: (**http://www.phenomenex.com/Products/AminoAcidDetail/EZfaast**)

A2.2.5 General AMDIS settings

GC-MS data were analysed in AMDIS in batch mode with the following settings:

- 60% minimum match factor
- Type of analysis: use retention time (show standards)
- Resolution: medium
- Sensitivity: low
- Shape requirements: low.

Features identified by comparing each feature's MS-spectral pattern with:
(i) customised spectral library specific with regard to CSF sample under investigation;
(ii) in-house spectral library specific to amino acids and organic acids with regard to commercial spiked plasma sample and normal, healthy urine sample, respectively, as discussed in the manuscript.

A2.2.6 Chemical Information Sheet

- 3-Phenylbutyric acid – $C_{10}H_{12}O_{2;}$ Mw: 164.21 g/mol; supplier: Fluka (25 g) (index no.78243). Precaution/hazard: avoid contact with skin and eyes.

- Sodium hydroxide – NaOH; Mw: 40.00 g/mol; supplier: Merck (500 g). Precaution/hazard: corrosive (causes severe burns).

- Hydrochloric acid (32%) – HCl; Mw: 36.36 g/mol; supplier: Merck (2.5 l). Precaution/hazard: corrosive (causes severe burns); irritating to respiratory system.

- Ethyl acetate – $CH_3COOC_2H_5$; Mw: 88.11 g/mol; supplier: Merck (2.5 l) (index no. 607-022-00-5). Precaution/hazard: highly flammable; causes drowsiness/dizziness; causes eye irritation; repeated exposure causes skin dryness/cracking.

- Diethyl ether – $(C_2H_5)_2O$; Mw: 74.12 g/mol; supplier: Merck (2.5 l) (index no. 602-022-00-4). Precaution/ hazard: extremely flammable; harmful if swallowed; causes drowsiness/dizziness; repeated exposure causes skin dryness/cracking; may form explosive peroxides.

- n-Hexane – $CH_3(CH_2)_4CH_3$; Mw: 86.18 g/mol; supplier: Merck (2.5 l) (index no. 601-037-00-0). Precaution/hazard: highly flammable; fatal if swallowed; causes skin irritation; toxic to aquatic life; causes drowsiness/dizziness; may cause infertility or damage to unborn child; may cause damage to organs through prolonged or repeated exposure.

- Sodium sulphate – Na_2SO_4; Mw: 142.04 g/mol; supplier: Merck (500 g).

- Chlorotrimethylsilane (TMCS) – C_3H_9ClSi; Mw: 108.64 g/mol; supplier: Flukka Analytical (100 ml) (index no. 92360). Precaution/hazard: highly flammable; corrosive (causes severe burns); reacts violently with water; harmful by inhalation/contact to skin; irritating to respiratory system.

- Pyridine – C_5H_5N; Mw: 79.10 g/mol; supplier: Flukka Analytical (1 l) (index no. 82703). Precaution/ hazard: highly flammable; harmful if inhaled or swallowed; harmful to skin.

Bis(trimethylsilyl)-trifluoracetamide (BSTFA) – $CF_3C=NSi(CH_3)_3OSi(CH_3)_3$; Mw: 257.40 g/mol; supplier: Supelco Analytical (25 ml) (index no. 3-3027). Precaution/hazard: flammable; irritant to eyes and skin; causes burns.

A2.3 Raw data matrices

Please find online, included as part of the SI, the following data matrices as discussed in the manuscript:

A2.3.1 TBM-infected CSF

A2.3.1.1 Raw data for both GC-MS-1 (398 features) and GC-MS-2 (239 features)

Table A2-1: Raw CSF data (concentrations).

Table A2-1: Raw CSF data (retention times).

A2.3.1.2 Reduced, common data (95 features)

Table A2-3: Common 95 CSF data.

A2.3.2 Urine from healthy individuals

Table A2-4: Raw Data - Urine - first attempt (3 analysts).

Table A2-5: Raw Data - Urine - second attempt (3 analysts).

A2.4 Statistical Data

A2.4.1 Common 95 CSF features

A2.4.1.1 Descriptive statistics

Table A2-6: Descriptive statistics (mean, standard deviation (SD) and coefficient of variation (CV)) of 95 common CSF features for GC-MS-1 and GC-MS-2.

	GC-MS-1			GC-MS-2		
	Mean	**SD**	**CV**	**Mean**	**SD**	**CV**
8.2332.minCSF1	4.076	1.714	42.054	1.898	0.725	38.213
8.3179.minCSF4	0.084	0.005	5.874	0.131	0.012	9.236
8.4358.minCSF1	2.367	0.975	41.207	2.117	1.347	63.622
8.4786.minCSF4	0.050	0.025	50.222	0.055	0.020	36.912
8.6480.minCSF1	0.228	0.082	36.092	0.364	0.208	57.153
8.7127.minCSF1	0.414	0.121	29.174	0.680	0.096	14.183
8.9936.minCSF1	105.079	9.622	9.157	71.245	3.035	4.260
9.0732.minCSF1	0.106	0.072	67.397	0.402	0.283	70.348
9.2649.minCSF1	0.261	0.154	58.889	0.165	0.120	72.531
9.3885.minCSF1	0.369	0.095	25.767	0.819	0.155	18.880

9.7987.minCSF1	0.863	0.161	18.625	0.275	0.041	14.758
9.9710.minCSF1	0.160	0.089	55.887	0.188	0.089	47.365
10.0997.minCSF1	0.191	0.154	80.736	0.603	0.354	58.782
10.2273.minCSF1	8.616	3.627	42.095	9.967	3.706	37.186
10.3612.minCSF1	2.487	0.646	25.984	4.239	2.557	60.329
10.7554.minCSF1	81.507	32.156	39.452	113.398	54.295	47.880
10.8539.minCSF1	6.620	1.564	23.622	2.347	1.790	76.300
10.8962.minCSF1	1.000	0.428	42.792	0.460	0.215	46.769
11.4364.minCSF1	0.461	0.253	54.903	0.680	0.492	72.373
12.2160.minCSF5	0.037	0.032	87.297	0.054	0.033	61.803
12.5670.minCSF1	3.756	1.896	50.478	4.019	1.661	41.340
12.8943.minCSF1	3.124	1.364	43.658	2.976	1.132	38.030
13.9729.minCSF1	1.719	0.727	42.289	1.623	0.587	36.167
14.1056.minCSF1	0.140	0.111	79.677	0.430	0.260	60.499
14.8352.minCSF2	0.111	0.037	33.521	0.038	0.025	65.959
14.9376.minCSF2	0.390	0.259	66.494	0.645	0.255	39.489
14.9702.minCSF1	0.706	0.276	39.103	2.070	0.375	18.128
15.6060.minCSF1	0.192	0.092	47.786	0.162	0.063	39.171
16.2239.minCSF1	0.451	0.269	59.655	0.835	0.267	32.013
16.2457.minCSF1	0.162	0.130	80.019	0.190	0.125	65.912
16.4047.minCSF1	1.037	0.705	68.026	1.130	0.719	63.653
16.6794.minCSF1	0.290	0.150	51.495	0.458	0.164	35.734
16.8190.minCSF1	0.176	0.115	65.455	0.425	0.171	40.189
17.1011.minCSF1	0.268	0.180	67.225	0.446	0.325	72.898
18.2592.minCSF2	0.124	0.038	31.085	0.184	0.104	56.352
18.6689.minCSF1	6.146	2.301	37.440	5.867	2.036	34.705
19.7910.minCSF1	0.250	0.120	48.163	0.496	0.241	48.688
19.8894.minCSF4	0.063	0.019	29.662	0.164	0.071	43.159
20.0565.minCSF1	0.275	0.071	25.694	0.224	0.062	27.511
20.4015.minCSF1	3.599	2.200	61.139	5.168	1.944	37.604
21.1614.minCSF1	0.801	0.378	47.214	0.809	0.477	58.901
21.8515.minCSF1	56.500	0.000	0.000	56.500	0.000	0.000
22.3688.minCSF1	1.163	0.526	45.264	1.025	0.399	38.880
23.3775.minCSF1	0.699	0.181	25.906	0.240	0.144	60.027
23.5635.minCSF1	2.129	0.610	28.672	1.834	0.523	28.542
23.9217.minCSF1	0.603	0.231	38.220	0.420	0.143	34.083
24.4962.minCSF1	3.968	3.558	89.665	3.367	1.513	44.926
25.3692.minCSF3	0.071	0.016	22.408	0.101	0.007	7.316
25.9547.minCSF1	0.223	0.127	57.086	0.136	0.065	47.952
26.6722.minCSF1	0.505	0.150	29.765	0.571	0.129	22.530
27.1134.minCSF1	0.136	0.041	30.284	0.075	0.047	62.764
27.8183.minCSF1	1.293	0.447	34.551	0.958	0.609	63.624
28.1451.minCSF2	0.241	0.165	68.734	0.337	0.171	50.750
28.3019.minCSF1	0.782	0.225	28.773	0.728	0.195	26.739
28.3866.minCSF1	0.096	0.029	30.193	0.096	0.024	25.254
28.4323.minCSF1	0.655	0.194	29.700	0.600	0.133	22.232
28.5319.minCSF1	0.425	0.129	30.404	0.298	0.056	18.923
28.5988.minCSF2	0.088	0.061	69.137	0.140	0.089	63.877
28.7110.minCSF1	0.213	0.129	60.369	0.174	0.141	80.854
28.7327.minCSF1	0.061	0.040	65.943	0.074	0.018	24.137
28.7665.minCSF1	0.107	0.039	35.945	0.076	0.045	59.694
28.9209.minCSF5	0.065	0.049	74.990	0.082	0.073	89.348
29.1155.minCSF1	0.086	0.028	32.993	0.099	0.022	21.829
29.2168.minCSF1	0.262	0.104	39.713	0.181	0.039	21.573

30.1821.minCSF1	6.714	2.545	37.907	5.584	1.803	32.291
31.3042.minCSF1	0.122	0.052	42.306	0.146	0.058	39.980
31.5148.minCSF1	12.828	2.722	21.216	12.896	2.524	19.573
33.4499.minCSF1	4.122	1.725	41.860	3.357	1.545	46.040
34.5314.minCSF1	0.423	0.231	54.566	0.277	0.158	57.205
34.6206.minCSF1	0.062	0.059	94.497	0.093	0.046	48.999
34.9886.minCSF1	1.129	0.492	43.552	0.960	0.399	41.558
35.5739.minCSF1	0.689	0.148	21.470	0.551	0.113	20.493
36.0878.minCSF1	16.603	6.825	41.107	8.507	3.982	46.810
36.1404.minCSF1	3.452	1.139	32.990	2.427	1.369	56.390
36.3853.minCSF1	6.661	2.055	30.858	5.058	1.983	39.198
36.4007.minCSF1	0.315	0.117	37.131	0.288	0.099	34.305
36.4539.minCSF1	0.812	0.433	53.239	0.339	0.158	46.598
37.2356.minCSF1	0.501	0.211	42.071	0.319	0.105	32.835
37.4599.minCSF1	1.112	0.194	17.430	0.835	0.282	33.796
37.6905.minCSF1	0.184	0.073	39.546	0.153	0.056	36.904
37.7923.minCSF1	0.155	0.095	60.896	0.129	0.081	62.746
38.0412.minCSF1	0.166	0.096	57.794	0.139	0.069	49.865
38.1087.minCSF1	6.916	2.632	38.057	5.959	2.126	35.688
38.5871.minCSF1	1.865	0.626	33.576	1.093	0.338	30.897
39.0609.minCSF1	0.165	0.056	33.758	0.148	0.068	46.175
39.7561.minCSF1	0.154	0.062	40.440	0.087	0.060	68.857
39.9827.minCSF1	3.305	0.979	29.636	1.942	1.272	65.502
42.2360.minCSF1	0.162	0.063	39.160	0.070	0.041	58.893
43.2087.minCSF1	0.054	0.036	65.736	0.037	0.013	35.558
44.1786.minCSF1	0.450	0.127	28.242	0.286	0.191	66.674
44.2461.minCSF1	0.582	0.191	32.755	0.443	0.264	59.557
44.5340.minCSF1	0.373	0.086	23.035	0.286	0.072	25.176
44.6404.minCSF1	3.778	1.466	38.815	2.330	0.887	38.065
45.3825.minCSF1	0.221	0.094	42.718	0.120	0.058	48.270
45.8443.minCSF1	2.812	1.062	37.757	1.747	0.688	39.396

A2.4.1.2 CV distributions

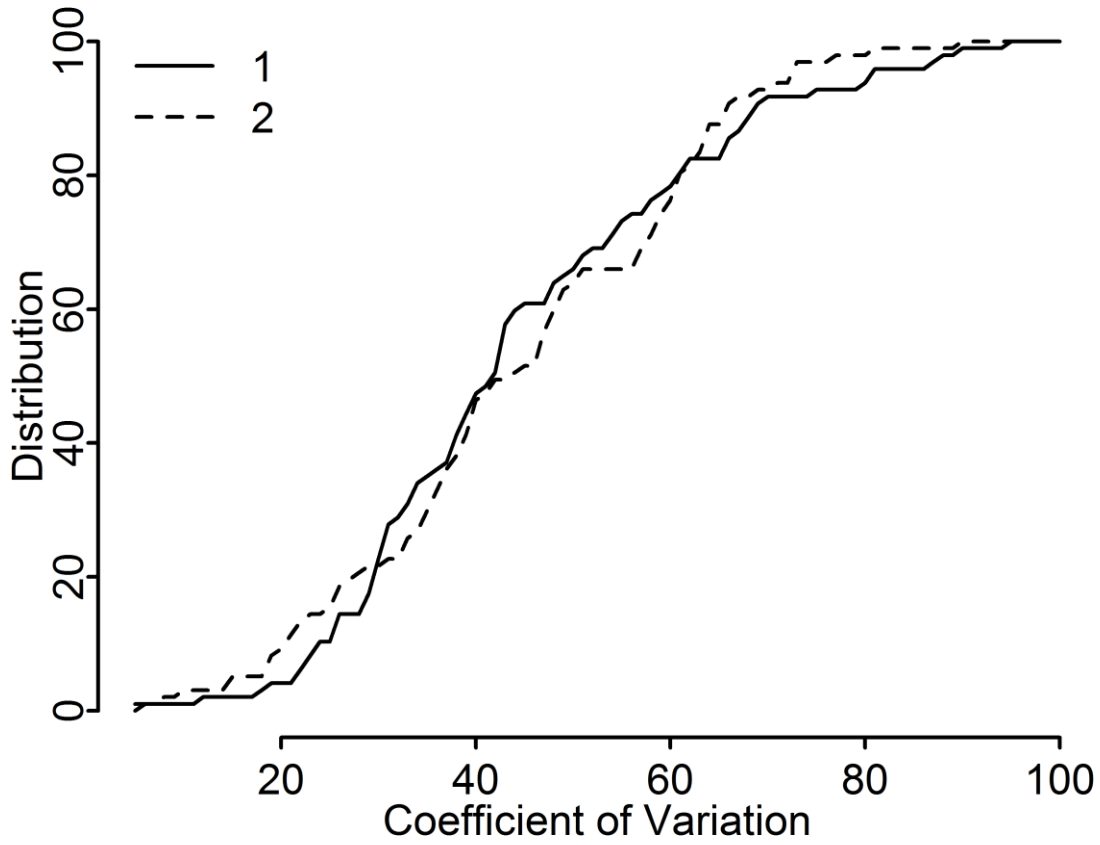


Figure A2-1: Overlay of distributions of CVs for GC-MS-1 (**—**) and GC-MS-2 (- - -).

A2.4.1.3 Normality and variance tests

The usual univariate hypothesis testing at a 5% significance level was performed to test for normality and allowable variance. For this we used a predefined CV value of 30% to determine the repeatability and reproducibility null distributions. During the three successive stages of the data reduction, the percentage of variables observed for repeatability with CV < 30% steadily increased from approximately 10% to approximately 22% for both GC-MS-1 and 2. This indicated the typical non-constant measurement errors that may be specific for individual metabolites, although the measurement errors of different metabolites may also be correlated. We subsequently determined the percentage of variables that satisfy the variance assumptions (that is, small within-variability). By contrast, it is clear that this measure progressively increased during the successive phases of data reduction, ranging from approximately 26%, through 56% to 59%, for both GC-MS instruments.

Table A2-7: Schematic outline of formal hypothesis testing of repeatability and reproducibility, and corresponding outcomes.

Description	Work-flow phase and GC-MS used (number of features)	% Variables with a CV<30%	% Variables that satisfy variance assumptions	% Variables that satisfy normality assumptions	% Variables that showed no significant difference in median values
Repeatability	1. Original data matrix: GC-MS-1 (398)	9.82	26.2	48.61	N/A
	GC-MS-2 (239)	11.76	25.63	44.96	N/A
	2. Reduced data matrix: GC-MS-1 (175)	21.26	57.47	89.08	N/A
	GC-MS-2 (112)	24.32	54.95	74.77	N/A
	3. Common features per GC-MS: GC-MS-1 (95)	22.34	64.89	90.43	N/A
	GC-MS-2 (95)	21.28	55.32	74.47	N/A
Reproducibility	GC-MS-1&2 (95)	N/A	39.36	78.72	82.98

The Shapiro-Wilk test for normality was used on each data set; the percentages of variables that satisfied normality were 48.6% to 89.0% to 90.4% for GC-MS-1, which outperformed the corresponding values for GC-MS-2 (44.9% to 74.7% to 74.4%). The percentage differences appear to be compatible with the different uses to which the two GC-MS instruments were put. The measures from the normality and variance tests thus clearly improve on the CV measure to evaluate repeatability in the generation of metabolomics data.

The measure of reproducibility of the 95 common features as determined on the two GC-MS instruments showed the same tendency as that observed for the repeatability measures; that is, the percentage of variables that satisfied normality assumptions > percentage of variables that satisfied variance assumptions (78.2% and 39.3%, respectively). Finally, we used the non-parametric Mann-Whitney U-test to compare reproducibility further. The appropriate test was used on all the features individually; the percentage of reproducible features was found to be 82.9%. We regard this as a good result, given the elaborate analytical procedures employed to identify the features in the repeats, and the diverse history of the GC-MS instruments.

A2.4.2 Amino acid analyses – KEMREP density plots

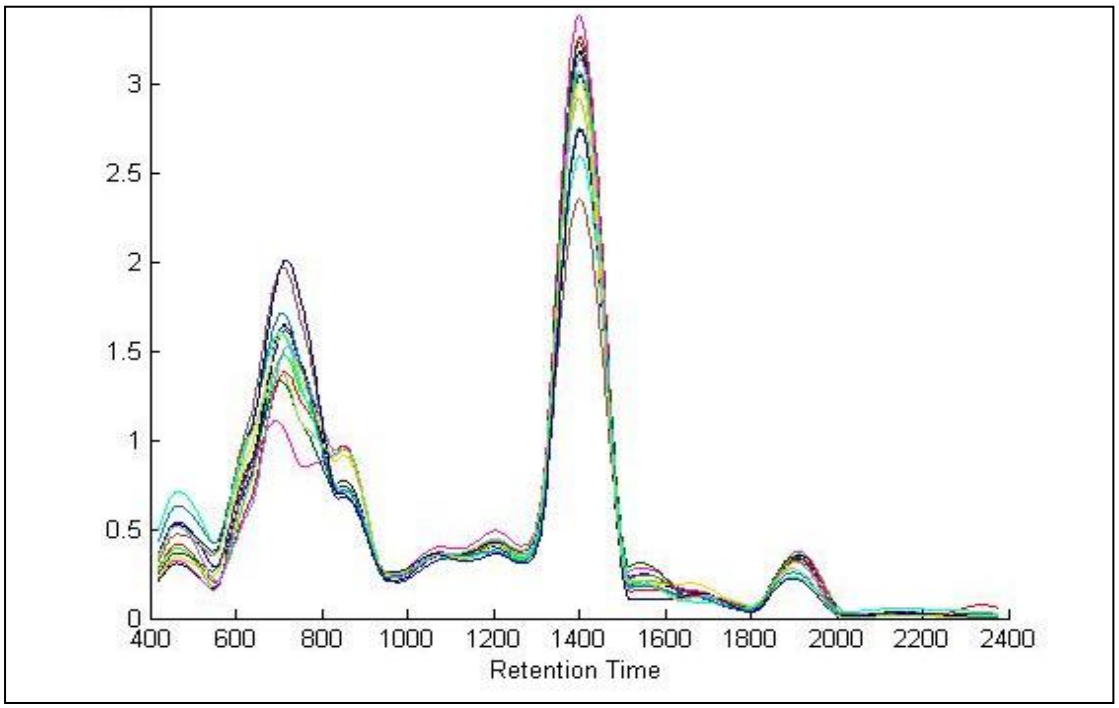


Figure A2-2: Overlay of smoothed chromatograms of 10 technical replicates of amino acid analysis of commercial lyophilized plasma sample spiked with known concentrations of amino acids.

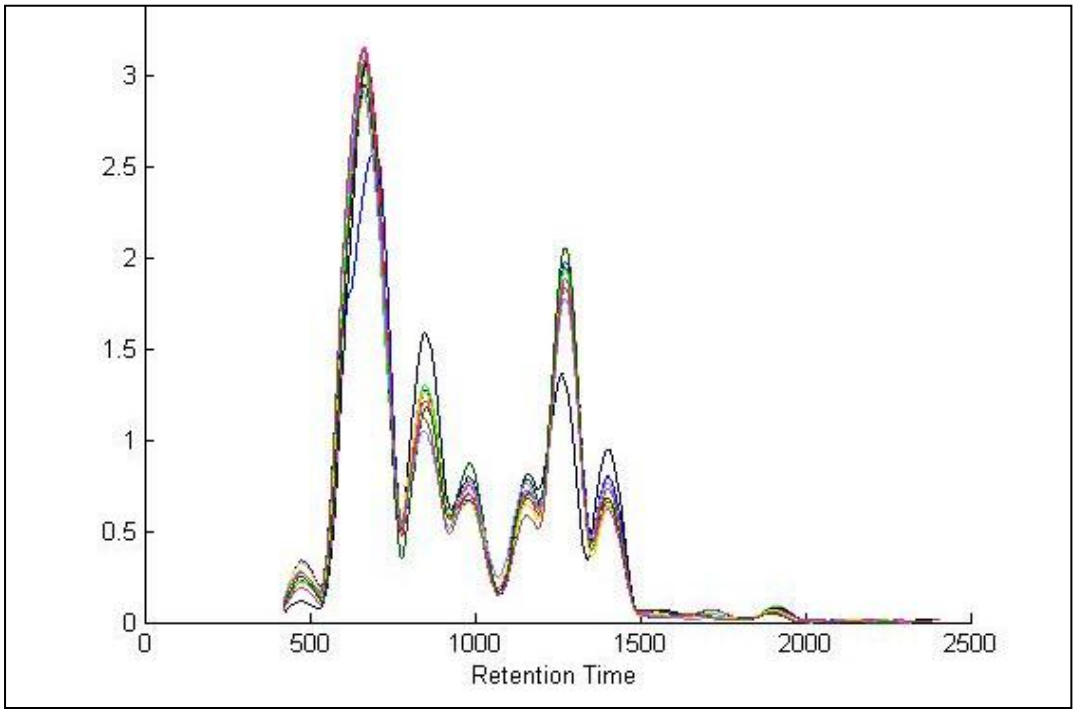


Figure A2-3: Overlay of smoothed chromatograms of 10 technical replicates of amino acid analysis of control urine sample.

A2.5 Biological Data

A2.5.1 Common 95 CSF features – biological interpretation

One of the primary aims of metabolite annotation is the subsequent identification of biomarkers, a biosignature, a metabolite profile or monitoring of treatment. These aims were not sought in the present investigation. However, classification of the 95 common features can provide insight into their origin in the CSF sample from the TBM patient. This classification is summarized in Table A2-8. Several organic acids in the CSF metabolome of the healthy human were detected in the experimental sample analysed. Some are among the most abundant common organic acids (e.g., acetic acid, acetoacetic acid, citric acid, lactic acid, and pyruvic acid) and others are of low abundance (e.g., homovanillic acid). Other organic acids (e.g., glutaric acid) and components of the tricarboxylic cycle (fumaric acid and malic acid) are sometimes associated with disorders or organic acid metabolism with neurological manifestations or defects in the mitochondrial respiratory chain. These observations are thus consistent with the severe disease state of the patient from whom the CSF sample was obtained.

Table A2-8: Biological classification of the 95 common features, including approximate match factor ranges of features compared with the NIST MS database.

Class Number	Class designation	Number of components	Range of match factors	Comments
1	Organic acids	29, including 3-phenylbutyrate (IS)	62.3–98.1	Include substances that may be indicators of disease
2	Fatty acids	5	88.5–93.8	Include exogenous substances from the diet
3	Alkanes	6	75.7–91.7	Origin unknown – exogenous or endogenous
4	Match factor <60%	6	49.8–59.3	Possible misclassified substances due to software inaccuracies
5	By-products	20	56.4–93.8	By-products of the derivatization reaction
6	Unknowns	29	61.6–95.7	Some catabolites possibly from medication or microorganisms
Total		**95**	**49.8–98.1**	

Although the origin of the alkanes is unclear, it is known that oxidative stress produces reactive oxidative species (ROS) with ROS-mediated lipid peroxidation as one of the main consequences of brain injury, resulting in the production of conjugated dienic hydroperoxides

that decompose into aldehydes, dienals or alkanes (Moore & Roberts 1998). In this regard formations of alkanes with a relatively low molecular weight have been observed in multiple sclerosis, one of the major neurological diseases (Toshniwald & Zarling 1992). The remainder of the 95 common components are related to features with a low (<60%) match factor, by-products of the derivatization process used, and unknowns. Among the last group several compounds seem to be related to catabolism of medication given to the patient, although indication of degradation products with a bacterial origin cannot be excluded. The outcome of the comprehensive metabolomics analysis of CSF from acutely ill, but untreated, TBM patients thus clearly holds promise for the identification of potential early biomarkers of this devastating infectious disease.

The NIST (National Institute of Standards and Technology) MS search program application for AMDIS was used tentatively to identify/name the final 95, reduced, common features based upon the best match of their MS spectra with those in the NIST library. The NIST names, as well the corresponding biological and IUPAC names, with their metabolite classification and their derivatized and un-derivatized molecular weights (g/mol) and chemical formulas, are given in Table A2-9.

Table A2-9: Detailed description of the 95 reduced, common features including: metabolite classification customised AMDIS library names with approximate match factor, NIST names, biological/IUPAC names, and corresponding derivatized and underivatized molecular weights (g/mol) and chemical formulas.

			Derivatized				Underivatized		
#	**Class**	**Custom Library ID**	**NIST Name**	**Match (%)**	**Formula**	**Mw**	**Biological/IUPAC Name**	**Mw**	**Formula**
1	1	8.2332.minCSF1	D-(-)-Lactic acid, trimethylsilyl ester	71.2	C6H14O3Si	162	Lactic acid/2-hydroxy-propanoic acid	90	$C_3H_6O_3$
2	1	8.4358.minCSF1	D-(-)-Lactic acid, trimethylsilyl ether	82.6	C6H14O3Si	162	Lactic acid/2-hydroxy-propanoic acid	90	$C_3H_6O_3$
3	1	8.7127.minCSF1	Malonic acid, bis(2-trimethyl-silylethyl ester	63.2	C13H28O4Si2	304	Malonic acid/prop-anedioic acid	104	$C_3H_4O_4$
4	1	10.2273.minCSF1	2,3-Butanediol, O-(trimethylsilyl)-monoacetate	76.6	C9H20O3Si	204	2,3-Butanediol mono-acetate/3-hydroxybutan-2-yl acetate	132	$C_6H_{12}O_3$
5	1	10.7554.minCSF1	Propanoic acid, 2-[(trimethyl silyl)oxy]-, trimethylsilyl ester	94.7	C9H22O3Si2	234	Lactic acid/2-hydroxy-propanoic acid	90	$C_3H_6O_3$
6	1	10.8962.minCSF1	Propanoic acid, 2-methyl-2-[(trimethylsilyl)oxy]-, trimethyl-silyl ester	88.6	C10H24O3Si2	248	2-Hydroxyisobutyric acid/2-hydroxy-2-methylpropanoic acid	104	$C_4H_8O_3$
7	1	11.4364.minCSF1	Propanoic acid, 2-oxo-3-(trimethylsilyl)-, trimethylsilyl ester	95.8	C9H20O3Si2	232	Pyruvic acid/2-oxo-propanoic acid	88	$C_3H_4O_3$

8	1	12.5670.min CSF1	Ethanedioic acid, bis(tri-methylsilyl) ester	95.9	C8H18O4Si2	234	Oxalic acid/ethanedioic acid	90	$C_2H_2O_4$
9	1	12.8943.min CSF1	Butanoic acid, 2-[(trimethyl silyl)oxy]-, trimethylsilyl ester	94.2	C10H24O3Si2	248	3-Hydroxybutyric acid/ 3-hydroxybutanoic acid	104	$C_4H_8O_3$
10	1	13.9729.min CSF1	Propanoic acid, 2-methyl-3-[(trimethylsilyl)oxy]-, trimethyl-silyl ester	89.8	C10H24O3Si2	248	3-Hydroxyisobutyric acid/3-hydroxy-2-methylpropanoic acid	104	$C_4H_8O_3$
11	1	14.1056.min CSF1	3-Acetoxybutyric acid, trimethylsilyl ester	78.4	C9H18O4Si	218	methyl 3-acetyloxybutanoate/methyl 3-methylbutanoate	116	$C_6H_{12}O_2$
12	1	15.6060.min CSF1	Ethanedioic acid, bis(trimethylsilyl) ester	76.8	C8H18O4Si2	234	Oxalic acid/ethanedioic acid	90	$C_2H_2O_4$
13	1	16.4047.min CSF1	Acetic acid, bis[(trimethyl-silyl)oxyl]-, trimethylsilyl ester	75.8	C11H28O4Si3	308	Acetic acid	60	$C_2H_4O_2$
14	1	16.6794.min CSF1	Acetoacetic acid, bis(trimethylsilyl)- deriv.	94	C10H22O3Si2	246	Acetoacetic acid/3-oxo-butanoic acid or diacetic acid	102	$C_4H_6O_3$
15	1	16.8190.min CSF1	Propanoic acid, 2-[(trimethylsilyl)oxy]-, trimethylsilyl ester	88.6	C9H22O3Si2	234	Lactic acid/2-hydroxy-propanoic acid	90	$C_3H_6O_3$
16	1	17.1011.min CSF1	Acetic acid, [(trimethylsilyl)oxy]-, trimethylsilyl ester	69.5	C8H20O3Si2	220	Acetic acid	60	$C_2H_4O_2$
17	1	18.6689.min CSF1	Butanedioic acid, bis(trimethylsilyl) ester	98.1	C10H22O4Si2	262	Succinic acid/ butanedioic acid	118	$C_4H_6O_4$
18	1	19.7910.min CSF1	2-Butenedioic acid (E)-, bis(trimethylsilyl) ester	95.8	C10H20O4Si2	260	Fumaric acid/ (E)-butenedioic acid	116	$C_4H_4O_4$
19	1	21.1614.min CSF1	Pentanedioic acid, bis(trimethylsilyl) ester	88.5	C11H24O4Si2	276	Glutaric acid/ pentanedioic acid	132	$C_5H_8O_4$
20	1	22.3688.min CSF1	(R*,S*)-3,4-Dihydroxybutanoic acid triTMS	91.7	C13H32O4Si3	336	3,4-Dihydroxybutyric acid/3,4-dihydroxy-butanoic acid	120	$C_4H_8O_4$
21	1	23.5635.min CSF1	DL-Malic acid, o-trimethylsilyl-, bis(trimethylsilyl) ester	93.2	C13H30O5Si3	350	Malic acid/hydroxyl-butanedioic acid	134	$C_4H_6O_5$
22	1	25.3692.min CSF3	Pentanedioic acid, 2-[(trimethylsilyl)oxy]-, bis(trimethylsilyl) ester	91.6	C14H32O5Si3	364	2-Hydroxyglutaric acid/2-hydroxy-pentanedioic acid	148	$C_5H_8O_5$
23	1	25.9547.min CSF1	bis(Trimethylsilyl)-2-[(trimethylsilyl)oxy]pent-2-enedioate	88.9	C14H30O5Si3	362	2-Ketoglutaric acid/ 2-oxopentanedioic acid	146	$C_5H_6O_5$
24	1	28.1451.min CSF2	Phenylpyruvic acid, tert-butyldimethylsilyl ether, tert-butyldimethylsilyl ester	62.3	C21H36O3Si2	392	Phenylpyruvic acid/ 2-oxo-3-phenyl-propanoic acid	164	$C_9H_8O_3$
25	1	28.7110.min CSF1	Trimethylsilyl [3-methoxy-4-(trimethylsilyloxy)phenyl]acetate	86.8	C15H26O4Si2	326	Homovanillic acid/2-(4-hydroxy-3-methoxy-phenyl)acetic acid	182	$C_9H_{10}O_4$
26	1	28.9209.min CSF5	Propanedioic acid, (trimethyl-silyl)-, bis(trimethylsilyl) ester	67	C12H28O4Si3	320	Tartronic acid/2-hydroxy propanedioic acid	120	$C_3H_4O_5$

27	1	30.1821.min CSF1	1,2,3-Propanetricarboxylic acid, 2-[(trimethylsilyl)oxy]-, tris-(trimethylsilyl) ester	89	C18H40O7Si4	480	Citric acid/2-hydroxy propane-1,2,3-tri-carboxylic acid	192	$C_6H_8O_7$
28	1	31.3042.min CSF1	Benzenepropanoic acid, α,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester	88.2	C18H34O4Si3	398	Hydrocinnamic acid/ benzenepropanoic acid	150	$C_9H_{10}O_2$
29	2	33.4499.min CSF1	Hexadecanoic acid, trimethylsilyl ester	93.8	C19H40O2Si	328	Palmitic acid/ hexadecanoic acid	256	$C_{16}H_{32}O_2$
30	2	34.5314.min CSF1	Heptadecanoic acid, trimethylsilyl ester	92.3	C20H42O2Si	342	Margaric acid/ heptadecanoic acid	270	$C_{17}H_{34}O_2$
31	2	36.0878.min CSF1	trans-9-Octadecenoic acid, trimethylsilyl ester	91.6	C21H42O2Si	354	trans-Oleic acid/(E)-, Octadec-9-enoic acid	282	$C_{18}H_{34}O_2$
32	2	36.3853.min CSF1	Octadecanoic acid, trimethylsilyl ester	92	C21H44O2Si	356	Stearic acid/ octadecanoic acid	284	$C_{18}H_{36}O_2$
33	2	39.0609.min CSF1	Eicosanoic acid, trimethylsilyl ester	88.5	C23H48O2Si	384	Arachidic acid/Icosanoic acid	312	$C_{20}H_{40}O_2$
34	3	37.2356.min CSF1	Octacosane	84.8	C28H58	394			
35	3	38.5871.min CSF1	Heneicosane	91	C21H44	296			
36	3	43.2087.min CSF1	Tridecane, 6-methyl-	75.7	C14H30	198			
37	3	44.6404.min CSF1	Octacosane	91.7	C28H58	394			
38	3	45.3825.min CSF1	Heptacosane	85.9	C27H56	380			
39	3	45.8443.min CSF1	Hexatriacontane	91.4	C36H74	506			
40	4	20.4015.min CSF1	Picolinic acid, trimethylsilyl ester	53.3	C9H13NO2Si	195	Picolinic acid/pyridine-2-carboxylic acid	123	$C_6H_5NO_2$
41	4	26.6722.min CSF1	l-Aspartic acid, bis(trimethylsilyl) ester	58.6	C10H23NO4Si2	277	Aspartic acid/2-amino-butanedioic acid	133	$C_4H_7NO_4$
42	4	34.6206.min CSF1	Decanoic acid, trimethylsilyl ester	54	C13H28O2Si	244	Capric acid/decanoic acid	172	$C_{10}H_{20}O_2$
43	4	36.4316.min CSF1	Cyclohexanecarboxylic acid, 1,3-dimethyl-2-[2-[3-(1-methyl-ethyl)phenyl]ethyl]-, methyl ester, (1α,2α,3α)-	50.9	C21H32O2	316			
44	4	35.5739.min CSF1	2(1H)-Naphthalenone, 7-ethynyl-4a,5,6,7,8,8a-hexahydro-1,4a-dimethyl-, (1α,4aβ,7β,8aα)-	59.3	C14H18O	202			
45	4	23.9217.min CSF1	Benzene, 1,1'-(1-methyl-ethylidene)bis[4-(2-propenyloxy)-	57.8	C21H24O2	308			

46	5	8.3179.minCSF4	Disiloxane, hexamethyl-	83.6	C6H18OSi2	162
47	5	8.6480.minCSF1	(Methoxymethyl)trimethylsilane	64.5	C5H14OSi	118
48	5	9.0732.minCSF1	Methoxydi(tert-butyl)silane	60.6	C9H22OSi	174
49	5	9.2649.minCSF1	Di-sec-butyl ether	71	C8H18O	130
50	5	9.3885.minCSF1	Pyridine, 3-trimethylsiloxy-	77.2	C8H13NOSi	167
51	5	9.7987.minCSF1	Acetamide, 2,2,2-trifluoro-N,N-bis(trimethylsilyl)-	61.8	C8H18F3NOSi2	257
52	5	9.9710.minCSF1	Silane, trimethylphenoxy-	93.8	C9H14OSi	166
53	5	10.0997.minCSF1	Silane, (2-ethoxyethoxy)-trimethyl-	80.8	C7H18O2Si	162
54	5	10.3612.minCSF1	Silanamine, 1,1,1-trimethyl-N-(trimethylsilyl)-N-[2-[(trimethylsilyl)oxy]ethyl]-	61.5	C11H31NOSi3	277
55	5	10.8539.minCSF1	tert-Butyl-(2-methoxyethoxy)-dimethylsilane	68	C9H22O2Si	190
56	5	12.2160.minCSF5	Silane, [(2-ethylhexyl)-oxy]trimethyl-	82.3	C11H26OSi	202
57	5	14.8352.minCSF2	Silane, 1,8-octanediyl-bis[trimethyl-	82.7	C14H34Si2	258
58	5	14.9376.minCSF2	Tetrasiloxane, decamethyl-	56.5	C10H30O3Si4	310
59	5	14.9702.minCSF1	Bis(trimethylsilyl)trifluoro-acetamide	67.2	C8H18F3NOSi2	257
60	5	16.2239.minCSF1	1,2-Bis(trimethylsiloxy)ethane	63.2	C8H22O2Si2	206
61	5	16.2457.minCSF1	Trisiloxane, octamethyl-	56.4	C8H24O2Si3	236
62	5	19.8894.minCSF4	Bis(trimethylsilyl)trifluoro-acetamide	66.8	C8H18F3NOSi2	257
63	5	24.4962.minCSF1	Bis(trimethylsilyl)trifluoro-acetamide	67.9	C8H18F3NOSi2	257
64	5	28.4323.minCSF1	Trimethyl[4-(1,1,3,3,-tetramethyl-butyl)phenoxy]silane	88.2	C17H30OSi	278
65	5	29.0938.minCSF1	Trimethyl[4-(1,1,3,3,-tetramethyl-butyl)phenoxy]silane	81.4	C17H30OSi	278
66	6	8.4786.minCSF4	3'-Methyl-2-benzylidene-coumaran-3-one	68.9	C16H12O2	236
67	6	8.9936.minCSF1	Tris(trimethylsilyl)borate	92.1	C9H27BO3Si3	278

68	6	18.2592.min CSF2	Trimethylsilyl ether of glycerol	86.4	C12H32O3Si3	308
69	6	20.0565.min CSF1	2,6-Bis(1,1-dimethylethyl)-4,4-dimethylcyclohexa-2,5-dien-1-one	70.6	C16H26O	234
70	6	23.3775.min CSF1	Butylated Hydroxytoluene	95.7	C15H24O	220
71	6	27.1134.min CSF1	Phenanthrene, 9,10-diethyl-3,6-dimethoxy-	73.4	C20H22O2	294
72	6	27.8183.min CSF1	1,6-Dioxacyclododecane-7,12-dione	72.3	C10H16O4	200
73	6	28.3019.min CSF1	Benzene, 2-[(tert-butyl-dimethylsilyl)oxy]-1-isopropyl-4-methyl-	65.9	C16H28OSi	264
74	6	28.3866.min CSF1	5H-Pyrano[2,3,4,5-lmn]-phenanthridin-5-one	70.2	C14H7NO2	221
75	6	28.5319.min CSF1	Propiophenone, 2'-(trimethyl-siloxy)-	68	C12H18O2Si	222
76	6	28.5988.min CSF2	1-Propene-1,2,3-tricarboxylic acid, tris(trimethylsilyl) ester, (E)-	89.6	C15H30O6Si3	390
77	6	28.7327.min CSF1	4,6-Dimethyl-2-thioxo-1,2-dihydro-3-pyridinecarbonitrile tbdms	78.7	C14H22N2SSi	278
78	6	28.7665.min CSF1	Benzestrol di-TMS derivative	71.4	C26H42O2Si2	442
79	6	29.1155.min CSF1	4-Methyl-1-tripropyl-silyloxybenzene	63.4	C16H28OSi	264
80	6	29.2168.min CSF1	4,6-Dimethyl-2-thioxo-1,2-dihydro-3-pyridinecarbonitrile tbdms	66.7	C14H22N2SSi	278
81	6	31.5148.min CSF1	Dibutyl phthalate	94.8	C16H22O4	278
82	6	34.9886.min CSF1	5β-Podocarpa-8,11,13-trien-16-oic acid, methyl ester	70.3	C18H24O2	272
83	6	36.4007.min CSF1	Benzamide, 2-methyl-N-phenyl-	74.4	C14H13NO	211
84	6	36.4539.min CSF1	Isopimaric acid TMS	61.6	C23H38O2Si	374
85	6	37.4599.min CSF1	As-Indacene, 1,2,3,6,7,8-hexahydro-1,1,6,6-tetramethyl-4-(1-methylethyl)-	68	C19H28	256
86	6	37.6905.min CSF1	1-Phenanthrenecarboxylic acid, 1,2,3,4,4a,10a-hexahydro-1,4a-dimethyl-7-(1-methylethyl)-, methyl ester, [1R-(1α,4aβ,10aα)]-	65.5	C21H28O2	312

87	6	37.7923.min CSF1	1-Phenanthrenecarboxylic acid, 7-ethenyl-1,2,3,4,4a,4b,5,6,7,8,10,10a-dodecahydro-1,4a,7-trimethyl-, methyl ester, [1R-(1α,4aβ,4bα,7α,10aα)]-, dihydro derivative.	61.6	C21H34O2	318
88	6	38.0412.min CSF1	Benz(a)anthracene, 7-methoxy-	65.3	C19H14O	258
89	6	38.1087.min CSF1	Dehydroabietic acid, trimethylsilyl ester	80.4	C23H36O2Si	372
90	6	39.7561.min CSF1	Octane, 3-ethyl-	72.4	C10H22	142
91	6	39.9827.min CSF1	1,2-Benzenedicarboxylic acid, diisooctyl ester	94.7	C24H38O4	390
92	6	42.2360.min CSF1	Oxalic acid, 2-ethylhexyl tetradecyl ester	73.7	C24H46O4	398
93	6	44.2461.min CSF1	1H-Azepine, hexahydro-1-(1-oxo-9-octadecenyl)-	71.5	C24H45NO	363
94	6	44.5340.min CSF1	2,6-Difluorobenzoic acid, 2-naphthyl ester	62.1	C17H10F2O2	284
95	7	21.8515.min CSF1	Butyric acid, 4-phenyl-, trimethylsilyl ester	68.5	C13H20O2Si	236

A2.6 R Script

The R script used to perform the required repeatability and reproducibility analysis is named “FOM.R”. A configuration file is used to specify all input to the R script. The configuration file (named “config.txt”) as well as an R script (called “interpcode.R”) containing the R code to perform monotonic spline interpolation should be sourced prior to executing the “FOM.R” script. In order to execute the KEMREP program simply open the R script “Runner.R” in R, change the location of the necessary files if needed, and run. The “Runner.R” script contains the following:

```
source("C:/Raw Data/config.txt")

source("C:/Raw Data/interpcode.R")

source("C:/Raw Data/FOM.R")
```

The data are required in two files, namely a retention time file and a concentration file. Table A2-10 illustrates representative data sets.

Table A2-10: First four columns of representative data sets.

Retention time file example

Batch	Case	8.2332.min	8.3179.min
1	1	8.2332	8.2973
1	2	8.2646	8.3224
1	3	8.257	8.3205
1	4	8.2578	8.3179
1	5	8.2583	8.3206
2	1	7.202	7.2758
2	2	7.2064	7.2778
2	3	7.202	7.2759
2	4	7.1977	7.2727
2	5	7.1962	7.2707

Concentration file example

Batc	Ca	8.2332.min	8.3179.min
1	1	4.879184	0.086521
1	2	6.079649	0.075452
1	3	4.738437	0.086139
1	4	1.780071	0.087649
1	5	2.900914	0.084357
2	1	2.537053	0.128458
2	2	2.488597	0.118794
2	3	2.066558	0.123601
2	4	0.784695	0.133367
2	5	1.615113	0.150206

An example of the configuration file required is given below.

```
config                      <-      list(

Data.dir                    =       "C:/Raw Data/" ,

RT.data.file                =       "RT.csv" ,

Concentration.data.file     =       "Conc.csv",

output.folder               =       "Output",

Apply.CASE.filter           =       FALSE ,

Percentage.of.detects       =       80 ,

Interpolation               =       1,

grid.size                   =       1000,

Ho.repeatability. rsd       =       0.3 ,

Ho.reproducibility.rsd      =       0.3 ,

CI.significance.level       =       0.05 ,

MC.number                   =       1000)
```

The configuration file must be specified by the user, with complete path description (see Data.dir). The output directory (see output.folder) is created automatically within the data directory and all output is directed to the specified directory. Checks for the data integrity are also performed, as described below:

- We assume that the data are in the following format: column x in the retention time file (specified using RT.data.file) contains the retention time of a specific variable, whereas the same column in the concentration file (specified using Concentration.data.file) contains the concentration of the same variable. A small test for this layout is conducted by comparing the variable names of these two files. If different, the script will exit with an appropriate message.
- Each case must correspond between the two files. To validate this, we check for any inconsistency in the first two columns of the concentration data when compared to the first two columns of the retention time data. If a discrepancy is observed, the script is terminated.
- The first two columns in both files should also be numeric and should contain the batch number (the reproducibility index) and the case index (the repeatability index). Data integrity is checked by first sorting both the retention time data file and the concentration data file according to column 1 (reproducibility index) and then according to column 2 (repeatability index). We then determine if cases with the same values of the reproducibility index and repeatability index occur; if so, the script is terminated.

After these initial data integrity checks, the data are further manipulated as follows:

- We assume that the data have been normalized prior to the execution of this script and we replace all missing data (retention times and concentrations) with zero.
- A case filter is applied using the retention times data file, if so specified in the configuration file (see Apply.CASE.filter). This filter is applied to all the reproducibility indexes individually. If a particular metabolite has been detected in at least one group (reproducibility item) with an acceptable percentage (see Percentage.of.detects), it will be retained. Reports are generated (in the output directory) containing one of three instances of cases, namely:

    1. The variables were detected in all reproducibility items with satisfactory levels.

2. The variables were detected in some of the reproducibility items with satisfactory levels. Note that these variables are retained for analysis, and are interesting because they contain information about non-reproducibility.
3. The variables were not detected in all of the reproducibility items with the specified satisfactory level (i.e. percentage of detects). These variables are not analysed further.

- Retention times that are zero are then input using weighted mean replacements, where each case that contains a non-zero retention time receives a weight of unity otherwise the weight is zero. This is performed for each reproducibility item, since different machines/persons/batches might have different retention times. Note that if a reproducibility item contains retention times that are all zero, we input with the weighted mean calculated by pooling all the data from all the reproducibility items.

Additional input that is required in the configuration file includes:

- Interpolation: Specify the type of interpolation used in the KEMREP method, where 1 represents linear interpolation and 2 is monotonic spline interpolation.
- grid.size: This specifies the size of the grid points where density estimates are required in the KEMREP method and is also a smoothing parameter.

Ho.repeatibility.rsd and Ho.reproducibility.rsd: This is used to specify a relative standard deviation that the user deems acceptable for repeatability and reproducibility, respectively. These values as well as those specified for the significance level (see CI.significance.level) and the Monte Carlo repetition number (see MC.number) are used to construct repeatability and reproducibility bounds in the KEMREP method. The RSD values specified here are also used as input to the goodness-of-fit methods.

Acknowledgement

We acknowledge the contributions of Ms. Natasha Booyens, Cindy Van Deventer and Cindy Irwin for the applications of KEMREP shown in the main manuscript (NB for Figures 7-6B and 7-6E and CvD for Figures 7-6C and 7-6F) and the supplementary information (CI for Figures A2-2 and A2-3).

ANNEXURE 3:

Supplementary information:

A putative urinary biosignature for diagnosis and follow-up of tuberculous meningitis in children: Outcome of a metabolomics study disclosing host–pathogen responses

Content

A3.1 Clinical information on the TBM and Non-TBM cases 276
Table A3-1 Clinical information on patients admitted to hospital as either confirmed TBM cases or suspected of meningitis but later confirmed negative (Non-TBM)

A3.2 Detailed experimental methods 277

A3.2.1 Organic acid analysis by GC-MS 277
A3.2.1.1 *Overview* 277
A3.2.1.2 *Reagents* 277
A3.2.1.3 *Sample preparation for urine analysis* 278
A3.2.1.4 *Settings for the GC and MS organic acid analyses* 279
Gas chromatograph (GC)
Mass spectrometer (MS)

A3.2.2 Amino acid analysis by GC-MS 280
A3.2.2.1 *Overview* 280
A3.2.2.2 *Reagents* 280
A3.2.2.3 *Sample preparation for urine analysis* 280
A3.2.2.4 *Settings for the GC and MS amino acid analyses* 281
Gas chromatograph (GC)
Mass spectrometer (MS)

A3.3 Detailed bioinformatics analyses 283

A3.3.1 Data pre-processing 283

A3.3.2 Data analysis 283
A3.3.2.1 *Cluster analysis for TBM and control cases using 180 variables* 283
Figure A3-1 Dendrograms from cluster analysis
A3.3.2.2 *Heat map for TBM and control cases using 180 variables* 284
Figure A3-2 Heat map of the cases and variables of TBM and controls
A3.3.2.3 *Outlier detection* 285
Figure A3-3 PCA scores plots for outlier detection
A3.3.2.4 *Multivariate analysis* 286
Figure A3-4 PCA and PLS-DA plots of the comparative multivariate scores

A3.4 Logistic regression model for the TBM biosignature 287
Table A3-2 Statistical validation values on the tentative TBM biosignature (SUM-6)
Table A3-3 Summary of model parameter estimates for the TBM biosignature (SUM-4)

A3.1 Clinical information on the TBM and Non-TBM cases

A summary of relevant clinical information pertaining to study population is shown in Table A3-1.

Table A3-1: Clinical information on patients admitted to hospital as either confirmed TBM cases or suspected of meningitis but later confirmed negative (Non-TBM).

	Final diagnosis	Ethnicity[1]	Age (year, month)	Gender	HIV status[2]	TBM stage[3]	TB contact	CXR* evidence of PTB	Evidence of Mtb outside CNS[2]
1	TBM	m	2y4m	m	neg	**2b**	y	no	neg
2	TBM	b	2y5m	f	neg	**2b**	n	no	pos gastric washing
3	TBM	m	3y2m	f	neg	**3**	y	no	pos gastric washing
4	TBM	m	11y3m	m	neg	**2a**	n	yes	not done
5	TBM	m	2y10m	m	neg	**1**	n	no	pos gastric washing
6	TBM	b	2y10m	f	neg	**2a**	y	yes	neg
7	TBM	m	3y10m	f	neg	**2a**	n	not done	neg
8	TBM	m	2y7m	m	neg	**2a**	n	yes	neg
9	TBM	b	3y2m	m	neg	**3**	y	no	neg
10	TBM	m	5y8m	m	neg	**1**	n	no	pos gastric washing
11	TBM	b	3y1m	m	neg	**3**	y	yes	pos gastric washing
12	TBM	m	3y9m	f	neg	**2b**	n	yes	pos gastric washing
1	Non-TBM	m	3y7m	f	neg	**n/a**	y	no	neg
2	Non-TBM	m	12y	f	neg	**n/a**	n	not done	not done
3	Non-TBM	m	1y6m	m	neg	**n/a**	n	no	neg
4	Non-TBM	b	2y8m	m	neg	**n/a**	y	no	neg
5	Non-TBM	b	2y1m	m	neg	**n/a**	y	yes	pos gastric washing
6	Non-TBM	b	10m	f	neg	**n/a**	y	no	neg
7	Non-TBM	m	3y5m	m	neg	**n/a**	n	not done	not done
8	Non-TBM	m	2y	f	neg	**n/a**	y	no	neg
9	Non-TBM	m	2y3m	m	neg	**n/a**	n	yes	not done
10	Non-TBM	m	4y7m	m	neg	**n/a**	n	no	neg
11	Non-TBM	b	3y5m	f	neg	**n/a**	n	no	neg
12	Non-TBM	m	2y1m	m	neg	**n/a**	n	yes	pos gastric washing
13	Non-TBM	m	1y4m	m	neg	**n/a**	n	yes	neg
14	Non-TBM	b	1y9m	m	neg	**n/a**	y	no	pos right mastoid
15	Non-TBM	m	1y8m	m	neg	**n/a**	n	no	neg
16	Non-TBM	b	6y3m	m	neg	**n/a**	n	no	neg
17	Non-TBM	b	5y4m	m	neg	**n/a**	y	no	neg
18	Non-TBM	m	2y1m	m	neg	**n/a**	n	no	neg
19	Non-TBM	m	4m	m	pos	**n/a**	n	no	neg
20	Non-TBM	m	3y11m	m	pos	**n/a**	n	no	neg
21	Non-TBM	m	4y	m	pos	**n/a**	n	yes	neg

Notation: *CXR = chest X-ray. [1]b = black, m = mixed race; [2]neg = negative, pos = positive; [3]stage 1 = Glasgow coma score (GCS) 15 and no focal neurology; stage 2a = GCS 15 plus focal neurology; stage 2b = GCS 11–14 with focal neurology; stage 3 = GCS <11.

A3.2 Detailed experimental methods

A3.2.1 Organic acid analysis by GC-MS

A3.2.1 1 Overview

Organic acids are isolated from physiological fluids with ethyl acetate and diethyl ether extractions. In order to compensate for the large differences between urine concentrations and so obtain an organic acid profile that appears similar regardless of this, a variable amount of urine, determined by urinary creatinine concentration, is extracted. The organic acid extract is evaporated to dryness under nitrogen; volatile trimethylsilyl (TMS) derivatives of the extracted organic acids are formed by heating with N,O-bis-(trimethylsilyl)trifluoraceteamine (BSTFA). The TMS derivatives are less than ideal for some classes of compounds such as acylglycine, which form mono- and di-TMS derivates, but they are still the most useful and versatile compounds for the wide range of chemical functional groups in organic acids. The derivatives are analysed by GC-MS.

GC-MS is able to separate the highly volatile organic acids by gas chromatography, followed by detection using mass spectroscopy, allowing rapid identification and quantification of organic acids with a high degree of sensitivity and chromatographic resolution.

A3.2.1 2 Reagents

Internal standard (3-phenylbutyric acid solution)

- Add precisely 26.25 mg 3-phenylbutyric acid to 50 ml distilled H_2O (dH_2O) and add 3 drops of 1M sodium hydroxide (NaOH)
- Sonicate overnight to ensure all of the 3-phenylbutyric acid is completely dissolved into solution.
- 5M Hydrochloric acid (HCl): 50 ml of 32 % conc. HCl (10M) added to 50 ml dH_2O

Ethyl acetate (HPLC grade) – distilled once to purify further

Diethyl ether (HPLC grade) – distilled once to purify further

Anhydrous sodium sulphate (Na_2SO_4)

Bis(trimethylsilyl)-trifluoracetamide (BSTFA) – stored at 4 °C

Trimethylchlorosilane (TMCS) – stored at 4 °C

Pyridine (HPLC grade)

Hexane (HPLC grade)

A3.2.1 3 Sample preparation for urine analysis

An aliquot volume of urine is added into large glass test tube according to creatinine values:

Creatinine < 100 mg % — use 1 mL urine.

Creatinine > 100 mg % — use 0.5 mL urine.

Creatinine < 5 mg % — use 2 mL urine.

Creatinine < 2 mg % — use 3 mL urine.

To each tube 6 drops 5M HCl are added to adjust to pH 1 (using glass pipette), followed by the internal standard of (IS) 5× creatinine mg% = volume in µL (e.g. if creatinine mg %.is 32.5 then add 5× 32.5 = 162.5 µL IS).

Procedure continues as follows:

1. Add 6 mL ethyl acetate.
2. Cap test tubes and check there is no leakage by inverting test tube (quality control step to ensure no sample is lost during next stage).
3. Mix for 30 min in Roto-torque and centrifuge at 3000 rpm for 3 min.
4. Aspirate the organic phase into clean large glass test tube (using glass pipette).
5. Add 3 mL diethyl ether to the aqueous (lower) phase.
6. Cap test tubes and check there is no leakage by inverting test tube.
7. Mix for 10 min in Roto-torque and centrifuge at 3000 rpm for 3 min.
8. Aspirate the organic phase and add to the ethyl acetate phase (using glass pipette).
9. Discard lower aqueous phase into appropriate organic waste container.
10. Add two level spatula scoops of anhydrous Na_2SO_4 to organic mixture.
11. Cap and vortex test tube for 5s (efficient dispersion of Na_2SO_4 ensures all water molecules removed from organic phase).
12. Centrifuge at 3000 rpm for 1min.
13. Pour/decant the organic phase into a clean small glass test tube.
14. Evaporate to dryness in heating block at 37 °C under nitrogen gas (~1 hour).
15. Using Hamilton glass syringe:

    Add BSTFA 2× creatinine mg% = volume in µL.

    Add TMCS 0.4× creatinine mg% = volume in µL.

    Add pyridine 0.4× creatinine mg% = volume in µL.

NOTE: Hamilton glass syringe is kept clean with pyridine; approximately 100 µL hexane is withdrawn into the syringe and discarded (five times) between the addition of each reagent (quality control step, to ensure glass syringe is clean and suffers no cross-contamination).

16. Cap test tubes and incubate at 60 °C for 1 hour (45 min at 70 °C).
17. Transfer approximately 100 µL sample to GC-MS vial. NOTE: Clean glass syringe with hexane (five times) after each transfer.
18. Cap GC-MS vial and place in autosampler.

A3.2.1 4 Settings for the GC and MS organic acid analyses

Gas chromatograph (GC):

- Oven program
    - 50 °C for 1 min
    - Then 20 °C/min to 60 °C
    - Then 5 °C/min to 120 °C
    - Then 7 °C/min to 280 °C
    - Post run: 300 °C for 1 min
- Run time 40 min
- Injection volume 1 µL
- Solvent wash (hexane): 4 µL pre-injection and post-injection
- Split mode (12:1)
- Split flow: 12.63 mL/min
- Pressure: ~8.2 psi (carrier gas = helium)
- Flow: 1.05 mL/min
- GC column:
    - DB-1MS
    - 340 °C: 30 m × 250 µm × 0.25 µm
    - In: front SS inlet He
    - Out: vacuum

Mass spectrometer (MS):

- MS source: 230 °C (max 250 °C)
- MS quad: 150 °C (max 200 °C)
- EM voltage: 1271 V
- Solvent delay: 7 min
- Scan range: 50–600 m/z
    - Low mass: 50.0
    - High mass: 600.0
    - Threshold: 15

A3.2.2 Amino acid analysis by GC-MS

EZ:faast™ Kit. Free (physiological) amino acid analysis by GC-MS – from Phenomenex.

A3.2.2 1 Overview

The EZ:faast amino acid analysis procedure consists of a solid phase extraction step followed by derivatization and liquid/liquid extraction; derivatized samples are quickly analysed by gas chromatography–mass spectrometry. The solid phase extraction is performed via a sorbent-packed tip that binds amino acids while allowing interfering compounds to flow through. Amino acids on sorbent tip are then extruded into the sample vial and quickly derivatized with reagent at room temperature in aqueous solution. Derivatized amino acids concomitantly migrate to the organic layer for additional separation from interfering compounds. Organic layer is then removed, evaporated, and re-suspended in re-dissolution solvent and analysed by GC-MS. Total sample preparation time takes around 8 minutes and analysis is performed in approximately 7 minutes for a total start to finish time of about 15 minutes.

The EZ:faast method has been developed by Phenomenex for the analysis of more than 60 aliphatic and aromatic amino acids, including primary and secondary amines.

A3.2.2 2 Reagents

Reagent 1 (internal standard solution) – norvaline 0.2 mM + N-propanol 10 %
Reagent 2 (washing solution) – N-propanol
Reagent 3A (eluting medium component 1) – sodium hydroxide
Reagent 3B (eluting medium component 2) – N-propanol
Reagent 4 (organic solution 1) – chloroform
Reagent 5 (organic solution 2) – iso-octane
NOTE: Reagents 1, 3B and 4 stored at 4 °C. Other reagents stored at room temperature. The toxicity of all reagents is low but all steps must be conducted in an exhaust-hood and appropriate protection must be worn. Hexane (HPLC grade)

A3.2.2 3 Sample preparation

1. For each sample, line up one glass sample preparation vial in the vial rack.
2. Pipette sample (100 µL urine) and 100 µL Reagent 1 (internal standard solution) into each sample preparation vial.

3. Attach a sorbent tip to a 1.5 mL syringe and loosen the syringe piston; immerse the tip and let the solution in the sample preparation vial pass through the sorbent tip by slowly pulling back the syringe piston, in small steps.
4. Pipette 200 µL Reagent 2 (washing solution) into the same sample preparation vial. Pass the solution slowly through the sorbent tip and into the syringe barrel. Drain the liquid from the sorbent bed by pulling air through the sorbent tip. Detach the sorbent tip and leave it in the sample preparation vial, then discard the liquid accumulated in the syringe.
5. Prepare the eluting medium as follows:

Number of samples	Reagent 3A eluting	Reagent 3B eluting
2	300 µL	200 µL
4	600 µL	400 µL
7	900 µL	600 µL
12	1.5 mL	1.0 mL
14	1.8 mL	1.2 mL
19	2.4 mL	1.6 mL
24	3.0 mL	2.0 mL
29	3.6 mL	2.4 mL
34	4.2 mL	2.8 mL

6. Pipette 200 µL freshly prepared eluting medium into same sample preparation vial.
7. Pull back the piston of a 0.6 mL syringe halfway up the barrel and attach the sorbent tip used in 3–8.
8. Wet the sorbent with the eluting medium; watch as the liquid rises through the sorbent particles and stops when the liquid reaches the filter plug in the sorbent tip.
9. Eject the liquid and sorbent particles out of the tip and into the sample preparation vial. Repeat steps 7 and 8 until the sorbent particles in the tip are expelled into the sample preparation vial.
10. Pipette 50 µL Reagent 4 (organic solution 1), using Drummond microdispensor, into the sample preparation vial.
11. Emulsify the liquid in the vial by vortexing for 5–8 seconds, allow to stand for 1 min (perform step 11 three times).
12. Pipette 100 µL Reagent 5 (organic solution 2), using Drummond microdispensor, into the sample preparation vial and repeat vortexing procedure in step 11.
13. After allowing the reaction to proceed for 1 min, transfer part of the (upper) organic layer (about 50–100 µL) using a Hamilton glass syringe into an autosampler vial for GC-MS analysis.

NOTE: clean Hamilton glass syringe after each transfer by collecting approximately 100 µL hexane and discarding (repeat 5 times).

A3.2.2 4 Settings for the GC and MS amino acid analyses

Gas chromatograph (GC):

- Temperature range: min 60 °C, max 325 °C
    - Increase 50 °C/min until 110 °C
    - Then 20 °C/min until 185 °C
    - Then 25 °C/min until 235 °C
    - Then 30 °C/min until 320 °C
- Run time: 11.58 min
- Splitless mode
- Pressure: 30.8 kPa
- Total flow: 9.4 mL/min (carrier gas = He)
- Injection volume: 3 µL
- Solvent washes (hexane): 4
- GC coil/column:
    - Phenom ZB-AAA
    - 10.0 m length, 250 µm diameter, 0.25 µm film thickness
    - Max temperature: 320 °C
    - Constant flow: avg 92 cm/sec

Mass spectrometer (MS):

- MS source: 240 °C (max 250 °C)
- MS quad: 180 °C (max 200 °C)
- Scan range: 40–450 m/z
    - Low mass: 40.0 m/z
    - High mass: 450.0 m/z
    - Threshold: 100 m/z
- EM voltage: 2000 eV
- Solvent delay: 2.50 min

A3.3 Detailed bioinformatics analyses

A3.3.1 Data pre-processing

Data from GC-MS are deconvoluted, identified, annotated and quantified using AMDIS (Automated Mass Spectral Deconvolution and Identification System), which is a software program for spectral extraction and compound identification of GC-MS data. Deconvolution allows overlapping signals to be separated into individual chemical peaks. Subsequent chromatographic peaks are identified and annotated according to relevant recorded spectral libraries, based upon similarities of spectra and retention times within the chromatogram. A report is generated by AMDIS, from which the concentration is calculated in µmoles per mmole creatinine. These reports are amalgamated into one (*n*x*p*) data matrix composed of *p* variable names (compound names) and *n* quantities or concentrations (*X*) in Excel format. The data matrix follows the format as shown below and is then used for statistical pattern recognition and for the generation of metabolomics data.

$$\underset{(n \times p)}{\boldsymbol{X}} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}$$

A3.3.2 Data analysis

A3.3.2 1 Cluster analysis for TBM and control cases using a matrix with 180 variables

Figures A3-1 and A3-2 show dendrograms based on hierarchical clustering of cases as well as variables. Cluster analysis is an unsupervised technique, that is, no group information is made available to the method making it robust against overfit. The analysis therefore identifies natural groupings of objects (cases or variables) based exclusively on the data matrix provided. The analyses performed above made use of Ward linkage and Euclidean distance. A dendrogram as presented in Figures A3-1 and A3-2 allows us to visualize the results of a cluster analysis with each pair of objects being connected by a line. The length of the line connecting two objects represents the distance or difference between them. The levels of the diagram therefore represent groupings since the shorter the line the more similar the objects.

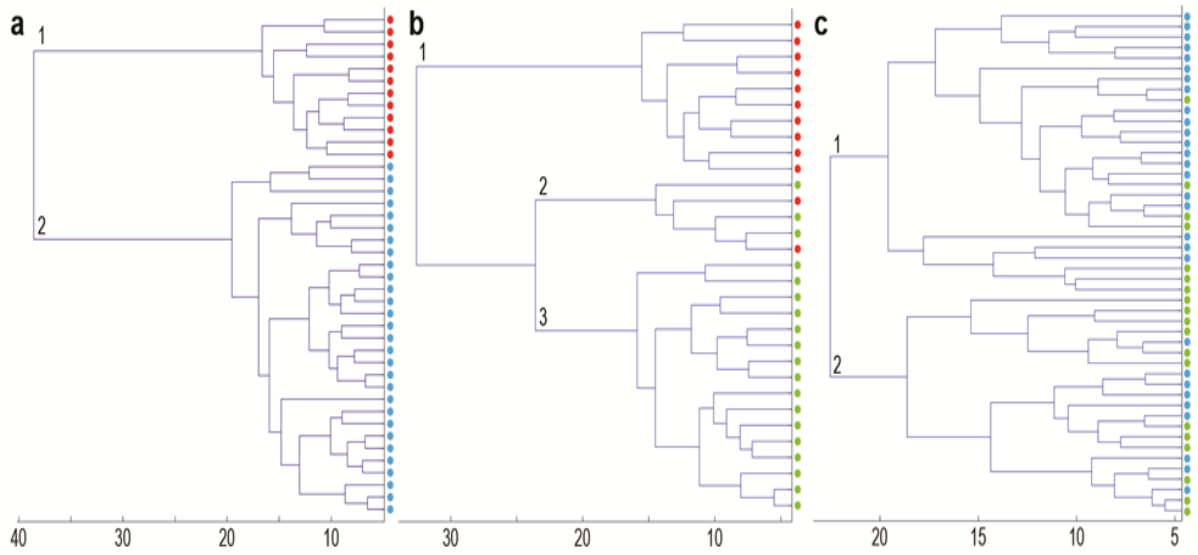


Figure A3-1 Dendrograms from cluster analysis indicating (**a)** complete separation of TBM (red dots) and controls (blue dots), (**b)** near-complete separation of TBM (red dots) and non-TBM (green dots), and (**c)** no class separation of non-TBM (green dots) and controls (blue dots).

A3.3.2 2 Heat map for TBM and control cases using 180 variables

The heat map shown in Figure A3-2 gives a colour-coded display of the values observed in the data matrix as indicated by the colour bar provided. Note that negative values were not observed, as was to be expected in GC-MS analysis. By combining such a visualization of the data with dendrograms showing the clustering of cases (left) and variables (top) we are able to identify regions in the data matrix that differ between groups as well as whether these regions have higher or lower observed values. This image was generated using Matlab functions clustergram and colorbar.

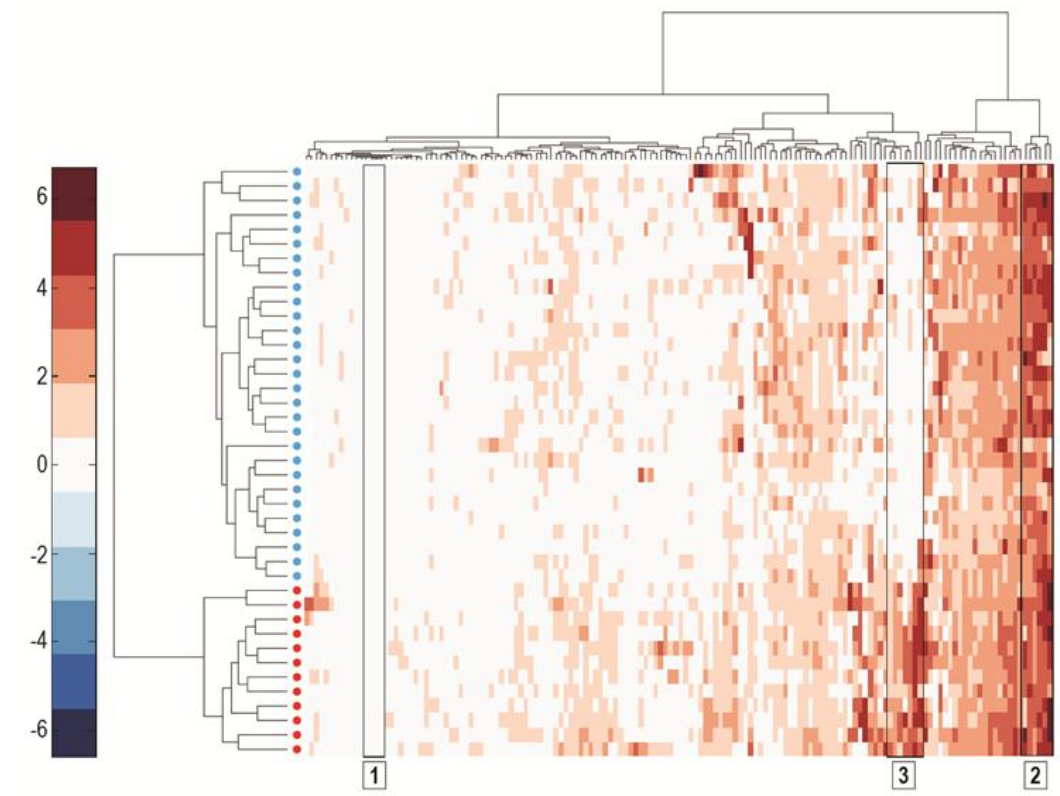


Figure A3-2 Heat map of the cases (left dendrogram) as well as variables (top dendrogram) of TBM (red dots) and controls (blue dots). The blocks numbered 1 to 3 indicate areas of particular interest in the figure: (1) variables with values in the range 0–1 across all cases (i.e. weak discriminators); (2) variables with values in the range 4–6 across all cases; and (3) variables with higher values for the TBM cases compared to the controls, indicating that these variables are good discriminators.

A3.3.2 3 Outlier detection

After performing a shifted log transformation with shift parameter equal to one, the data was used for outlier detection by looking at the scores from a principal component analysis (PCA) of each group, as shown in Figure A3-3, in conjunction with the resulting Hotelling's T^2 values.

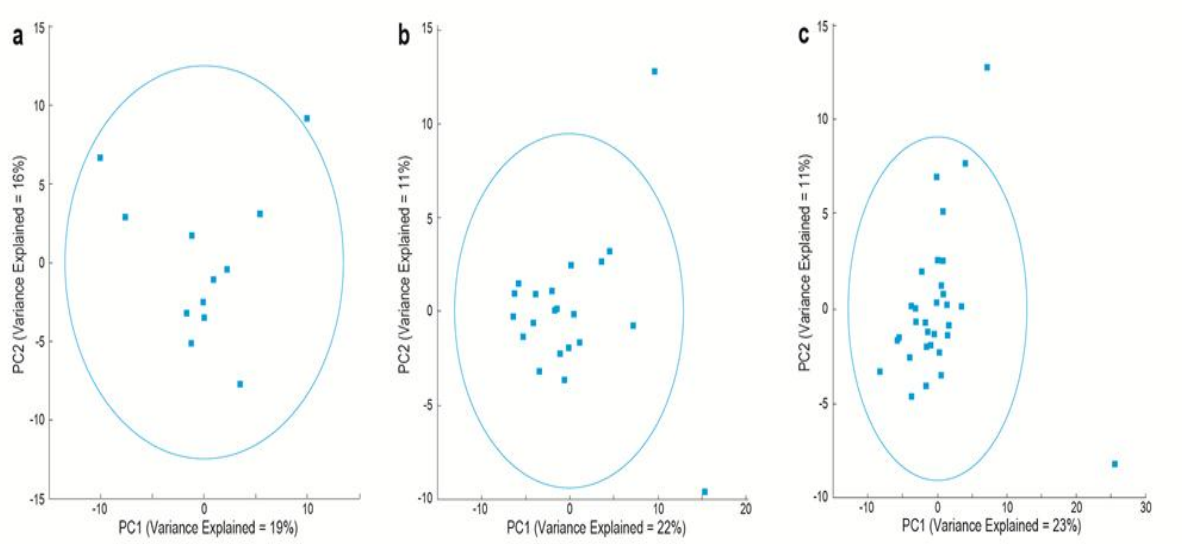


Figure A3-3 PCA scores plots for outlier detection, where (**a)** represents the TBM patients, (**b**) the non-TBM cases, and (**c**) the controls. The ellipses represent 90 % confidence intervals. Cases beyond the 90% confidence intervals were regarded as outliers

A3.3.2 4 Multivariate analysis

Unsupervised PCA was performed to compare the variability of the three experimental groups pairwise, and the results are shown in Figures A3-4(a-c). Subsequently, a supervised PLS-DA was performed to maximize the discrimination between the groups and to identify the metabolites which discriminated between the groups. The outcome of these analyses are shown in Figures A3-4(d-e). A total of 26 metabolites were identified as being the most important discriminatory variables on the basis of the VIP values (VIP > 0.9)from the PLS-DA and univariate assessments (Fold Change (FC): TBM/control > 1.0 with a significant difference at *p* value < 0.05 based on Wilcoxon–Mann-Whitney test).

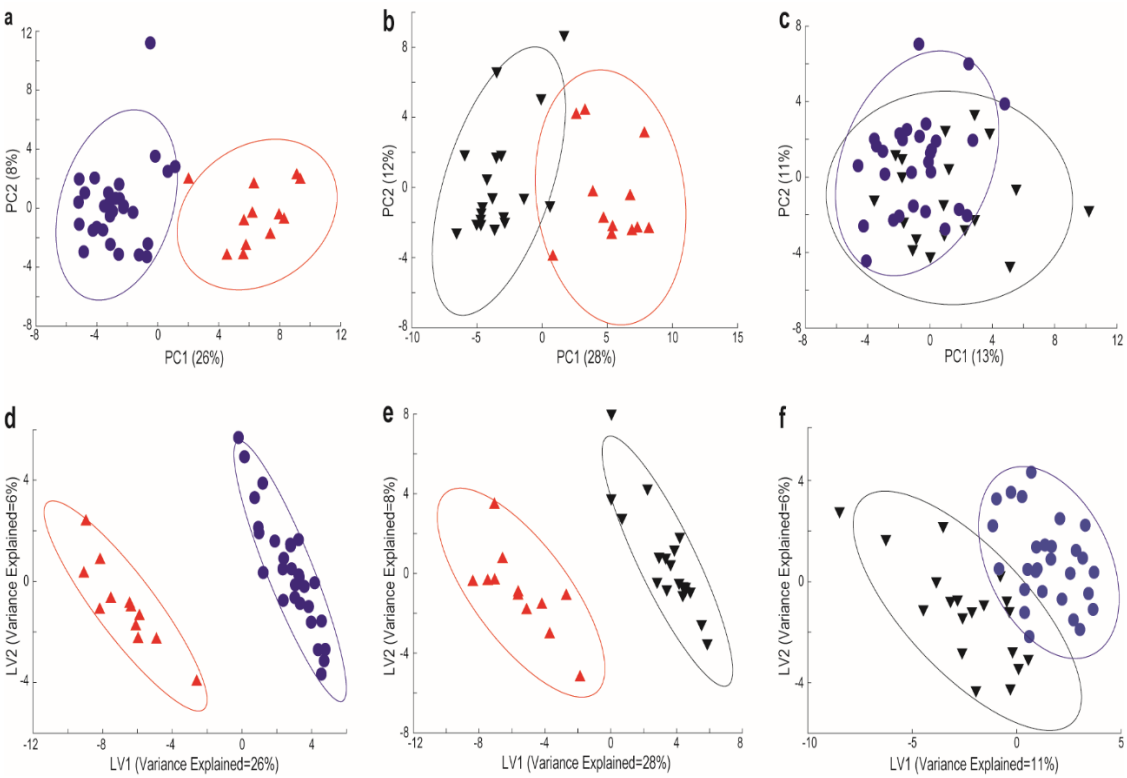


Figure A3-4 PCA and PLS-DA scores plots based on 150 variables for each pair of groups. Data from PCA (**a** to **c**) and PLS-DA (**d** to **f)** scores are shown for three experimental groups: (**a** and **d**) for the TBM and control groups; (**b** and **e**) for the TBM and non-TBM groups; (**c** and **f**) for the non-TBM and control groups. The $R^2(X)$ and $Q^2(Y)$ values for the three PLS-DA models were: $R^2(X)$ = 89.5 %, $Q^2(Y)$ = 82.4 % for **d**; $R^2(X)$ = 86.0 %, $Q^2(Y)$ = 77.4 % for **e** and $R^2(X)$ = 53.0 %, $Q^2(Y)$ = 13.8 % for **f**

A3.4 Logistic regression model for the TBM biosignature

We evaluated the respective contribution of each of the six biomarkers individually towards a TBM diagnostic model. For this purpose we (1) fitted a logistic regression model (2) looked at the Hosmer–Lemeshow statistical test for goodness of fit for the logistic regression model, and (3) performed a ROC analysis for each of the individual biomarkers constituting SUM-6. We did not make use of stepwise logistic regression to select predictors, because adding predictors individually would result in a slightly more complex model to apply in practice. The outcome of the logistic regression analysis and PCA-power data (Table A3-2) indicated that quinolinic, methylcitric, 4-hydroxyhippuric and 2-ketoglutaric acids outperformed the two other components as the best discriminating metabolites for a TBM biosignature.

Table A3-2: Statistical validation values regarding metabolites that comprise the SUM-6 biosignature for TBM. Note the *p*-value reported for the AUC does not relate to testing for a drop in AUC after cross-validation as reported previously. Here we test whether the AUC is different from 0.5.

Biomarker (selected metabolite)	AUC (%)	AUC (*p*-value)	H&L test (*p*-value)	Model Coeff. (*p*-value)	PCA (power)
Methylcitric acid	99.7	<0.0001	0.98	0.015	0.53
Quinolinic acid	99.4	<0.0001	0.993	0.027	0.57
4-Hydroxyhippuric acid	95.4	<0.0001	0.452	0.008	0.54
2-Ketoglutaric acid	91.7	<0.0001	0.038	0.012	0.3
Kynurenic acid	75	0.013	n/a	n/a	0.17
Isocitric acid	73	0.022	0.025	0.452	0.17

Table A3-3 provides the parameter estimates of the logistic regression model used in the ROC analysis of SUM-4 when distinguishing TBM patients on first encounter from controls. The table indicates that both the model intercept term as well as the parameter estimate for SUM-4 are significant, with $p = 0.0035$ and $p = 0.0192$, respectively.

Table A3-3: Summary of model parameter estimates for TBM vs Control based on a binary logistic model with the sum of the four markers as predictor.

Analysis of maximum likelihood estimates					
Parameter	Degree of freedom	Estimate	Standard error	Wald Chi-square	*p*-value
Intercept	1	-4.450	1.526	8.505	0.0035
Sum-4	1	0.123	0.052	5.486	0.0192

ANNEXURE 4:

Supplementary information:

Cerebrospinal fluid in tuberculous meningitis exhibits only the L-enantiomer of lactic acid

Contents:

A4.1 Detailed sample preparation and UPLC-ESI-MS/MS analysis 290
Sampling 290
Chemicals 290
Sample preparation and UPLC-ESI-MS/MS analysis 291

A4.2 Additional statistical analyses 292
Parametric correlations 292
Nonparametric correlations 293
Regression 294

A4.1 Detailed sample preparation and UPLC-ESI-MS/MS analysis

Sampling

The experimental group consisted of infants and children (<13 years of age) (n = 20) from the Western Cape region of South Africa, who were directed from local clinics to the paediatric unit of Tygerberg Hospital, Cape Town, on suspicion of meningitis, based on clinical symptoms. Upon admission to hospital, a lumbar CSF sample was taken for differential diagnosis, a portion of which was stored at –80°C for this study, which was used to confirm a diagnosis of TBM. A diagnosis of TBM was based on the uniform research case definition of Marais *et al.* 2010. Only children with 'definite' and 'probable' TBM were included. TBM was classified as 'definite' when CSF demonstrated acid-fast bacilli on microscopy, a positive Mtb culture and/or passed a positive CSF Mtb commercial nucleic acid amplification test in a child with symptoms or signs suggestive of the disease. TBM was classified as 'probable' according to a scoring system based on clinical, cerebrospinal fluid and neuroimaging criteria, as well as evidence of extraneural TB. Clinical details of the patients are described in the supplementary information of Mason *et al.* 2015. For the purpose of the present study the stage of TBM, as well as the glucose concentration in 16 out of the 20 CSF samples, were made available, as shown in Fig. 2. We also collected a urine sample from our subjects upon admission to hospital. Urine is often the biofluid of choice for investigating the various potential sources of lactic acid, as well as providing for a non-invasive mode of sample collection. A limitation in using urine remains the unpredictable fluctuation in the concentration of targeted metabolites linked to the disease state of the patients. Informed consent was obtained from all participants and this study was approved by the Human Research Ethics Committee of Stellenbosch University, South Africa (study no. N11/01/006).

Chemicals

The chemicals used as standards were: sodium L-lactic acid ≥99.0% (Sigma-Aldrich 71718, CAS: 867-56-1); sodium D-lactic acid ≥99.0% (Sigma-Aldrich 71716, CAS: 920-49-0); sodium L-lactic acid-3,3,3-d3 (CDN isotopes D-2646, CAS: 79-33-4).

The following chemicals were used in sample preparation and analyses: (+)-O,O'-Diacetyl-L-tartaric anhydride (DATAN) (Sigma-Aldrich 358924, CAS: 6283-74-5); acetonitrile HPLC supragrade (Biosolve 01203502, CAS: 75-05-8); MilliQ water (Millipore, CAS: 7732-18-5); acetic acid 100% (Merck 1.00063.1000, CAS: 64-19-7); ammonium formate (Sigma-Aldrich 25204, CAS: 540-69-2); dichloromethane (Lab-Scan AR1040A, CAS: 75-09-2); perchloric acid (Sigma-Aldrich 244252, CAS: 7601-90-3); formic acid 98-100% (Merck 1.00264.1000, CAS: 64-18-6).

Sample preparation and UPLC-ESI-MS/MS analysis

A dilution series of L- and D-lactic acid and a 2.5 mM L-lactic acid-d_3 internal standard (IS) solution were prepared in advance for the calibration curve. A fresh diacetyl-L-tartaric anhydride (DATAN) derivatization solution was prepared by dissolving 250 mg of DATAN in 4 ml dichloromethane and 1 ml acetic acid. Samples were prepared by combining 50 µL CSF / 100 µL urine with 20 µL internal standard solution and 300 µL acetonitrile (ACN) in an Eppendorf tube. Samples were vortexed and centrifuged for 10 minutes at 3000 RPM in order to remove proteins and other macromolecules. The supernatant was transferred to a clean glass vial and dried under nitrogen gas at 40 °C. To this, 100 µL DATAN solution was added, vortexed and incubated at 75 °C for 30 minutes. Once again the solution was dried under nitrogen gas at 40 °C and then reconstituted in 200 µL ACN / H_2O (1/2 v / v) for UPLC-ESI-MS/MS analysis. For the UPLC: eluent A (2.5 mM ammonium formate, pH = 3.6) and eluent B (100% acetonitrile) were used. Samples were analysed on a Waters Acquity UPLC hyphenated to a Quattro premier XE triple quadrupole mass spectrometer using negative ion electrospray ionization (ESI), and the following multiple reaction monitoring (MRM) parameters:

Parent ion	Daughter ion	Dwell (s)	Cone (eV)	Collision (eV)	Delay (s)	Compound
305	89	0.200	10	8	0.005	L- & D-lactic acid
308	92	0.200	10	8	0.005	L-lactic acid-d_3

The UPLC column used was an Acquity UPLC bridged ethylene hybrid (BEH) C18 reverse phase with 1.7 µm particle size, 2.1 mm diameter and 100 mm length. Sample temperature was kept at 10 ºC in the autosampler and column temperature at 50 ºC. Injection volume was 10 µL. Source and desolvation temperatures were set at 150 ºC and 450 ºC, respectively. Capillary voltage was set at 1.50 kV. Cone and desolvation gas flow 50 L/h and 900 L/h, respectively. Sample run consisted of, first, blank + IS, followed by 5 calibration curve series samples, blank and then the prepared samples After derivatization with DATAN, L- and D-lactic acids were separated from each other on the UPLC column and measured by MS/MS in MRM mode. Use of solvent delay time avoided the introduction of salts in the MS. Results were quantified in terms of the stable isotope-labelled analogue of L-lactic acid. The peak areas corresponding to component-specific transitions were converted via calibration lines to concentrations.

A4.2 Additional statistical analyses

These statistics showed a significant correlation between Glasgow coma score and TBM stage at the 0.01 level. However, there was a significant correlation at the 0.05 level between Glasgow coma score and glucose, between TBM stage and poor prognosis, and between TBM stage and glucose. The significance of glucose is also seen in the regression model. Lactate did not show up as statistically significant in this small sample group.

Parametric correlations

		Glasgow coma score	Outcome severity score	TBM stage	lactate	glucose
Glasgow coma score	Pearson correlation	1	-.404	-.862**	.029	-.526*
	Sig. (2-tailed)		.120	.000	.915	.036
	N	16	16	16	16	16
Outcome severity score	Pearson correlation	-.404	1	.514*	-.122	.215
	Sig. (2-tailed)	.120		.041	.654	.425
	N	16	16	16	16	16
TBM stage	Pearson correlation	-.862**	.514*	1	.129	.502*
	Sig. (2-tailed)	.000	.041		.634	.047
	N	16	16	16	16	16
lactate	Pearson correlation	.029	-.122	.129	1	-.213
	Sig. (2-tailed)	.915	.654	.634		.428
	N	16	16	16	16	16
glucose	Pearson correlation	-.526*	.215	.502*	-.213	1
	Sig. (2-tailed)	.036	.425	.047	.428	
	N	16	16	16	16	16

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Nonparametric correlations

			Glasgow coma score	Outcome severity score	TBM Stage	lactate	glucose
Kendall's tau_b	Glasgow coma score	Correlation Coefficient	1.000	-.253	-.770**	.075	-.331
		Sig. (2-tailed)		.240	.000	.704	.096
		N	16	16	16	16	16
	Outcome severity score	Correlation Coefficient	-.253	1.000	.442*	-.117	.225
		Sig. (2-tailed)	.240		.042	.560	.263
		N	16	16	16	16	16
	TBM stage	Correlation Coefficient	-.770**	.442*	1.000	.085	.409*
		Sig. (2-tailed)	.000	.042		.671	.042
		N	16	16	16	16	16
	lactate	Correlation Coefficient	.075	-.117	.085	1.000	.025
		Sig. (2-tailed)	.704	.560	.671		.892
		N	16	16	16	16	16
	glucose	Correlation Coefficient	-.331	.225	.409*	.025	1.000
		Sig. (2-tailed)	.096	.263	.042	.892	
		N	16	16	16	16	16
Spearman's rho	Glasgow coma score	Correlation Coefficient	1.000	-.292	-.834**	.089	-.465
		Sig. (2-tailed)		.272	.000	.742	.069
		N	16	16	16	16	16
	Outcome severity score	Correlation Coefficient	-.292	1.000	.500*	-.162	.268
		Sig. (2-tailed)	.272		.048	.548	.316
		N	16	16	16	16	16
	TBM stage	Correlation Coefficient	-.834**	.500*	1.000	.109	.523*
		Sig. (2-tailed)	.000	.048		.689	.038
		N	16	16	16	16	16
	lactate	Correlation Coefficient	.089	-.162	.109	1.000	.001
		Sig. (2-tailed)	.742	.548	.689		.996
		N	16	16	16	16	16
	glucose	Correlation Coefficient	-.465	.268	.523*	.001	1.000
		Sig. (2-tailed)	.069	.316	.038	.996	
		N	16	16	16	16	16

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Regression

Model Summary

Model	R	R square	Adjusted R square	Std. error of the estimate
1	.557[a]	.311	.205	.9770

a. Predictors: (Constant), glucose, lactate

ANOVA[a]

Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	5.591	2	2.796	2.929	.089[b]
	Residual	12.409	13	.955		
	Total	18.000	15			

a. Dependent variable: TBM stage

b. Predictors: (Constant), glucose, lactate

Coefficients[a]

		Unstandardized coefficients		Standardized coefficients			95.0% Confidence interval for B	
Model		B	Std. error	Beta	t	Sig.	Lower bound	Upper bound
1	(Constant)	1.054	.759		1.390	.188	-.585	2.694
	lactate	.103	.098	.247	1.049	.313	-.109	.315
	glucose	.450	.191	.555	2.355	.035	.037	.863

a. Dependent variable: TBM stage

ANNEXURE 5: COPYRIGHT LICENCING AGREEMENTS

Part 1: cover image:

The College of Physicians of Philadelphia

Image Information

	Artist / Author	Title / Description of Requested Images	Accession, Library or Catalog #	Proposed Image Usage (e.g. cover or illustration of text)	Image Resolution Requested:
1	H. Koplik	Diseases of Infancy and Childhood, 4th ed. (1918)	Fb159c	Text	~~[ADD CHECK BOXES] – EVI WILL EDIT THIS SECTION~~
2					
4					
5					
6					
7					
8					
9					
10					

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specific conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement. This institution reserves the right to refuse a copying or reproduction order, if, in its judgment, fulfillment of the order would involve violation of copyright law. The user is responsible for conforming to the laws of libel, publicity rights, and copyright, which may be involved in the use of these materials. Fair use under the Copyright Act may be found in §107.

College use only: One-time publication rights are hereby conferred to the User named above, subject to the manner set forth above and the TERMS AND CONDITIONS described in this Agreement and incorporated by reference herein.

Date Request Received: 2/19/16
Estimated Fee: $ 0
Approved: Yes ☑ No ☐
Signature: Beth Landy
Title: College Librarian
Date: February 19, 2016

Chapter 1: Background to tuberculous meningitis

Figure 1-1: Global incidence rates of TB for 2013 as reported by the World Health Organization (WHO) in the Global Tuberculosis Report 2014 (WHO 2014).

According to WHO website: “If you wish to use the extract for research, private study or in a noncommercial document with limited circulation (such as an academic thesis or dissertation), you may do so without seeking permission. Our only requirement is that the WHO source should be appropriately acknowledged.”
URL: http://www.who.int/about/licensing/extracts/en/

Figure 1-2: Possible outcomes of Mtb infection (with permission from Young et al. 2008).

Order Details:
Licensee: Shayne Mason
License Date: Sep 1, 2015
License Number: 3700220168417
Publication: Nature Reviews Microbiology
Title: Systems biology of persistent infection: tuberculosis as a case study
Type Of Use: reuse in a dissertation / thesis.
Your order details and publisher terms and conditions are available by clicking the link below:
http://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=f1be2a26-c6d7-4487-9786-70ab378358bf

Figure 1-3: Life cycle of the granuloma — from formation to lysis (with permission from Russell et al. 2010).

Order Details:
Licensee: Shayne Mason
License Date: Sep 1, 2015
License Number: 3700200427848
Publication: Science
Title: Tuberculosis: What We Don't Know Can, and Does, Hurt Us
Type Of Use: Thesis / Dissertation.
Your order details and publisher terms and conditions are available by clicking the link below:

http://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=12250684-ed73-4e97-aa92-228cbd03f570

Figure 1-4: Predicted pathways of propionate metabolism in Mtb as proposed by Savvi et al. (2008).

Permissions Request

ASM authorizes an advanced degree candidate to republish the requested material in his/her doctoral thesis or dissertation. If your thesis, or dissertation, is to be published commercially, then you must reapply for permission.

Figure 1-5: Side chain β-oxidation by Mtb on cholesterol (taken from Thomas *et al.* 2011).

To whom it may concern,

It is the policy of the American Society for Biochemistry and Molecular Biology to allow reuse of any material published in its journals (the Journal of Biological Chemistry, Molecular & Cellular Proteomics and the Journal of Lipid Research) in a thesis or dissertation at no cost and with no explicit permission needed. Please see our copyright permissions page on the journal site for more information.

Best wishes,
Sarah Crespi
American Society for Biochemistry and Molecular Biology
11200 Rockville Pike, Rockville, MD
Suite 302
240-283-6616
JBC | MCP | JLR

URL: http://www.jbc.org/site/home/about/permission_letter.pdf

Figure 1-6: Integrated model of routes and regulation in the Mtb citric acid cycle as proposed by Baughn et al. (2009).

Chapter 2.6: Template paper outlining the first step in biomarker identification

Disclosure of a putative biosignature for respiratory chain disorders through a metabolomics approach

Chapter 4: A hypothetical astrocyte–microglia lactate shuttle derived from a ^{1}H NMR metabolomics analysis of cerebrospinal fluid from a cohort of South African children with tuberculous meningitis

Chapter 7: KEMREP: A new qualitative method for the assessment of an analyst's ability to generate a metabolomics data matrix by gas chromatography–mass spectrometry

Order Details:
Licensee: Shayne Mason
License Date: Sep 8, 2015
License Number: 3704110022545
Publication: Current Metabolomics
Title: KEMREP: A New Qualitative Method for the Assessment of an Analyst's Ability to Generate a Metabolomics Data Matrix by Gas Chromatography- Mass Spectrometry

Type Of Use: Thesis/Dissertation

Your order details and publisher terms and conditions are available by clicking the link below:
http://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=e2ff4cf0-7598-4571-b28e-4b2960d49b7d

Chapter 8: A putative urinary biosignature for diagnosis and follow-up of tuberculous meningitis in children: Outcome of a metabolomics study disclosing host–pathogen responses

Submitted: *Metabolomics,* 2016.

Part 4: cover image

Shutterstock hereby grants you a non-exclusive, non-transferable right to use, modify and reproduce Visual Content worldwide, in perpetuity, as expressly permitted by the applicable license and subject to the limitations set forth.

Chapter 9:

Figure 9-1: Schematic representation of ANLS model. Glu=glutamate, Gln=glutamine, GluR=glutamatergic receptor. (taken from Bélanger *et al.* 2011).

Order Details
Licensee: Shayne Mason
License Date: Mar 22, 2016
License Number: 3834250833741
Publication: Cell Metabolism
Title: Brain Energy Metabolism: Focus on Astrocyte-Neuron Metabolic Cooperation
Type Of Use: reuse in a thesis/dissertation
Your order details and publisher terms and conditions are available by clicking the link below:
http://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=8f85796c-0c05-487e-aae3-ac7c5089798d

Chapter 10.1: Cerebrospinal fluid in tuberculous meningitis exhibits only the L-enantiomer of lactic acid

Accepted: *BMC Infectious Diseases,* 2016 – in press.