Host serum protein biomarkers of tuberculosis in sub-Saharan Africa

Dr Thomas C Morris MA (Cantab) MRCP

Department of Infectious Disease

Imperial College London

Submitted for the degree of:

Doctor of Medicine (Research)

April 2020

Declarations

Originality

The content of this thesis is my own original work, except for the biosignature analyses for which statistical input was received. These analyses were performed by Dr Clive Hoggart PhD, Imperial College London and Prof Martin Kidd PhD, University of Stellenbosch. Results from these are acknowledged at the relevant points in the thesis.

In addition, this work relied upon the significant prior work of others, specifically:

a) Recruitment of patients to the European Union-funded Tuberculosis (EU-TB) study by the ILULU consortium;
b) Recruitment of patients to "the longitudinal study" by Dr Elizabeth Whittaker and colleagues;
c) Discovery of a novel 7-protein diagnostic signature and 22 individual biomarkers of TB in the African European TB Consortium study;
d) Discovery of novel proteomic diagnostic signatures of TB in the EU-TB cohort by Drs Shea Hamilton and Lachlan Coin, on behalf of the ILULU-TB consortium.

All of the immunoassays and statistical analyses other than those listed above were performed by me. Throughout the thesis, the pronoun "we" is used to reflect the impersonality of science: all discoveries result at some level from the contributions of multiple people, and one can only build upon what has been discovered before. As an unnamed poet cited by Claud Bernard in his "Introduction to the Study of Experimental Medicine" in 1865, "Art is I; Science is we."

Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

It would not have been possible to complete this thesis without help, input and support of many kinds from a large number of people. I would like to thank and acknowledge the following parties for their contributions to each of the following areas:

i) Supervision
   a. Dr Shea Hamilton (Department of Infectious Disease, Imperial College London);
   b. Prof Robert J Wilkinson (Department of Infectious Disease, Imperial College London; Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, South Africa; The Francis Crick Institute, London);
   c. Prof Mike Levin (Department of Infectious Disease, Imperial College London);

ii) Funding
   a. The Wellcome Trust (104803);
   b. The Royal College of Physicians (James Maxwell Grant Prophit Fellowship 2016);

iii) Serum/plasma
   a. ILULU Consortium (listed below);
   b. Dr Elizabeth Whittaker (Department of Infectious Disease, Imperial College London);

iv) Laboratory Facilities
   a. Sena Plevnik (Laboratory Manager, Section of Paediatric Infectious Disease, Imperial College London);
   b. Jane Webb (Secretary, Section of Paediatric Infectious Disease, Imperial College London);
   c. St Mary's Flow Cytometry Core Facility, Imperial College London;

v) Statistical Analysis
   a. Dr Clive Hoggart (Department of Infectious Disease, Imperial College London);
   b. Prof Martin Kidd (Centre for Statistical Consultation, Stellenbosch University, South Africa);

vi) Collaboration
   Prof Novel Chegou and Prof Gerhard Walzl (DST-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Stellenbosch University, South Africa);

vii) Progress Review Panel
   a. Dr Brian Robertson (Department of Infectious Disease, Imperial College London);
   b. Prof Charles Bangham (Department of Infectious Disease, Imperial College London).

To expand, I would like to thank each of my supervisors in turn for the significant amount of time and effort they have given me. To Shea: for your untiring dedication to every aspect of this thesis. From the research plan, through supervising the laboratory work, to monitoring my progress, teaching about proteomics, and generally keeping me on track. This also included reviewing every word document and powerpoint presentation I ever wrote, often in your own time, and always with an incredibly prompt and thorough response. Thank you. To Rob: I cannot thank you enough for inviting me into and introducing me to the world of medical research. I would not have gained access to any of the educational and scientific opportunities that I have had over the last four years if it weren't for you. In addition, your huge generosity in keeping me going financially in between grants, and your commitment to regular personalised lab meetings, reviewing grant applications, manuscript iterations and chapters of this thesis has been immense. And to Mike: thank you so much for everything you did in hosting me for this work. It was a very valuable experience to be part of your research group for the time at Imperial, and eye-opening to hear the discussions in Friday lab meetings about

coordinating thousands of samples from across the world! Thank you for your oversight, kindness and support. Lastly, I would like to mention Clive, who has been a rock of statistical advice and guidance, always impartial and committed to the facts! Thank you for your collaboration and patient explanation of statistics. I am excited to have grown in my understanding of and confidence in data analysis and interpretation during this degree.

I would also like to thank all those who taught me laboratory skills, especially Hania Khoury, Rachel Lai, Luiz Carvalho and Shea Hamilton; all the other sources of wisdom and encouragement that I encountered at Imperial, especially Victoria Wright, Fadil Bidmos and Carolina Herrera; and those that have provided valuable friendship, including multiple members of the Levin group at the St Mary's Campus, especially Rahmeen Rahman and Nathalie MacDermott, and colleagues at the Leicester Royal Infirmary, notably Andrew Rosser and the late Navin Venkatraman who is sorely missed. Thank you to my father, who, on top of everything else, jumped into the foray at the last minute to help me with the formatting. And last, but the furthest from least: my wife Tamsin, who has been an incredible source of encouragement, love and support over the last four years. She is the informal collaborator, in the true sense of the word. Thank you for everything you have done!

The ILULU Consortium

Institute of Infectious Diseases and Molecular Medicine, University of Cape Town Nonzwakazi Bangani, Lizl Bashe, Melina Carr, Hannah P. Gideon, Rene Goliath, Yekiwe Hlombe, Vanessa January, Bekekile Kwaza, Suzaan Marais, Marc Mendelson, Tolu Oni, Fadheela Patel, Ronnett Seldon, Relebohile Tsekela, Katalin A. Wilkinson, Robert J. Wilkinson, Kathryn Wood; *London School of Hygiene & Tropical Medicine/Karonga Prevention Study* Lyn Ambrose, Amelia C. Crampin, Hazel M. Dockrell, Neil French, Lumbani Munthali, Bagrey Ngwira, Amos Phiri, Femia Zgambo; *Red Cross War Memorial Children's Hospital, University of Cape Town* Margaret Cooper, Brian Eley, Mabel Gcuwa, Spasina King, Glynis Kossew, Karen McCabe, Wonita Petersen, Sandra Pienaar, Vashini Pillay; *Liverpool School of Tropical Medicine/Malawi-Liverpool-Wellcome Trust Clinical Research Programme, University of Malawi College of Medicine* Benjamin Allubha, George Chagaluka, Angeziwa Chunga, Janet Dube, Robert S. Heyderman, Annie Joabe, Martha Kalemba, Anne Kerr, Monica Matola, Rachel Mlotha, Agnes Mwale, David Mzinza; *Brighton and Sussex Medical School, University of Sussex* Suzanne T. Anderson, Gillian Baker, Claire M. Banwell, Terry Bishop, Natalie Chaplin, Julian Golland, Florian Kern, Susan Poore, Jayne Wellington; *Imperial College London* Andrew J. Brent, Lachlan J. Coin, Hariklia Eleftherohorinou, Shea Hamilton, Myrsini Kaforou, Paul R. Langford, Michael Levin, Stephanie Menikou, Victoria J. Wright.

Abstract

Rapid, accurate, affordable, user-friendly and robust diagnostic tests are needed for use in sub-Saharan Africa, where the incidence and case fatality rates of TB are highest. Pathogen detection is such settings is challenging, owing to low abundance in sputum, and the associated need for laboratory-based assays. An alternative to pathogen detection are host-derived biomarkers, from which accurate diagnostic biosignatures can been derived. Of these, host serum proteins are a favourable option owing to their large diversity and abundance, range of clinical sources, response to pathological processes and amenability to immunocapture, including by multiplexed lateral flow immunoassay.

In this thesis, I evaluated two approaches to discovering immunocapture-based serum protein diagnostic signatures for TB in cohorts recruited from Malawi and South Africa, primarily from secondary care. The first is to re-test 22 proteins from which a 7-protein signature was recently discovered in a cohort recruited from primary care in Africa. This signature showed potential for use as a screening test in this setting. In our study, a 9-protein signature is discovered that may perform in a similar manner in secondary care. We also studied the levels of a subset of these 22 proteins in a smaller paediatric cohort before and after TB treatment. The second approach was to evaluate the translatability of serum protein biomarkers that were discovered using a mass spectrometry-based platform in our cohort to immunocapture. Results on three such proteins are presented, each with differing but informative outcomes.

Patients recruited to our studies were characterised with regards to demographic and clinical variables. In the final part of the thesis, I explore associations between serum protein levels in TB and these variables in order to gain further insight into the biology of TB. Novel associations were discovered with age, body mass index, smear grade and CD4+ count.

CONTENTS

Declarations 2

Acknowledgements 3

Abstract 5

LIST OF FIGURES 9

LIST OF TABLES 12

CHAPTER I: INTRODUCTION 16

1. Background 16

i) History 16

ii) The Current Challenge 16

2. The Limitations of Current Diagnostic Tests in sub-Saharan Africa 17

i) Microbiological culture 18

ii) Smear microscopy 18

iii) Chest x-ray 18

iv) GeneXpert MTB/RIF 19

v) Loop-mediated isothermal amplification platform 21

vi) Urine lipoarabinomannan 21

vii) Interferon-gamma release assay 22

3. A role for host-derived sputum-free serum protein biosignature-based tests 23

i) "Omics"-based biomarker discovery 23

ii) Variable selection methods for biosignatures 23

iii) Host serum proteins as infectious disease biomarkers 24

iv) Host serum proteins as POCT biomarkers 24

v) Proteomic biomarker discovery (an overview) 25

vi) Surface-enhanced laser desorption ionisation time-of-flight mass spectrometry 27

vii) Important considerations for study design for a host protein-based diagnostic signature of TB for use in sub-Saharan Africa 28

viii) previous studies that have sought host serum protein diagnostic signatures for TB 30

4. The EU-TB study 35

i) Clinical recruitment 35

ii) Proteomic workflow 39

iii) Forward selection partial least squares 39

iv) EU-TB SELDI diagnostic signatures 39

5. Clinical variables and the Biology of TB 40

6. Hypothesis 41

7. Primary Aims and Objectives 42

8. Secondary Aim and Objectives 42

9. Longitudinal cohort 42

CHAPTER II: MATERIALS AND METHODS 45

1. Introduction 45

2. Materials 45
i) Sample biobanks 45
ii) Equipment and consumables used for Luminex assays 46
iii) Additional equipment and consumables used for ELISA 46
iv) Luminex panels/kits: 46
v) ELISA kits: 47
vi) Software: 47
3. Methods 47
i) Luminex assays of the 22 proteins in the AE-TBC study 47
ii) ELISA of the three proteins in EU-TB diagnostic signatures derived from SELDI-TOF mass spectrometry 49
iii) Statistical analyses 49
4. Analysis of serum protein levels before and after treatment in a paediatric cohort 52
5. Analyses of markers for biologically meaningful associations with clinical variables 53

CHAPTER III RESULTS: Re-test of 22 protein biomarkers from the AE-TBC in the EU-TB cohort 55
1. Introduction 55
2. Individual serum protein levels and ROC AUCs (TB vs. OD) 59
3. ROC AUCs of individual proteins to distinguish TB from LTBI 89
4. Results of analyses of biosignatures to distinguish TB from OD 90
i) Application of the AE-TBC 7-protein signature to the EU-TB cohort 91
ii) *Discovery of the best new signature in the EU-TB cohort;* 94
5. Discussion 99

CHAPTER IV RESULTS: Levels of AE-TBC proteins before and after treatment in a paediatric TB cohort 103
1. Background 103
2. Results 105
3. Discussion 111

CHAPTER V RESULTS: Evaluation of performance of protein biomarkers from Surface-enhanced laser desorption ionisation time-of-flight mass spectrometry (SELDI) signatures by immunoassay 115
1. Introduction 115
2. Results 116
i) Transthyretin 116
ii) Beta-2-Microglobulin 120
iii) Alpha-1-antichymotrypsin 125
3. Discussion 130

CHAPTER VI RESULTS: host serum proteins and the biology of TB 134
1. Introduction 134
2. The TB disease state and the impact of HIV 136
i) The TB disease state 136
ii) The impact of HIV 138
3. Associations between the 22 AE-TBC biomarkers and clinical variables in TB 142
i) Smear grade 142
ii) Body Mass Index 150

iii) CD4+ count 156
iv) Age 163
v) Gender 166
4. **Discussion** 171
i) The TB disease state 171
ii) The impact of HIV 176
iii) Smear grade 178
iv) Body Mass index 181
v) CD4+ count 183
vi) Age 184
vii) Gender 185
5. Conclusion 186

CHAPTER VII GENERAL DISCUSSION 188
1. Host serum protein diagnostic biomarkers of TB 188
2. Response of host serum proteins to treatment. 191
3. Host serum proteins and the biology of TB. 192
4. Closing remarks 196

REFERENCES 197
Chapter I 197
Chapter II 202
Chapter III 202
Chapter IV 203
Chapter V 204
Chapter VI 205
Chapter VII 214

APPENDICES 217
1. Introduction 217
2. Correspondence 217

LIST OF FIGURES

Figure 1	**Recent and predicted trends in global TB incidence (from Dye et al, '13)[5].**	17
Figure 2	**Examples of multiplexed lateral flow immunoassay (LFA) formats.**	24
Figure 3	**Top-down vs. bottom-up proteomic approaches to protein identification.**	26
Figure 4	**An illustration of the workflow in surface-enhanced laser desorption ionisation mass spectrometry (SELDI-MS).**	27
Figure 5	**An ideal study design for discovering a diagnostic signature for TB in sub-Saharan Africa based on host serum proteins.**	29
Figure 6	**Classification criteria used for cohort study by the African-European TB Consortium (AE-TBC, from Chegou et al, '16)**	35
Figure 7	**The diagnostic process followed in phase 1 of the EU-TB study to identify TB, OD and LTBI cases (from Kaforou et al, '13)**	37
Figure 8	**Numbers of patients recruited to each clinical group of the phase 1 EU-TB cohort.**	38
Figure 9	**Flow diagram showing the process followed from patient recruitment and serum collection onwards for discovery of new bio-signatures in the EU-TB cohort.**	51
Figure 10	**Serum levels of procalcitonin in each of the six clinical groups in the EU-TB study measured by Luminex.**	60
Figure 11	**Serum levels of ferritin in each of the clinical groups in the EU-TB study measured by Luminex.**	61
Figure 12	**Serum levels of tissue plasminogen activator (tPA) in each of the clinical groups in the EU-TB study measured by Luminex.**	62
Figure 13	**Serum levels of fibrinogen in each of the clinical groups in the EU-TB study measured by Luminex.**	63
Figure 14	**Serum levels of serum amyloid A (SAA) in each of the clinical groups in the EU-TB study measured by Luminex.**	64
Figure 15	**Serum levels of alpha-2-macroglobulin (alpha-2-M) in each of the clinical groups in the EU-TB study measured by Luminex.**	65
Figure 16	**Serum levels of serum amyloid P (SAP) in each of the clinical groups in the EU-TB study measured by Luminex.**	66
Figure 17	**Serum levels of haptoglobin in each of the clinical groups in the EU-TB study measured by Luminex.**	67
Figure 18	**Serum levels of C-reactive protein (CRP) in each of the clinical groups in the EU-TB study measured by Luminex.**	68
Figure 19	**Serum levels of matrix metalloproteinase-2 (MMP-2) in each of the clinical groups in the EU-TB study measured by Luminex.**	69
Figure 20	**Serum levels of matrix metalloproteinase-9 (MMP-9) in each of the clinical groups in the EU-TB study measured by Luminex.**	70
Figure 21	**Serum levels of apolipoprotein A-I (apo-AI) in each of the clinical groups in the EU-TB study measured by Luminex.**	71
Figure 22	**Serum levels of apolipoprotein C-III (apo-CIII) in each of the clinical groups in the EU-TB study measured by Luminex.**	72

Figure 23	**Serum levels of transthyretin (TTR) in each of the clinical groups in the EU-TB study measured by Luminex.**	73
Figure 24	**Serum levels of complement factor H (FH) in each of the clinical groups in the EU-TB study measured by Luminex.**	74
Figure 25	**Serum levels of interferon gamma (IFN-gamma) in each of the clinical groups in the EU-TB study measured by Luminex.**	75
Figure 26	**Serum levels of interferon gamma-inducible protein-10 (IP-10) in each of the clinical groups in the EU-TB study measured by Luminex.**	76
Figure 27	**Serum levels of interferon alpha-2 (IFN-alpha-2) in each of the clinical groups in the EU-TB study measured by Luminex.**	77
Figure 28	**Serum levels of vascular endothelial growth factor (VEGF) in each of the clinical groups in the EU-TB study measured by Luminex.**	78
Figure 29	**Serum levels of interleukin-1 receptor antagonist (IL-1 RA) in each of the clinical groups in the EU-TB study measured by Luminex.**	79
Figure 30	**Serum levels of transforming growth factor alpha (TGF-alpha) in each of the clinical groups in the EU-TB study measured by Luminex.**	80
Figure 31	**Serum levels of tumour necrosis factor alpha (TNF-alpha) in each of the clinical groups in the EU-TB study measured by Luminex.**	81
Figure 32	**Areas under the receiver operating characteristic curves (ROC AUCs) of each protein to distinguish TB from OD in the EU-TB cohort.**	82
Figure 33	**ROC curves for the accuracy of complement factor H after stratification by each of HIV status and site.**	83
Figure 34	**Individual ROC AUCs for each of HIV uninfected and infected halves of the cohort.**	84
Figure 35	**Individual ROC AUCs for each protein at each of the Karonga and Cape Town sites.**	85
Figure 36	**ROC AUCs of each protein to distinguish TB from OD in order of the size of the difference in performance between the two studies.**	88
Figure 37	**Levels of each of the 22 proteins from the AE-TBC study in the TB groups of the two cohorts (AE-TBC and EU-TB).**	92
Figure 38	**Plasma levels of complement factor H (FH) in a South African paediatric TB cohort (n = 39) before and after TB treatment, as measured by Luminex.**	105
Figure 39	**Plasma levels of apolipoprotein A-I (apo A-I) in a South African paediatric TB cohort (n = 39) before and after TB treatment, as measured by Luminex.**	106
Figure 40	**Plasma levels of apolipoprotein C-III (apo-CIII) in a South African paediatric TB cohort (n = 39) before and after TB treatment, as measured by Luminex.**	107
Figure 41	**Plasma levels of transthyretin in a South African paediatric TB cohort (n = 15) before and after TB treatment, as measured by Luminex.**	108
Figure 42	**Plasma levels of complement C3 in a South African paediatric TB cohort (n = 24) before and after TB treatment, as measured by Luminex.**	109
Figure 43	**Euler diagram showing numbers of patients with protein levels that failed to normalise after treatment, by protein.**	111
Figure 44	**Serum levels and diagnostic performance of transthyretin (TTR) as measured by ELISA (N = 167).**	117

Figure 45 Correlation with SELDI intensity and diagnostic performance of transthyretin as measured by Luminex (N = 437). 118
Figure 46 Serum levels of beta-2-microglobulin (beta-2-M) as measured by ELISA (N = 172). 121
Figure 47 Comparison of levels of beta-2-microglobulin (beta-2-M) in HIV uninfected and infected individuals. 122
Figure 48 Serum levels of beta-2-M as measured by SELDI intensity and stratified by HIV status. 123
Figure 49 Translation of the 4,357 m/z alpha-1-antichymotrypsin (ACT) fragment to ELISA to distinguish TB from LTBI/OD. 126
Figure 50 Translation of the 4,468 m/z alpha-1-antichymotrypsin (ACT) fragment to ELISA to distinguish TB from OD. 127
Figure 51 Box plots showing serum levels of alpha-1-antichymotrypsin (ACT) in each of the six clinical groups by ELISA and SELDI. 129
Figure 52 Distribution of smear grades in each of TB, HIV- and TB, HIV+. 143
Figure 53 Box plots of protein level by smear grade category for the three proteins that correlated significantly with smear grade as an ordinal variable, regardless of HIV status. 145
Figure 54 Serum MMP-2 levels in TB mono-infected patients (HIV -ve) with smear grades < 2+ vs. >/= 2+. 147
Figure 55 Levels of fibrinogen, MMP-9, apo-AI and TGF-alpha in HIV co-infected TB patients that were smear negative (-ve) vs. positive (+ve). 149
Figure 56 XY scatter plots of serum protein level vs. BMI in the TB patients (HIV- and HIV+ together). 153
Figure 57 Levels of proteins in TB, HIV+ patients by CD4+ count as a continuous variable. 158
Figure 58 Levels of tPA in TB, HIV+ patients with CD4+ counts either < or >/= 200 cells/uL. 160
Figure 59 Distribution of ages in patients with active TB. 163
Figure 60 Correlation of serum levels of Haptoglobin with age in patients with active TB. 165
Figure 61 Comparison of levels in female vs. male TB patients, regardless of HIV status or site. 167
Figure 62 Comparison of levels in female vs. male TB patients after stratification by HIV status. 169
Figure 63 Comparison of CD4+ count by gender in TB, HIV+ patients. 170

LIST OF TABLES

Table 1	**A review of all host serum proteomic studies to date that have sought diagnosis of active pulmonary TB in adults vs. symptomatic controls.**	32
Table 2	**Surface chemistry and protein binding affinities of the four array types used in the EU-TB project.**	39
Table 3	**The EU-TB host protein signatures with performances as derived from the SELDI data shown as ROC AUC.**	40
Table 4	**List of demographic and clinical variables in the EU-TB project for which information was gathered on participants.**	41
Table 5	**A description of which sample biobanks were used for which experiments in this thesis.**	45
Table 6	**A list of which proteins were assayed as part of each Luminex panel.**	47
Table 7	**A description of the serum dilution process that was followed for the Luminex assays on the EU-TB sera.**	48
Table 8	**Demographic and clinical features for the 438 participants randomly selected for this study from the EU-TB cohort.**	56
Table 9	**Major clinical diagnoses in the Other Diseases groups.**	57
Table 10	**Serum concentrations of each protein in each of the three main clinical groups.**	59
Table 11	**Diagnostic accuracy of protein biomarkers individually.**	87
Table 12	**ROC AUCs for each protein to distinguish active TB from LTBI.**	89
Table 13	**Numbers of patients with TB and OD that were included in the biosignature analyses, stratified by HIV status and site.**	90
Table 14	**Performance of the seven-protein signature from the AE-TBC study in the ILULU-TB cohort.**	91
Table 15	**Comparison of levels of apo-AI, complement FH, apo-CIII, transthyretin and procalcitonin in the LTBI group of the EU-TB cohort with normal ranges taken from the literature.**	93
Table 16	**Performance of a new five-protein signature derived from Generalised Discriminant Analyses (GDA).**	95
Table 17	**Performance of a new 22-protein signature derived from Random Forests analyses.**	96
Table 18	**Performance of a nine-protein signature derived from Parallel Regularised Regression Model Search.**	97
Table 19	**ROC AUCs of the nine-protein signature after stratifying for each of HIV status and site.**	98
Table 20	**Demographic and clinical features of the patients and healthy controls randomly selected for this study from the paediatric cohort of Dr Elizabeth Whittaker.**	104
Table 21	**Difference in levels of apo-AI, transthyretin and complement factor H in a South African paediatric TB cohort before and after TB treatment, as measured by Luminex.**	110
Table 22	**Areas under the ROC curves (ROC AUCs) for transthyretin for distinguishing TB from each of LTBI/OD and OD alone when measured by ELISA or Luminex.**	119
Table 23	**ROC AUCS for transthyretin when measured by ELISA for distinguishing TB from each of LTBI/OD and OD in each of HIV uninfected and infected patients.**	119
Table 24	**ROC AUCS for transthyretin when measured by ELISA for distinguishing TB from each of LTBI/OD and OD in patients from Karonga and Cape Town.**	119

Table 25	**Correlation of ELISA and SELDI data and diagnostic accuracy of beta-2-microglobulin (beta-2-M) levels by ELISA (ROC AUC) after stratifying the patients by HIV status.**	124
Table 26	**Comparison of interquartile ranges of beta-2-M levels as measured by ELISA between HIV uninfected and infected individuals.**	124
Table 27	**ROC AUCS for beta-2-microglobulin when measured by ELISA for distinguishing TB from LTBI/OD in patients from Karonga and Cape Town.**	125
Table 28	**Correlation of ELISA and SELDI data for each of the two alpha-1-antichymotrypsin fragments and corresponding ROC AUCs by each of ELISA and SELDI.**	128
Table 29	**A list of the clinical variables that were included for analysis of association with the AE-TBC biomarkers within the TB group.**	135
Table 30	**A comparison of serum protein concentrations in TB with those in LTBI, focusing on HIV uninfected patients.**	137
Table 31	**A comparison of serum protein concentrations in TB, HIV+ with those in LTBI, HIV+ patients.**	139
Table 32	**A comparison of serum protein concentrations in TB, HIV- with those in TB, HIV+ patients.**	141
Table 33	**Correlation of serum protein levels with smear grade in patients with active TB, regardless of HIV status (n = 150).**	144
Table 34	**Protein levels in TB mono-infected patients (HIV -ve) with smear grades <2+ and >/= 2+.**	146
Table 35	**Serum protein levels in smear -ve vs. smear +ve patient with active TB and HIV. Median and interquartile ranges**	148
Table 36	**Correlation of serum protein levels with body mass index (BMI) in the EU-TB active TB patients (n = 150).**	152
Table 37	**Numbers of TB patients with BMI < 18.5 and >/= 18.5 when stratified by HIV status, and in total.**	154
Table 38	**Comparison of protein levels between TB, HIV- patients with BMI < 18.5 vs. >/= 18.5.**	154
Table 39	**Comparison of protein levels between TB, HIV+ patients with BMI < 18.5 vs. >/= 18.5.**	155
Table 40	**Distribution of TB, HIV+ patients by CD4+ count category.**	156
Table 41	**Correlation of serum protein levels with CD4+ count as a continuous variable in the TB, HIV+ patients (n = 66).**	157
Table 42	**Protein levels in TB, HIV+ patients with CD4+ counts < 200 vs. >/= 200 cells/uL (n = 66).**	159
Table 43	**Comparison of biomarkers significantly associated with each of the clinical markers of severity (smear grade, BMI and CD4+ count).**	160
Table 44	**Numbers of TB, HIV+ patients who were smear -ve vs. +ve by each of CD4+ count < vs. >/= 200 cells/uL.**	161
Table 45	**Multivariate analysis of associations between smear grade and BMI with TGF-alpha level in all TB patients.**	161
Table 46	**Multivariate analysis of associations between smear grade and CD4+ count with apo-AI level in TB, HIV+ patients.**	162
Table 47	**Correlations of serum protein levels with age in patients with active TB.**	164
Table 48	**Comparison of levels between female and male TB patients, regardless of HIV status or site.**	166

Table 49 **Results of a multivariate linear regression analysis to assess independent association of multiple different variables with serum ferritin level.** 167
Table 50 **Results of a multivariate linear regression analysis to assess independent association of multiple different variables with serum ferritin level in HIV uninfected TB patients.** 168
Table 51 **Results of a multivariate linear regression analysis to assess independent association of multiple different variables with serum ferritin level in HIV co-infected TB patients.** 168
Table 52 **Results of a multivariate linear regression analysis to assess independent association of multiple different variables with serum ferritin level in HIV co-infected TB patients, with CD4+ count and ART status excluded.** 169
Table 53 **Numbers of TB, HIV+ patients on antiretroviral therapy (ART) by gender.** 170
Table 54 **The source(s), causative cytokines, biological functions and possible effects in TB of proteins whose serum concentrations are altered in TB.** 175
Table 55 **A comparison of serum protein biomarkers that were found to associate with measures of baseline sputum bacterial load or associated variables in the EU-TB and a published study.(94)** 180

CHAPTER I: INTRODUCTION

1. Background

i) History

In 1993, tuberculosis (TB) was declared a global emergency by the World Health Organization.[1] An unstable situation resulting from the manifold effects of poverty in sub-Saharan Africa together with a severe lack of funding for national TB control programmes had been further derailed by the emergence of HIV-1 (HIV). Over 95% of TB cases were in the developing world, however international travel and immigration to industrialized countries had led to rising TB incidences there as well.[1] A further issue was the rise of drug-resistant strains of *Mycobacterium tuberculosis* (Mtb), contributed to by the fall of the Soviet Union in 1991 and subsequent degradation of its healthcare systems.[2] A marker of the general disarray in global TB control was that, by 1992, more than one third of Mtb strains isolated in New York City were resistant to at least one first line drug, one fifth were multi-drug resistant (resistant to both rifampicin and isoniazid, the two most effective first line drugs) and over 100 cases had been identified in which isolates were resistant to all known drugs.[1]

The uniquely compelling arrival of the year 2000 prompted the largest assembly of world leaders in history in the form of The United Nations Millenium Summit. Also held in New York, this summit focussed on the improvement of quality of life and health in the poorest countries of the world. The sixth of eight Millenium Development Goals (MDGs) was to "combat HIV/AIDS, malaria, and other diseases", with targets including "reversing the incidence of malaria and other major diseases by 2015". The Stop TB Partnership was also then founded in 2001 as an international body comprised of over 1500 partners with representation from international organisations, NGOs, local governments, research and funding agencies, private companies and community groups. It aimed to accelerate progress in seven areas: expansion of the recently developed "DOTS (directly observed treatment, short course)" strategy, the aim of which was to decrease transmission by supervising treatment of existing cases with chemotherapy; drug-resistant TB; HIV-associated TB (HIV/TB); new TB drugs; new vaccines; new diagnostic tests and a global laboratory initiative.[3] The result of these concerted efforts was that, by 2013, the tide had indeed turned with incidence decreasing by around 2% per year, which has also been sustained in the last few years.[4]

ii) The Current Challenge

The remaining amount of work to be done, however, is immense. In the Global Plans to Stop TB released by the Stop TB Partnership (2006 and 2010), the primary goal was to eliminate TB by 2050, defined as an incidence of less than 1 case per million people. As depicted in Figure 1, to reach this goal, we need to increase the rate of reduction in incidence ten-fold. This equates to a 1000-fold decrease in incidence from the present day, bringing the global burden down to the level that it is in low incidence countries currently. As judged by the Stop TB partnership themselves, however, in order to achieve this we cannot simply rely on rollout of existing technologies as this rate of decline has never been achieved in any location at any time in history.[5] They state that we must continue to direct funding and research efforts towards the development of a "more potent armoury of diagnostics, drugs and vaccines". More immediate targets that have been set more recently by the WHO as part of their "End TB strategy" include a 90% reduction in incidence and a 95% reduction in mortality by 2035 compared to 2015.[6] They suggest an even more imminent need for "technological breakthroughs by 2025" in order to achieve these targets. One of the priority areas identified for research and development was better diagnostic tests, including new rapid, user-friendly, accurate and affordable systems that deliver results at point-of-care (POC).[6]

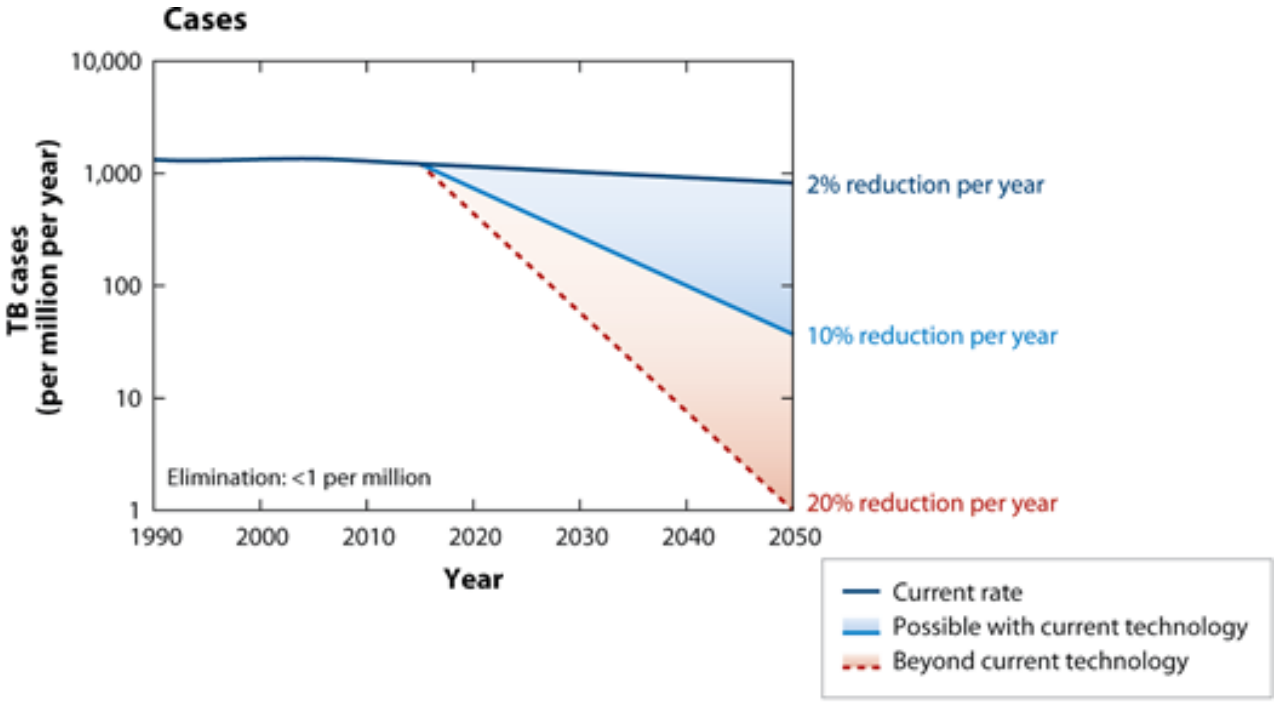


Figure 1 Recent and predicted trends in global TB incidence (from Dye et al, '13)[5].
Permission sought for re-use from journal (see Appendix).

2. The Limitations of Current Diagnostic Tests in sub-Saharan Africa

The WHO region with the highest incidence and case fatality rate is Africa where, in 2016, there were 2.5 million incident cases of which 27% died.[4] Within this, the subregion of sub-Saharan Africa (SSA) has the highest incidence rates and accounts for the majority of the burden of TB disease. SSA is also the poorest of the WHO world regions, with the highest gross domestic product of any African country still falling in the lowest World Bank category for gross domestic product.[7] In addition, the rate of "extreme poverty" in SSA is still 41% and, whilst that has decreased from 51% in 1981, owing to population expansion it is the only World Bank region where the absolute number of people in this bracket is increasing.[8] The region most in need of a cheap, user friendly POC test (POCT), and focus of this thesis, is therefore SSA. The WHO assert that such a test for TB for use in resource-constrained settings should ideally meet the "ASSURED" criteria, which are shown below in Box 1.

A - Affordable
S - Sensitive
S - Specific
U - User-friendly (simple to perform in a few steps with minimal training)
R - Rapid and robust (results available in less than 30 mins)
E - Equipment-free
D - Deliverable to end users

Box 1. Criteria for a point-of-care diagnostic test for use in resource-constrained settings: the "ASSURED criteria".
From Peeling *et al*, '06.[9]

An additional comment that should be made with regard to Africa as a whole is that the average incidence of co-infection with HIV is 29.5%, and over 50% in much of southern Africa.[4] Such a POCT should therefore ideally perform equally as well in co-infected patients, especially as the mortality rate in this population is almost three-fold higher at 36%.[4] The limitations of currently available diagnostic tests for TB will now be reviewed, with a focus on their suitability to use in SSA.

i) Microbiological culture

The gold standard for TB diagnosis historically has been culture of Mtb on solid, selective Lowenstein-Jensen medium. This results in the growth of characteristic colonies, but cultures that become positive take an average of 32 days to grow.[10] Use of liquid culture systems such as the BACTEC™ MGIT™ automated mycobacterial detection system (Becton Dickinson) shortens that delay to an average of 17.5 days. Liquid culture is also more sensitive than solid and is compatible with downstream species confirmation and drug susceptibility testing.[11] This delay is still far from ideal, however, and insufficient for interrupting transmission and impacting morbidity and mortality with the speed that is required. Further reduction in times to detection by culture are unlikely to be possible given that Mtb is a slow-growing mycobacterium, and co-infection with HIV leads to a lengthening of culture times.[12] Culture is also not always readily available in peripheral levels of healthcare systems in SSA, including district hospitals.[13] It does continue to function as the reference standard for new diagnostics research, however, and is likely to remain as the final confirmatory test in most settings.

ii) Smear microscopy

The rapid diagnostic test (RDT) that had been available prior to the plans and targets outlined above is bright field microscopy of Ziehl-Neelsen stained sputum smears (smear microscopy). Following the DOTS expansion, smear microscopy is in use in endemic settings, but the sensitivity of this is unreliable, ranging between 32 and 94%.[14] In addition, up to three separate samples are required in order to achieve those rates, meaning that repeated clinic visits are required. A proportion of patients default from this diagnostic process, with 15% of patients that had already given a sample generating a positive smear result being lost in one study in Malawi.[15] Smear microscopy is also a skilled process and requires frequent retraining of technicians in order to sustain quality assurance, possibly explaining the variability in performance.[16] Microscopes themselves also require maintenance, as do supplies of materials for smear preparation and staining, which may not always be continuous.[17] Furthermore, co-infection with HIV leads to a further decrease in diagnostic sensitivity to below 50% compared to sputum culutre.[18] Adaptation of the smear microscopy process by the addition of fluorescent light emitting diodes and use of auramine stain increases sensitivity by around 10%,[14] but the other limitations just described make this an unsatisfactory approach.

iii) Chest x-ray

The use of chest x-ray (CXR) as a standalone diagnostic test for TB was discouraged by the WHO on account of its poor specificity after the implementation of DOTS, which promoted microscopy as the test of choice.[19] Its place in diagnostic algorithms was therefore downstream of microscopy and as an aid to clinical diagnosis if microscopy was negative and a course of empiric antibiotic therapy did not resolve symptoms.[20] In combination with smear microscopy, however, it has been shown to perform well, with a combined sensitivity and specificity in one study in Kenya of 89% (95% confidence interval (CI) 86% – 91%) and 97% (84% – 90%) if CXR was performed first, or 93% (91% – 95%) and 62% (57% – 67%) if microscopy was performed first.[21] This may itself seem like a pragmatic solution, however the disadvantages of microscopy have already been described and radiography is also a skilled discipline, requiring specialised equipment and reliable power supplies. A two-test process even before culture is also not likely to be

particularly rapid and doubles the degree of inter-reader variability which is inherent with both techniques. CXR sensitivity was also lower in HIV co-infected patients in this study, in keeping with previous literature.[22]

iv) GeneXpert MTB/RIF

Of the new technologies that have recently emerged from the research and development pipeline and gained endorsement from the WHO, the GeneXpert® MTB/RIF assay (Xpert, Cepheid, Sunnyvale, CA) has undoubtedly been the most successful in terms of purchase and usage in endemic settings. By the end of 2016, 6,659 Xpert modules and 6.9 million test cartridges had been purchased in the public sector by 130 of the 145 countries deemed eligible for concessional pricing.[23] Xpert is a cartridge-based, sputum-based real-time PCR-based assay that detects Mtb DNA in less than two hours of laboratory time.[24] Diagnostic performance is highly appealing, with a sensitivity and specificity of 89% (95% CI: 85% - 92%) and 99% (98% - 99%) respectively in a meta-analysis for culture positive cases under operational conditions from testing of a single sputum sample.[25] Sensitivity smear negative, culture positive cases is 67%, providing a clear increase in detection compared to microscopy. In addition, sensitivity was not significantly lower in people living with HIV (PLWH) in this study at 79% (70% – 86%). This validates the findings of an early multi-national feasibility and accuracy study in 2011 which led the WHO to recommend it as a new first-line test in place of smear microscopy for patients either known to be or at high risk of being infected with HIV.[18, 26] Xpert can also simultaneously detect mutations conferring 95% of prevalent rifampicin resistance.[27] Rifampicin is one of the four first line anti-tuberculous therapy (ATT) drugs and is a cornerstone of the “short course” six month regimen. Resistance to rifampicin is also associated with resistance to isoniazid, and is thus a useful marker of multi-drug resistant (MDR-) TB especially in settings with high incidence of MDR-TB.[28] In 2017, a new cartridge for Xpert was released, the GeneXpert MTB/RIF “Ultra”, which has an even lower limit of detection at approximately 1 $\log_{10}$ below the original model and a run time of less than 80 minutes. This translated to even greater sensitivity in a recent clinical trial setting, particularly for smear negative, culture-confirmed cases and for HIV-associated TB.[29]

Whilst Xpert is clearly an extremely valuable addition to the diagnostic armoury, there are significant limitations with regards to its usage specifically as a POCT universally in high prevalence settings, including those with a high co-prevalence of HIV. Firstly, it is expensive. Even with compassionate pricing, the cost of purchase is $17,500 for a four test module with laptop and $9.98 per test cartridge.[30] Additional costs, however, include shipping (depends on destination), installation costs ($19,200), annual calibration overseas ($1,800), technician salary and fees for training and technical assistance. The grand total for purchase plus a year of usage, in fact, is estimated to be over $100,000 per module.[31] It may be affordable as the first line diagnostic test for all comers in some countries, for example South Africa, but the median budget per capita for health as a whole in low and middle income countries in SSA in 2016 was $78.[32] In addition, for TB funding globally it is estimated that there is already an annual funding shortfall of approximately $2 billion for low- and middle-income countries.[33] This would likely explain why countries such as Congo and Central African Republic had purchased less than 20 modules as of December 2016 (or Gabon and Congo less than 5,000 cartridges) despite having over 300 cases per 100,000 per year.[23]

Secondly, even for countries that can afford Xpert, it may not necessarily improve outcomes. Following the endorsement in 2011 and subsequent rollout in many African countries, two large studies were done. Both were in primary healthcare settings and are described below.

a) The TB-NEAT study compared the outcomes for patients in South Africa, Zambia, Zimbabwe and Tanzania who had been randomised to either smear microscopy or Xpert testing at POC of two spot sputum samples (n = 1,502).[34]

Fluorescence microscopy was employed at all sites except for in Tanzania were Ziehl-Neelsen staining was used. 60% of patients were co-infected with HIV, of which 26% were on antiretroviral therapy (ART) at enrolment. 2% of culture-positive cases were MDR-TB. By the end of day 1, more patients in the Xpert group that eventually became culture positive were on ATT than in the smear microscopy group (81% vs. 43%, $p < 0.0001$), but there was no overall impact on morbidity or mortality (morbidity was assessed by "TBscore", a composite of clinical features known to associate with severity) and Karnofsky Performance Score (KPS) at two and six months). There was more empiric prescribing in the microscopy group (n = 197 vs. 130), and by 9 days after randomisation, there were still significantly more culture positive patients in the Xpert arm on treatment, but only just ($p = 0.047$). This was almost certainly due to the availability of chest x-ray (CXR) at all sites, since 94% of empirically-treated patients had a CXR deemed compatible with active pulmonary TB (PTB). There were significantly more patients treated who did not become culture positive for TB (143/197 vs. 121/130, $p <0.0001$), however seemingly not enough to adversely impact morbitidy and mortality. CXR apparently improved sensitivity particularly for patients co-infected with HIV, since 87% of patients treated empirically in the microscopy group were HIV-infected. After a multivariate analysis, baseline TBscore ($p<0.0001$; OR 1.41, 95% CI 1.26 – 1.58) and HIV status ($p = 0.0095$, OR 2.01, 1.20-3.48) were significant predictors of death, but not sputum testing group ($p = 0.68$). No assessment was made of financial expenditure in this study, however the provision of an Xpert module at every primary healthcare centre would surely have been more expensive.

b) The XTEND study compared outcomes for patients at primary healthcare clinics in South Africa only who had been randomised to either fluorescence smear microscopy (two sputa) or Xpert (one sputum) performed at a hospital laboratory (n = 4,656).[35] Through partnership with the South African government, Xpert was rolled out in a step-wise fashion to facilitate a real world study. 62% of subjects were HIV infected (self-reported) of which 33% were on ART and median CD4+ counts were 310 cells/uL (IQR 183 – 470). 4% of patients with a positive Xpert test had detectable rifampicin resistance. Patients whose sputa tested negative by either initial test were subjected to a protocol including CXR if HIV infected or antibiotics followed by CXR if no improvement if HIV uninfected. As in the TB-NEAT study, there was no statistically significant difference in mortality between the two testing groups (91/2324 Xpert vs. 116/2332 microscopy, OR 0.86 (0.56 – 1.28)). There was no difference in the number of patients receiving ATT either, suggesting that empiric prescribing rates were again higher in the microscopy group and that Xpert replaced rather than supplemented the empirically treated culture positives. With regard to cost per patient, there was actually very little difference, with microscopy patients costing an average of $298.58 each (95% confidence intervals $246.35 – $350.82) and Xpert patients costing $312.58 ($252.46 - $372.70). The extra cost of the Xpert process itself was offset by the cost of the extra two sputum cultures per patient that had to be done in the microscopy group, together with the extra interim category II treatments that had to be given for the microscopy patients with a history of previous TB. Given that lack of increased health benefit, however, a subsequent analysis unsurprisingly revealed a less than 3% probability that Xpert was cost-effective in that population.[36]

In summary, the superior performance of Xpert was offset by very effective empiric prescribing by physicians with a high degree of experience of TB, in the context of adjunctive CXR use. This would seem to particularly oppose usage of Xpert at POC at primary healthcare level in SSA, which would require a much greater number of modules overall. With regard to hospital-based usage, there are no studies of impact or cost-effectiveness specific to this setting yet but availability of CXR is likely to be even greater than in primary care. Extrapolation of the above findings of "no benefit to morbitidy and mortality from Xpert" to hospital settings should be done with caution, however, since both TB and non-

TB patients are likely to be more unwell in secondary care. Studies of Xpert impact and relative cost in such settings are warranted.

A third limitation of Xpert is that it requires relatively well-developed infrastructure including a stable electricity supply, a maximum ambient temperature of 30 degrees centigrade, adequate storage space, protection from dust, security against theft and a bio-safety facility.[31] These issues may represent hidden costs, especially at the lowest resource sites where such features are often not already in existence, as was found in one study of Xpert installation in Nigerian district hospitals.[37] Fourthly, whilst training is relatively straightforward, dedicated personnel are needed to run the assays, representing a significant human resource allocation. Fifthly, because of the infectiousness of sputum in TB and the need for bio-safety, the test is best suited to laboratory usage. This adds further delay to the run time. In fact, one detailed study in a South African primary care centre of human resource requirements and total time required for running Xpert found that 2.5 people were needed to process 15 samples per day, and that the last time a test could be ordered for a same-day result was three hours before closing time.[38] In a hospital setting it could be envisaged that the journey time and time to sample processing might both be longer. The new Ultra cartridge may shorten run time to 80 minutes, however further shortening seems unlikely since significant amplification of the target material (Mtb DNA) is required.

One final and recent development of the Xpert platform is the Xpert "Omni", which is a small, single-cartridge module with a rechargeable battery that is better able to cope with heat and dust, designed to optimise desirability for use at POC. There are currently no studies evaluating its performance, however, and release of further data has been delayed by unforeseen technical problems, as reviewed by Walzl *et al*, '18.[39]

v) Loop-mediated isothermal amplification platform

Another recent test to have received WHO endorsement since 2000 is the loop-mediated isothermal amplification platform (LAMP) assay (Eiken Chemical Co., Japan). To some extent this is a cheaper and more user-friendly version of Xpert in that it is PCR and sputum-based but takes less than 2 hours and generates results that can be read with the naked eye.[40] Sensitivity and specificity were comparable to Xpert in one study, but rates of co-infection with HIV were low in this setting.[41] In addition, in a larger, multinational study, diagnostic accuracy was quite variable between sites and also between operators within each site ($p < 0.01$).[42] The authors point out that LAMP still requires a laboratory space as the sample has to be heated to 90^0C for five minutes to inactivate and lyse the mycobacteria. Uninterruptible power supplies were needed at all sites and it was felt that two separate work areas were also required to minimise DNA cross-contamination. Similar laboratory-related drawbacks to Xpert would be associated with this technology as well, making it less affordable and rapid.

vi) Urine lipoarabinomannan

Detection of the water-soluble Mtb cell wall lipopolysaccharide lipoarabinomannan (LAM) in urine has emerged in recent years for the diagnosis of active TB in PLWH in the form of the Alere Determine TB LAM (AlereLAM) test. 60uL of urine is applied to a lateral flow immunoassay strip which is then incubated at room temperature for 25 minutes. The strip is then inspected by eye and, if a band is seen, the intensity is then compared with that of four bands of increasing intensity printed on a reference card. Sensitivity of the assay, however, correlates with the degree of immunosuppression. In a recent Cochrane Review of studies in medical inpatients, in patients with CD4+ ≤ 100, pooled sensitivity was 56% (41% - 70%).[43] If CD4+ > 100, however, sensitivity fell to 26% (16% - 46%). Specificity in each of those two groups was 90% (81% - 95%) and 92% (78% - 97%), respectively. The WHO has now endorsed it as a "rule-in" test for TB, but specifically only in HIV co-infected patients with a CD4+ < 100, an unknown CD4+ count or severely disturbed vital signs.[44] The

biology behind this is that the presence of LAM in the urine in TB is likely to reflect renal seeding secondary to mycobacteraemia, rather than simply filtration of freely circulating LAM.[45-47] Further enhancement in sensitivity of the assay may close the gap slightly, and in fact a development in LAM detection technology by immunocapture in urine has recently been developed by Fujifilm (Tokyo, Japan) which does appear to do this: a study of biobanked urine from hospitalised patients with HIV found that the Fujifilm TB LAM test ("FujiLAM") had a sensitivity of 70·4% (95% CI 53·0 to 83·1) in relation to culture, as opposed to 42·3% (31·7 to 51·8) for the AlereLAM test.[48] The FujiLAM test device incorporates two buttons to release a reducing agent followed by a silver ion reagent for signal amplification, hence there is a chemically definable enhancement compared to the AlereLAM device, however these features may increase cost. A 40-minute sample incubation step is also required before loading the device, hence further studies are needed of feasibility, as well as robustness on storage in ambient conditions, user-friendliness and impact on clinical outcomes. There was a significant increase in sensitivity in patients with CD4+ count > 200 as well as < 100 in the recently published study, however it was lower in the > 200 group at 44·0% (29·7–58·5).[48] Generalisation to use in HIV uninfected patients (or "all-comers") with TB-like illness therefore remains unlikely, and indeed risk of Mtb mycobacteraemia is itself highly correlated with degree of immunosuppression and is extremely low in HIV uninfected patients.[49, 50]

vii) Interferon-gamma release assay

Interferon-gamma release assay (IGRA) of peripheral whole blood (Quantiferon Gold (QFT)) or peripheral blood mononuclear cells (T-SPOT.TB) were developed as tests for latent TB infection (LTBI) and as an upgrade to the tuberculin skin test (TST). Unlike the TST, patients do not have to return to the clinic at a specified time in order to receive a result, and a positive result is specific for Mtb whereas exposure to non-tuberculous mycobacteria can generate a positive TST. The IGRA does still require an overnight incubation step, however, and is thus fundamentally not suited to POC. Furthermore, multiple meta-analyses have shown that, especially in areas of high TB incidence, IGRA does not accurately distinguish active TB from clinically similar diseases. In one meta-analysis that focussed on studies performed in low and middle-income countries, pooled sensitivity was 76% for T-SPOT, 60% for QFT and pooled specificity 61% for T-SPOT and 52% for QFT.[51] The poor specificity was attributed to the inability of IGRA to distinguish active TB from LTBI, the latter being highly prevalent in such settings. Sensitivity in HIV-infected individuals was particularly poor at 68% for T-SPOT and 65% for QFT. As a result of these findings, the WHO have now advised against the use of IGRA to diagnose active TB in these settings.[52]

3. A role for host-derived sputum-free serum protein biosignature-based tests

A fundamental problem underlying the majority of the limitations in the techniques described above is the fact that Mtb bacilli are low abundance, slow growing, intracellular pathogens that primarily infect the deep tissues of the respiratory tract. Samples are therefore often difficult to obtain and require relatively advanced technology in order to generate the required sensitivity. As a result, and because sputum is also infectious material, such assays will be expensive and have a tendency to need to be based in laboratories, which are away from the patient. They will also require laboratory containment capability, which in turn necessitates infrastructure and dedicated staff, further increasing the expense. Attempting to meet the ASSURED criteria with a pathogen-based detection system would therefore be extremely challenging and may not be realised by the End TB or even Stop TB targets. Host-derived biomarkers, by contrast, are of higher abundance and more accessible, being present in many biofluids other than sputum. Consideration will therefore now be given to host serum protein biomarker discovery, with the following items discussed in turn:

i) "omics"-based biomarker discovery;
ii) variable selection methods for biosignatures;
iii) host serum proteins as infectious disease biomarkers;
iv) host serum proteins as POCT biomarkers;
v) proteomic biomarker discovery (an overview);
vi) surface-enhanced laser desorption ionisation time-of-flight mass spectrometry;
vii) important considerations for study design for host protein-based diagnostic signatures for TB for use in sub-Saharan Africa;
viii) previous studies that have sought host serum protein diagnostic signatures for TB.

i) "Omics"-based biomarker discovery

Following the completion of the human genome project and development of transcriptomic, proteomic and metabolomic technologies, there has been an increasing focus in medical research on analysis of these various "omes" and their association with different disease states. Such approaches can shed important new light on the biology of diseases and their response to treatment, but can also be a powerful tool in diagnostics research. Host-derived biomarkers by their levels individually are almost always non-specific to any one disease state, and so when a selection of markers that are all independently associated with the clinical situation or diagnosis of interest are discovered, the resulting combination or "signature" will be more discriminatory than any of the markers alone.

ii) Variable selection methods for biosignatures

Various statistical methods have been employed with the aim of generating the best diagnostic accuracy from an available pool of biomarker levels. Common themes among these are multiple regression (linear or logistic) and decision tree-based analyses. Linear and logistic regression involve a mathematical equation where the levels of each individual biomarker all contribute to varying degrees (and positively or negatively) to the probability of the diagnosis being the disease of interest. The conventional way of generating a test output is that a probability (p) > 0.5 constitutes a positive test result, with anything lower being a negative result. Decision tree analyses also use the levels of each marker to generate an output but do so sequentially, with each marker level representing the next branch. These methods are susceptible to over-fitting, but various techniques exist to mitigate this. One important example is the "random forest" method, where the average (the mode for categorical dependent variables) of the outputs from each of multiple different trees is taken as the test result. The process of finding the best possible signature for either of the regression or decision

tree methods is done using supervised machine learning. This is ideally performed using data from a discovery cohort of patients that can then be tested on a validation cohort to confirm reproducibility, but in the absence of two distinct cohorts, a single cohort can be divided into "training" and "test" sets.

iii) Host serum proteins as infectious disease biomarkers

The host serum proteome is an extremely diverse resource for biomarker discovery with over 10,000 different proteins identified to date.[53] Levels of serum proteins are also dynamic and representative of almost all different tissue types, supporting the notion that the serum proteome provides "a window into the state of one's organs in health and disease".[54] An early demonstration of the ability of the serum proteome to distinguish between different infectious diseases came from a study in Angola on the diagnosis of human african trypanosomiasis (HAT). Using surface-enhanced laser desorption ionisation time-of-flight mass spectrometry (SELDI-TOF MS; explained below), a signature derived from principal component analysis of levels of 206 different protein peaks was identified that could distinguish HAT (n = 85) from controls (n = 146) with a sensitivity and specificity of 100 and 98.6% respectively. The control group comprised patients with other protozoan, neurological or systemic infections (n = 113) and healthy controls (n = 33). 19 of the biomarkers had individual t-test p-values of $< 1 \times 10^{-5}$ when comparing levels between HAT and control groups. Use of just these 19 proteins still led to 96.3% of patients in the test set (n = 109) being correctly classified.[55]

iv) Host serum proteins as POCT biomarkers

Host proteins are also particularly appealing as constituents of a biosignature-based test since technology already exists for POC detection in the form of lateral flow immuno-assays (LFA). LFA would certainly facilitate an ASSURED criteria satisfying test, as evidenced by the widespread and effective usage in low resource settings of POCTs for *Plasmodium spp.* detection following the incorporation of these into WHO malaria guidelines over ten years ago.[56] Combined HIV antigen and antibody LFA tests are also WHO endorsed and widely available also. The urine LAM test (mentioned above) is an LFA for TB, and with regard to host proteins, the home pregnancy urine test is a widely established example. LFAs can also be multiplexed, allowing a combination of biomarkers to be assayed. Possible formats for this are illustrated in Figure 2.

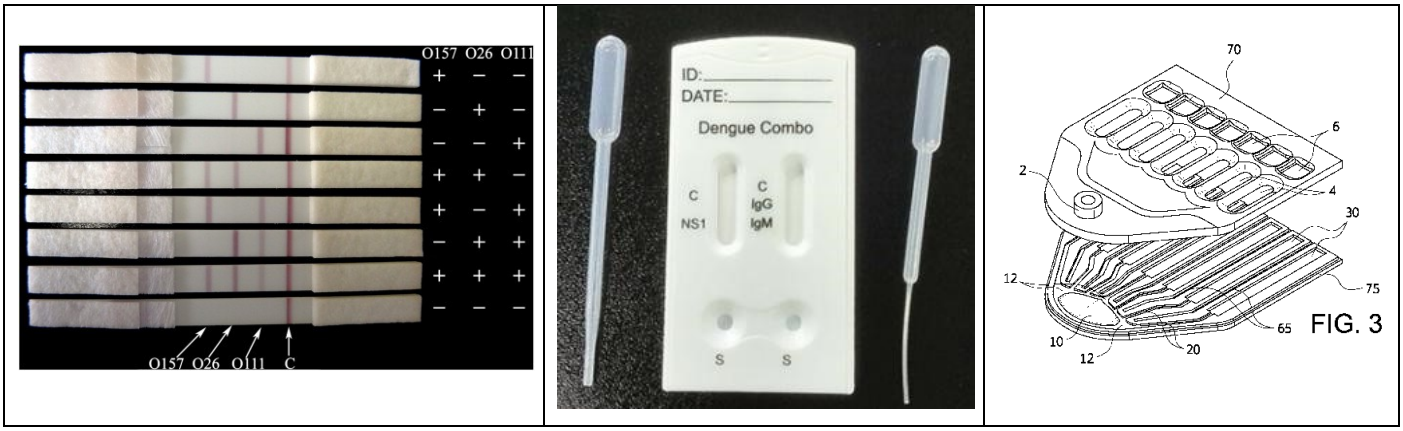


Figure 2 Examples of multiplexed lateral flow immunoassay (LFA) formats.

Left: detection of three serogroups of Shiga toxin-producing *Escherichia coli* by multiplexed LFA in series.[57] Middle: detection in parallel and series of Dengue nonstructural protein 1 (NS1) in one lane and anti-Dengue IgG and IgM in another.[58] Right: a putative design for multiplexed detection in parallel following diffusion from a single well.[59]

Finally, LFA can also be quantitative, especially when utilising highly sensitive reporting mechanisms such as up-converting phosphor (UCP) fluorescence.[60] In a study of levels of the chemokines interferon-gamma inducible protein 10 (IP-10) and C-C motif chemokine 4 (CCL4) in stimulated whole blood supernatants from patients with TB-like illnesses across six sites in SSA, UCP-LFA kits were found to be accurate between 100 and 100,000 pg/ml. Correlation of measurements of levels in patient samples by UCP-LFA with levels by ELISA was very strong (Pearson rho > 0.9). UCP-LFA assays were done independently at each of the six sites, and inter-site variation of measurement of different quantities of quality controls was also very low (Pearson > 0.98). Distribution of reagents to these sites was done without use of a cold chain, thereby also demonstrating robustness and applicability of UCP-LFA to low resource settings.[60]

v) Proteomic biomarker discovery (an overview)

Any biomarker discovery method can be targeted or untargeted. In targeted approaches, the identities of the markers to be quantified are pre-specified. The advantages of this are primarily that the often lengthy process of protein identification is circumvented, but also that markers can be selected for which POC-applicable detection technologies (e.g. immunocapture) are already commercially available. Truly novel and comprehensive biomarker discovery, however, requires untargeted methods that survey all or a vast majority of the relevant proteome.

Untargeted protein biomarker discovery processes have two main components: variable selection (explained above) and protein identification. The order in which these two processes are done depends on the type of proteomic approach that is taken. These can be broadly divided into “bottom-up” and “top-down” approaches. As illustrated in Figure 3, bottom-up involves breaking the proteins into smaller constituent parts (usually with trypsin digestion to form peptides), which are then accurately sequenced using mass spectrometry. This is also known as “shotgun” proteomics. The identities and quantities of the original proteins are then derived using proteomic database matching, at which point variable selection can then be performed. In top-down proteomics, the intact proteins are quantified and selected, before full identication has been done. A small number of target proteins are then fragmented or digested (and peptide sequenced) in order to obtain further identification (Figure 3).

From a pure preparation/isolate of a single protein, bottom-up identification is the more statistically powerful method as it allows no preconceptions to influence the final reported identify. When starting with a complex sample such as serum, however, pure bottom-up approaches become limited by the number of possible parent protein identities to which any individual peptide could be allocated. In addition, the higher number of starting peptides leads to decreased peptide recovery, leading in turn to decreased peptide coverage (and therefore certainty of identification) of any purported parent protein. The benefits of top-down are that it is avoids the time-consuming trypsin digestion step and retains post-translational modifications. In practice, protein biomarker discovery can employ either one of these approaches, or use them in combination. After either approach, final confirmation of protein identity is often gained using another method, e.g. immunodepletion from the starting sample with antibody-linked to protein A or G-coupled beads.

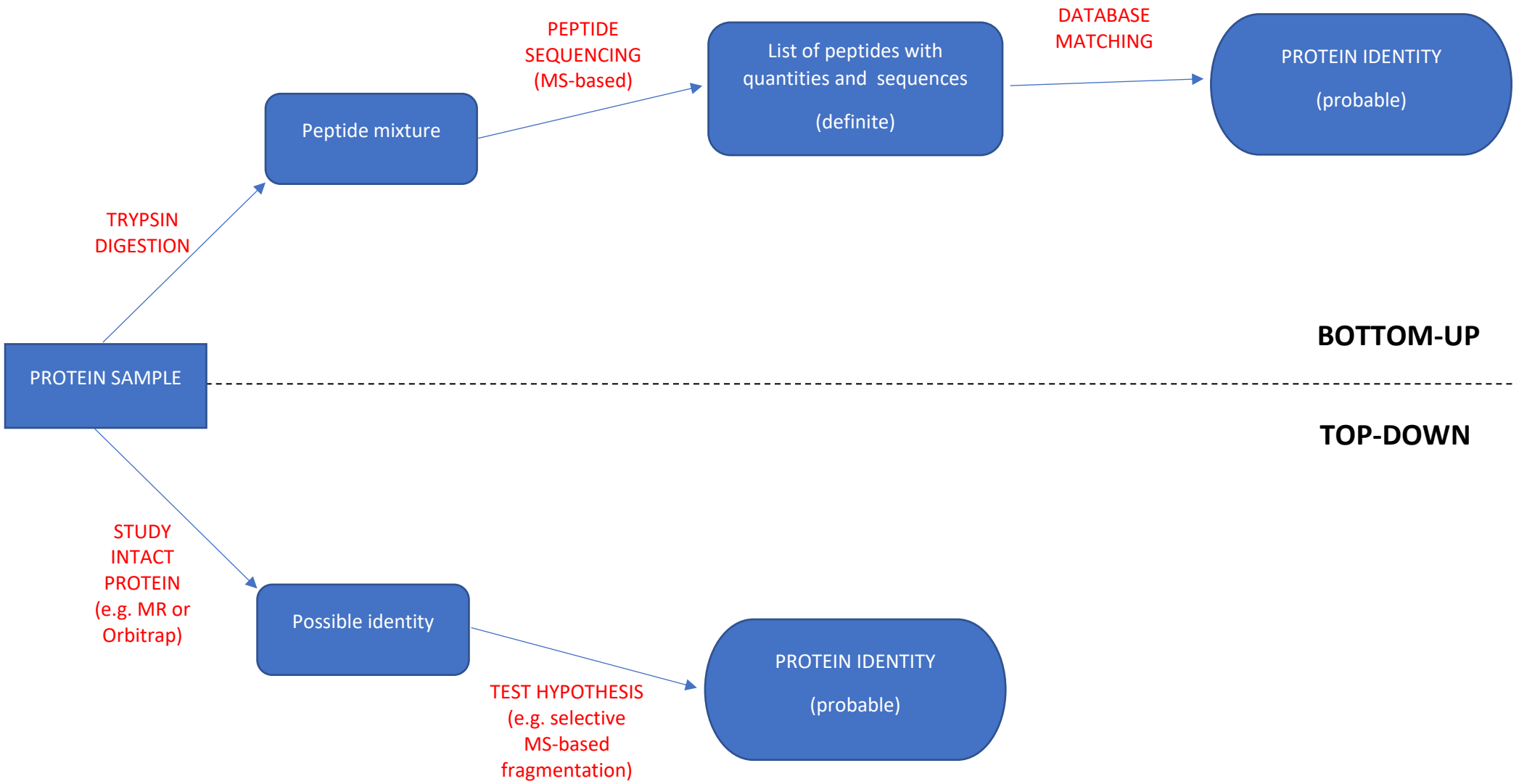


Figure 3 **Top-down vs. bottom-up proteomic approaches to protein identification.**

In "bottom-up" proteomics (top half of diagram), a protein is digested into constituent peptides whose amino acid sequences are then determined by mass spectometry (MS) with fragmentation. Protein identity is then inferred from the combination and abundance of sequenced peptides. In "top-down" approaches, information is gained about an intact protein, following which an hypothesis as to the identity can be tested with further experiments (e.g. specific MS-based fragmentation).

vi) Surface-enhanced laser desorption ionisation time-of-flight mass spectrometry

One technology that has been frequently used for proteomic biomarker discovery from clinical samples is surface-enhanced laser desorption ionisation mass spectrometry (SELDI-TOF MS, or "SELDI"). Originally developed by Ciphergen, this technique built on the existing technology of matrix-associated laser desorption ionisation (MALDI)-MS by upgrading the chemistry of the protein binding surface to have affinity for whole proteins.[61] As displayed in Figure 4, the protein binding surface is arranged as a series of small chromatographic spots such that 96 spots (in 12 strips or "chips" with eight spots each) can be assayed from one plate (or "array"). Protein chips are available with a variety of different surface chemistries, each with affinity for different categories of protein: hydrophobic, hydrophilic, anionic or cationic. This increases the overall sensitivity of the approach, but also the reliability of quantification, since a more specific binding surface leads to a lower number of proteins bound overall, thereby decreasing competition for ionisation, which is a major source of variability in MS (the "matrix effect"). A large number of patient samples can therefore be analysed in each run (e.g. 92 after addition of protein and peptide standards and controls) and on multiple different chromatographic surfaces. This makes SELDI a high throughput, highly sensitive and highly accurate means of analysing any given proteome in an untargeted manner.

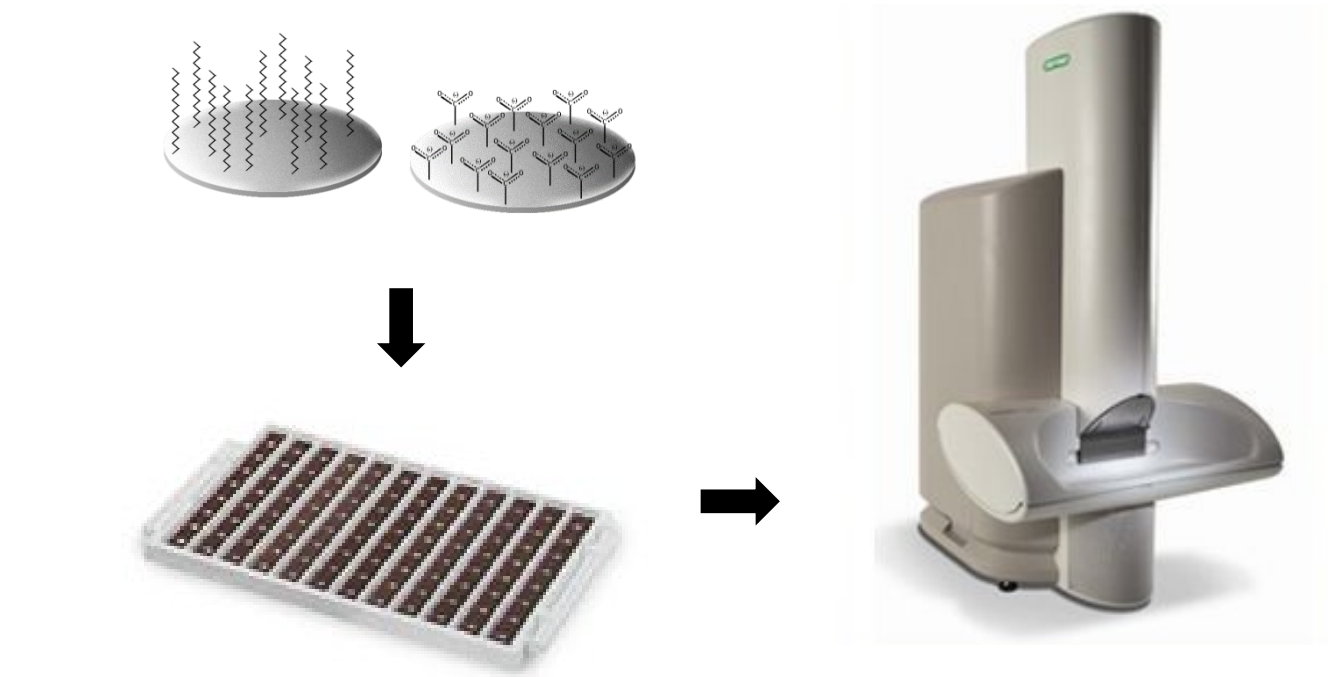


Figure 4 An illustration of the workflow in surface-enhanced laser desorption ionisation mass spectrometry (SELDI-MS).

Chips of chromatographic protein-binding spots (top left, showing reversed phase or cation-exchange coated spots) are loaded into arrays of 96 spots (eight spots per "chip", bottom left). Spots are chemically primed before a fixed amout of serum sample is added (not shown). After drying, an ionisation enhancing buffer is added to each spot ("matrix") and arrays are loaded into the SELDI mass spectrometer for analysis (right).

vii) Important considerations for study design for a host protein-based diagnostic signature of TB for use in sub-Saharan Africa

As explained above, the ideal POC diagnostic test for use in SSA would meet the ASSURED criteria, with the added requirement of doing so regardless of HIV status. Because of the variation in prevalence of other infectious and inflammatory conditions across the region, however, a further consideration is reproducibility across different geographical sites. For example, west, central and east African countries have high rates of malaria, whereas the south (including the majority of South Africa) is malaria-free.[62] Next, for a test to be truly helpful to a diagnostician, samples need to be taken at a point of "differential diagnosis". This is when the initial clinical features (i.e. history and examination but not laboratory results) are suggestive of TB, but other diagnoses are also possible. At this point the clinician intends to test for TB, and it is at this exact moment that a POCT would be applied. Healthy controls are informative as a separate comparison, but biomarker selection needs to be restricted to the cases (i.e. TB) and the sick controls (i.e. other diseases (ODs) that presented in a manner suggestive of TB). Untargeted proteomics is the more comprehensive strategy, but some test of translatability to immunocapture is also desirable. An ideal workflow is therefore illustrated in Figure 5, and a critical appraisal of the literature will then be performed against this standard.

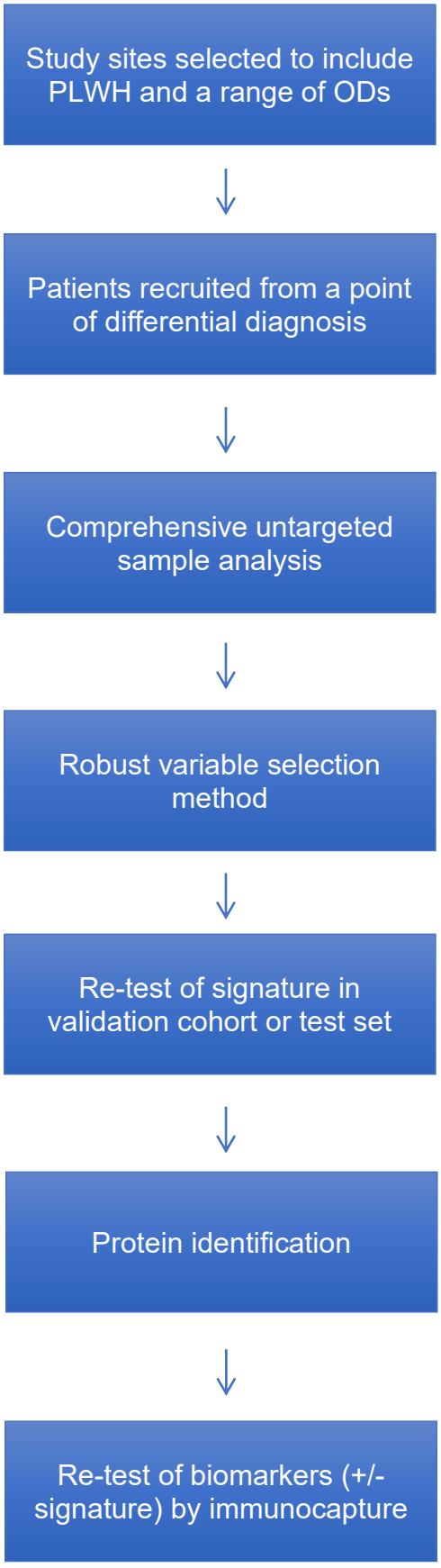


Figure 5 An ideal study design for discovering a diagnostic signature for TB in sub-Saharan Africa based on host serum proteins.

PLWH: people living with HIV; ODs: other diseases that presented in a manner suggestive of TB.

viii) previous studies that have sought host serum protein diagnostic signatures for TB

A summary of the studies to date that have had similar aims to the main aim of this thesis are shown in Table 1. In each case, information is given on each of the above criteria (in text and Figure 5), as well as the number of patients recruited, countries of recruitment, the proteomic technology that was used and the diagnostic accuracy.

Study	N	Controls	Recruitment sites	Included PLWH?	Technology	Targeted or untargeted	No. proteins identified	Re-tested by immuno-capture?	Accuracy	From a test set or validation cohort?
Agranoff, '06	249	Selection of 149 ODs and 21 HCs	Uganda; The Gambia; Angola; UK (London)	Y	SELDI	Untargeted	2	Yes (2)	All 219 protein peaks: Sens 94%, spec 95% Top 3-protein signature: Sens 82%, spec 86%	Y
Liu JY, '11	149	Selection of 36 ORDs and 32 HCs	China	N	SELDI	Untargeted	0	-	Sens 80% Spec 84%	Y
Zhang, '12	264	Selection of 69 ORDs and 66 HCs	China	N	SELDI	Untargeted	1	No	Sens 97% Spec 98%	N
Sandhu, '12	266	110 ORDs with symptoms suggestive of TB	Peru	N	SELDI	Untargeted	0	-	Sens 84% Spec 90%	Y
Liu J, '13	391	Selection of 120 ORDs and 91 HCs	China	N	SELDI	Untargeted	1	No	Sens 80% Spec 84%	Y
Song, '14	57	Selection of 30 ODs and 1 HC	S. Korea	N	Nano-UPLC-MS/MS	Untargeted	>500	No	ROC AUC 0.95 (top marker)	N
Xu, '15	160*	Selection of 80 ORDs and 40 HCs	China	N	iTRAQ-coupled 2D LC-MS/MS	Untargeted	434	Yes (3)	Vs. pneumonia: Sens 90%; spec 88 Vs. Lung ca: Sens 85%; spec 93%	N
Achkar, '15	215	ORDs with symptoms suggestive of TB	USA	Y	LC-MS/MS for discovery, MRM-MS for verification	Untargeted	18	No	ROC AUC: 0.96 for HIV(-) TB 0.95 for HIV(+) TB	Y

Study	N	Controls	Recruitment sites	Included PLWH?	Technology	Targeted or untargeted	No. proteins identified	Re-tested by immuno-capture?	Accuracy	From a test set or validation cohort?
Chegou, '16	716	Protocol-based (cough >/= 2 weeks)	Uganda Malawi South Africa Namibia The Gambia	Y	Luminex	Targeted	22	Yes (22)	Train set (70%): Sens 87% Spec 85% Test set (30%): Sens 81% Spec 80%	Y

Table 1 A review of all host serum proteomic studies to date that have sought diagnosis of active pulmonary TB in adults vs. symptomatic controls.

Characteristics that are suggested as being suboptimal from the point of view of this thesis specifically are in red font. *In the study by Xu *et al*, '15, sera from every 10 patients in each clinical group were pooled.[69] OD: other disease; ORD: other respiratory disease; HC: healthy control; UPLC: ultra-high pressure liquid chromatography; iTRAQ: isobaric tagging for relative and absolute quantification; MRM: multiple reaction monitoring; ROC AUC: area under the receiver operating characteristic curve.

The first study to apply an untargeted proteomics approach to finding a serum protein diagnostic signtuare for TB was by Agranoff *et al*, '06.[63] In this study, two cohorts were recruited. To the first, patients with culture-confirmed pulmonary TB, with and without HIV co-infection, were recruited retrospectively in Uganda and The Gambia (from sample banks) and prospectively from St George's Hospital, London, UK (n = 179). Controls were recruited from the same sites and in the same fashion, as well as prospectively from Angola (n = 170). The control group comprised a range of inflammatory conditions with clinical features that "can overlap with" those of TB (e.g. sarcoidosis, inflammatory bowel disease and autoimmune conditions). 21 healthy controls were also included. Sera were loaded onto weak cationic chromatographic chips (CM10) and analysed by SELDI. A support vector machine (SVM) algorithm-based approach was then used which generated very high sensitivity and specificity in the test set using all 219 protein peaks (Table 1). Limitations of this study, however, were firstly that control group comprised a selection of ODs rather than being differential diagnosis-based, secondly that 12% of the control group were healthy which probably falsely boosted the results, thirdly that the prevalence of HIV in the OD group was very low (3%) and fourthly that the two processes of protein identification and re-assay by immunocapture were only performed on two of the proteins (transthyretin and serum amyloid A). The performance of these two markers was then only assessed in combination with two other proteins taken from outside the study (neopterin and CRP).[64]

The next two studies listed in Table 1 were both limited by the fact that they recruited a selection of OD cases and did not include PLWH. Furthermore, Liu *et al*, '11, did not identify any of their biomarkers and Zhang *et al*, '12, identified one (orosomucoid) but didn't test their signature in a test set or validation cohort.[64, 65]

Sandhu *et al*, '12, undertook a well-designed study in which 266 adult patients presenting to a community clinic in Peru with symptoms suggestive of pulmonary TB were prospectively recruited.[66] The focus of the study was on discriminating active TB from LTBI given that the prevalence of positive IGRAs was high in that setting, as it is in SSA, however they did also look at discrimination of active TB from ORD including with use of a test set. The generalisability to SSA, however, is limited by the fact that no patients with HIV-coinfection were included and that none of the proteins were identified or re-tested by immunocapture.

The limitations of the study by Liu J *et al*, '13, are that the control group was a selection of ODs, included healthy controls, and whilst one protein was identified (a fibrinogen fragment), the performance even of this one individual biomarker was not re-tested by immunocapture.[67] Song *et al*, '14, used a shotgun proteomic method and didn't seek a signature but did identify a biomarker (alpha-1-antitrypsin) with an individual area under the receiver operating characteristic curve (ROC AUC) of 0.95.[68] This was not validated in a test set or validation cohort, however, and so together with the low numbers and poor recruitment design is unlikely to be clinically valuable.

Xu *et al*, '15, generated very impressive results using the accurate relative quantitation technology of "isobaric tagging for relative and absolute quantification (iTRAQ)".[69] A major downside of iTRAQ, however, is that it can only compare quantities of peptides between a maximum of 8 samples at a time, and so sera in this study had to be pooled from ten patients at a time within each clinical group. On top of the suboptimal study design and lack of validation in a test set, this aspect of their study means that the findings would likely be far less reliable and generalisable in terms of diagnostic accuracy at the single patient level.

The study by Achkar *et al*, '15, had a slightly unusual study design in that there were three phases of biomarker testing (discovery, verification and "classification"), but the discovery cohort was very small in number (n = 39), did not include

any unwell controls and comprised only HIV negative individuals.[70] The proteomic approach was bottom-up for the discovery phase, namely trypsin digestion followed by nano-UPLC coupled to tandem mass spectrometry for the purposes of identification and quantitation by fragmentation. A multiple reaction monitoring (MRM) platform was then set up on a separate mass spectrometer for more accurate quantification of biomarkers in the following two phases (a targeted approach). This involved the use of 244 synthetic peptides. A total of 87 proteins were taken forward for testing in the following cohorts, of which 18 had been selected from other studies in the literature which had included HIV co-infected patients at discovery level. In the verification and classification phases, HIV infected and uninfected patients were analysed separately. For each of HIV uninfected and infected patients, two signatures containing 10 and eight proteins respectively were derived. The protein signatures had very high combined ROC AUC values in the classification set, as shown in Table 1. Despite the intense MS-based workflow in this study, however, there are significant limitations. Firstly, the protein identities were not confirmed by immunodepletion or non-mass spectometry techniques, and so to some extent remain incompletely confirmed. Secondly, their performance was not re-tested by immunocapture, so it is difficult to know if they would all actually be translatable. Thirdly, having two separate signatures to take forwards for POCT testing is cumbersome, and a single POCT capable of detecting a total of 16 proteins (two being common to both signatures) would also be challenging to develop. The two proteins that were common to both signatures, however (CD14 and SEPP1), would be interesting to test in other cohorts.[70]

The final study that was published before the commencement of this thesis was by the African-European TB Consortium (AE-TBC).[71] In this large, multi-centre cohort study led by Stellenbosch University, a total of 716 adult patients that had presented to primary care health centres in SSA were recruited. Inclusion criteria were a cough of at least 2 weeks' duration plus any one of the following symptoms: fever, malaise, recent weight loss, night sweats, knowledge of close contact with a patient with TB, haemoptysis, chest pain or loss of appetite.[71] Patients were recruited across five countries: The Gambia, Namibia, South Africa, Malawi and Uganda. Patients were recruited regardless of HIV status and overall 24% were HIV infected. Patients were allocated to one of four categories: definite TB, probable TB, questionable TB and no TB based on clinical, radiological and microbiological criteria (Figure 6). Sera were collected and analysed centrally at Stellenbosch University using a Luminex platform (multiplexed immunoassay) to measure the levels of 22 pre-defined host proteins: interleukin 1 receptor antagonist (IL-1ra), transforming growth factor (TGF)-α, IFN-γ, IFN-γ-inducible protein (IP)-10, tumour necrosis factor (TNF)-α, IFN-α2, vascular endothelial growth factor (VEGF), matrix metalloproteinase (MMP)-2, MMP-9, apolipoprotein A-I (Apo A-I), Apo-CIII, transthyretin, complement factor H (FH) (Merck Millipore, Billerica, Massachusetts, USA), and C-reactive protein (CRP), serum amyloid A (SAA), serum amyloid P (SAP), fibrinogen, ferritin, tissue plasminogen activator (TPA), procalcitonin (PCT), haptoglobin and α-2-macroglobulin (A2M). These proteins had all previously been found to contribute to the diagnosis of TB in signatures from pilot studies. Patients were divided randomly into train and test sets (70:30), and two variable selection methods employed: Generalised Discriminant Analyses (GDA; a multivariate linear regression approach) and Random Forests. The top signature was a 7-protein test comprising CRP, transthyretin, IFN-γ, complement factor H, apolipoprotein A-I, IP-10 and serum amyloid A. In the test set, this signature had 81% sensitivity (95% CI: 69% - 90%) and 80% specificity (72% - 86%). With a view to implementing the signature as a community-based screening test, the cut-off for diagnosing TB was then lowered to increase sensitivity at the expense of specificity. In the test set, this produced a sensitivity of 94% (84% - 98%) and a specificity of 73% (65% - 80%). This translated into a positive and negative predictive value of 61% (50% - 70%) and 96% respectively (91% - 99%). The authors concluded that use of such a test in primary care situations would decrease the number of sputum-based investigations (Xpert or culture) by 75%.[71]

Classification	Definition
Definite TB	Sputum culture-positive for MTB. OR 2 positive smears and symptoms responding to TB treatment. OR 1 positive smear plus CXR suggestive of PTB.
Probable TB	1 positive smear and symptoms responding to TB treatment. OR CXR evidence and symptoms responding to TB treatment.
Questionable	Positive smear(s), but no other supporting evidence. OR CXR suggestive of PTB, but no other supporting evidence. OR Treatment initiated by healthcare providers on clinical suspicion only. No other supporting evidence.
No-PTB	Negative cultures, negative smears, negative CXR and treatment never initiated by healthcare providers.

CXR, chest X-ray; MTB, *Mycobacterium tuberculosis;* No-PTB, non-'pulmonary TB'; TB, pulmonary TB.

Figure 6 Classification criteria used for cohort study by the African-European TB Consortium (AE-TBC, from Chegou et al, '16)

Permission granted to re-use by BMJ Journals (see Appendix) [71]

While the AE-TBC study used an entirely targeted method of protein biomarker discovery, out of those in the literature to date (Table 1), it is the most interesting from the point of view of this thesis. This is firstly because it is the only published host serum proteomic study to have sought a signature for use in SSA and irrespective of HIV status; and secondly the protein identities are known and were by nature of the study design proven to be amenable to immunocapture. It was, however, a dedicated primary care or "community"-based study, with "all comers" recruited such that culture-negative patients were included, and just those with "questionable TB" excluded. Confirmation of OD (or "no-PTB") diagnosis was not actively sought, however the authors commented that a large number comprised upper respiratory tract infections plus exacerbations of COPD or asthma.[71] Not all such patients would be as likely to present to hospital facilities, hence generalisability to these settings is not guaranteed. A POCT for use in hospitals in low resource settings, be it sub-Saharan Africa or other low- and middle-income areas would be equally valuable, whether as a standalone diagnostic test or a POCT screening test upstream of Xpert and/or mycobacterial culture.

4. The EU-TB study

i) Clinical recruitment

Between 2007 and 2011, a European Union (EU)-funded study led by Imperial College London recruited 674 adults in Karonga District, Malawi and Cape Town, South Africa who had presented with clinical features suggestive of active TB to a large case-control study (the EU-TB study). The purpose was to discover novel host biomarkers of TB, both from the blood transcriptome and the serum proteome, that could form the basis of a cheap, accurate and user-friendly rapid diagnostic POC test for TB. Findings from the transcriptomic arm of the study have already been published.[72]

At both sites, the entry criterion was intention of the local clinician to test for TB on the basis of the clinical presentation. In Karonga, patients with TB or other diseases (OD) were recruited at Karonga District Hospital. In Cape Town, TB patients were recruited from an outpatient clinic in Khayelitsha township (Khayelitsha site B), and OD patients from two hospitals: GF Jooste (Khayelitsha) and Groote Schuur (central Cape Town). In contrast to the AE-TBC study, recruitment was therefore almost entirely set in secondary care, with all of the OD patients being recruited from hospitals.

As described in Kaforou, *et al* (’13), all patients were thoroughly investigated for TB. Two sputum samples (expectorated spontaneously or induced) were collected from each patient, examined by smear microscopy and then cultured for Mtb using solid media (both sites) and liquid media (Cape Town). Patients also underwent CXR and HIV testing, with cultures of blood, CSF or urine performed where clinically indicated. Biopsies for histology were also performed where necessary with samples sent for TB culture. Patients were allocated to the TB group if Mtb was cultured from one or more sputum sample(s). Patients were allocated to the OD group if an established alternative diagnosis was made or if there was an observed improvement in symptoms without TB treatment. Follow-up was performed at 26 weeks to confirm correct disease categorisation. For healthy controls, healthy individuals with LTBI (TST and IGRA postive) were recruited from the same catchment areas as the TB and OD patients at each site. In Karonga, this was done by random community selection, and in Cape Town at antiretroviral clinics.[72]

This process was followed until sufficient and equal numbers of patients in each of the following clinical groups and at each site had been recruited (HIV+: HIV infected; HIV-: HIV uninfected):

a) TB, HIV-
b) TB, HIV+
c) LTBI, HIV-
d) LTBI, HIV+
e) OD, HIV-
f) OD, HIV+

A detailed representation of the recruitment process is shown in Figure 7. The total number of patients recruited to each group and at each site together with their demographic and clinical details is then shown in Figure 8. The total number of patients recruited to the study was 674.

This episode of recruitment was the first of two phases, with the second phase recruiting “all-comers” with suspected TB, including culture negative TB patients (clinically diagnosed). Only the first phase (“phase 1”) of the EU-TB study will be referred to hereon in in this thesis.

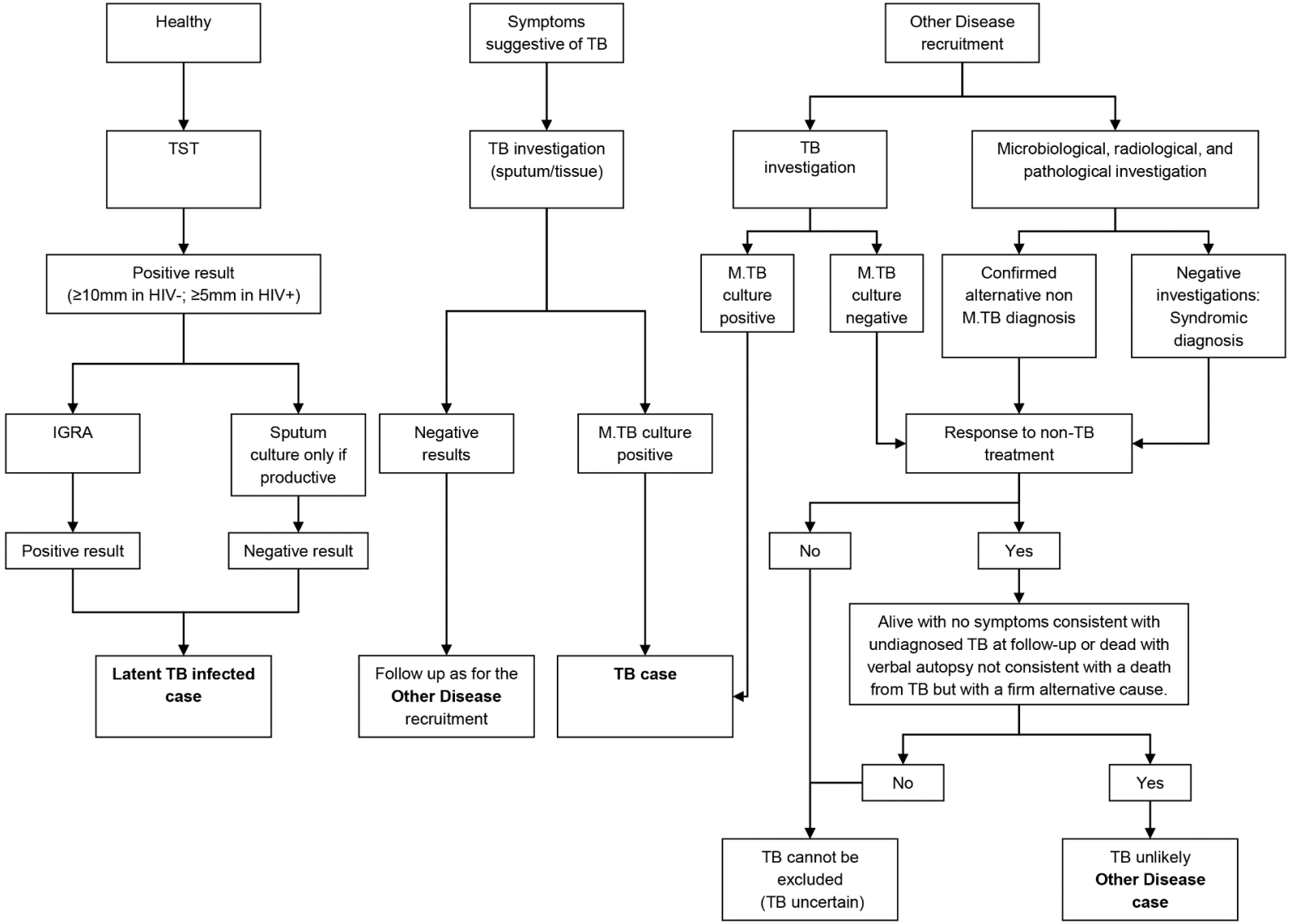


Figure 7 The diagnostic process followed in phase 1 of the EU-TB study to identify TB, OD and LTBI cases (from Kaforou et al, '13)

Permission to re-use granted under the Creative Commons Attribution License.[72]

Group	TB HIV+		TB HIV−		LTBI HIV+		LTBI HIV−		OD HIV+		OD HIV−	
Location	**SA**	**Malawi**	**SA**	**Malawi**	**SA**	**Malawi**	**SA**	**Malawi**	**SA**	**Malawi**	**SA**	**Malawi**
Number	49	60	47	59	48	41	50	36	68	38	49	39
Age in years median (IQR)	33.7 (29.0–38.3)	34.5 (29.6–43.2)	32.1 (26.3–42.7)	35.6 (26.2–53.1)	31.5 (27.9–37.4)	43.8 (35.4–49.4)	20.6 (19.1–23.4)	38.9 (32.3–50.9)	33.6 (28.6–37.9)	33.8 (29.4–41.3)	40.4 (28.7–53.5)	43.0 (27.0–53.9)
Sex (male, %)	40	52	70	58	27	22	42	53	38	34	45	28
Duration of symptoms/days median (IQR)	21 (0–33)	60 (14–210)	30 (21–30)	60 (30–240)	NA	NA	NA	NA	21 (6–90)	7 (3–90)	42 (7–130)	7 (2–365)
BMI (kg/m^2) median (IQR)	22.6 (19.5–25.2)	18.5 (16.9–20.7)	19.5 (18.0–22.5)	18.7 (16.5–20.2)	24.2 (20.6–28.4)	21.2 (18.6–23.9)	22.2 (21.4–25.7)	22.0 (20.2–23.4)	21.4 (20.0–24.6)	19.8 (18.3–22.2)	22.6 (18.4–24.9)	21.1 (19.6–22.2)
CD4 count/mm^3 median (IQR)	174 (64.7–293)[a]	128 (35–314)	NA	NA	326 (231–555)	312 (240–418)	NA	NA	197 (92–357)[b]	198 (111–270)	NA	NA
Anti-retroviral therapy	4 (8%)	14 (23.3%)	NA	NA	1 (2%)	0 (0%)	NA	NA	26 (38.2%)	16 (42.1%)	NA	NA
Tuberculin skin test induration (mm) median (IQR)	20 (15.5–22)[c]	ND	ND	ND	16 (10–20)	17 (0–25)	15 (12–20)	13 (11–17)	ND	0 (0–0)	ND	0 (0–9)
IGRA positive (see Methods)	ND	ND	ND	ND	48 (100%)	22 (53.7%)	50 (100%)	13 (36.1%)	ND	ND	ND	ND
Malaria positive	NA	2 (3.3%)	NA	2 (3.4%)	NA	1 (2.4%)	NA	0 (0%)	NA	3 (7.9%)	NA	2 (5.1%)

BMI, body mass index; HIV−, HIV-uninfected; HIV+, HIV-infected; IQR, inter quartile range; LTBI, latent TB infection; NA, not applicable; ND, not done; OD, other diseases (see Table 2); SA, South Africa; TB, active TB;
[a]Four missing values.
[b]Ten missing values.
[c]33 missing values, not routinely performed in the work up of TB+/HIV+ patients.
doi:10.1371/journal.pmed.1001538.t001

Figure 8 Numbers of patients recruited to each clinical group of the phase 1 EU-TB cohort.

Numbers are shown before sample collection. Information on demographic and clinical features is also shown (from Kaforou, *et al* '13).[72] Permission to re-use granted under the Creative Commons Attribution License.

Blood was collected from each patient into clot-activated serum tubes, centrifuged and aliquots stored at -80 degrees centigrade (deg. C.) in the respective laboratories. They were then shipped to the laboratory of Professor Mike Levin at Imperial College London for further storage at -80 deg. C. until the time of proteomic analysis by SELDI-TOF MS. A description of the proteomic workflow used in this study will now be given. These experiments were led by Dr Shea Hamilton.

ii) Proteomic workflow

Sera were thawed in ice and then aliquoted for analysis on each of the five separate chromatographic protein-binding SELDI arrays: CM10 (pH 4.0 and 6.0), Q10 (pH 7.5 and 9.5) and IMAC (Cu^{2+}). The chemistry and resultant binding affinity of each of these is shown in Table 2. Fixed amounts of high and low molecular weight standards were used in each run. The clinical groups were numbered 1 - 6 in order to blind the operators to the disease category of each patient.

Array type	Surface chemistry	Protein affinity
CM10	Weak anionic carboxylate groups	Positively charged
Q10	Quaternary ammonium groups	Negatively charged
IMAC	Nitrilotriacetic acid groups (bind metal cations that are applied in solution)	Metal-binding

Table 2 Surface chemistry and protein binding affinities of the four array types used in the EU-TB project.

IMAC: immobilised metal affinity chromatography. In the case of this project, copper cations were used.

The protein peaks from each of the five analyses were then amalgated for each patient and analysed in order to find the best signatures to diagnose TB, regardless of HIV status or site. In this project, two separate signatures were sought: one to distinguish active TB from OD and LTBI combined (TB vs. LTBI/OD), and the other to distinguish active TB from OD alone (TB vs. OD). Both were designed to do so irrespective of HIV status or site.

iii) Forward selection partial least squares

The statistical method used in this study was forward selection partial least squares (FS-PLS) linear regression.[73] Partial least squares regression extracts and works with the principal component of the distribution of each variable. The effect of this simplification is to minimise the chances of over-fitting. Forward selection means that the model (or signature) is built by starting with the variable (or protein) that is the most discriminatory individually, and then adding variables one by one that improve the combined ROC AUC. This process is continued until the overall predictive accuracy of the model is not enhanced further by the additional of further variables. In this study, patients were divided randomly into training and test sets (80:20), with the signatures being derived from the training set.

iv) EU-TB SELDI diagnostic signatures

The results of the two biosignature discovery experiments are shown in Table 3 below. Each protein is referred to first of all by its mass to charge ratio (m/z), with the final identity (where known) shown in the right hand column.

TB vs LTBI/ OD, HIV-/+						
Protein	**m/z mean**	**Intensity diff**	**P-value**	**aucTrain**	**aucTest**	**Protein Identity**
Q10_pH7.5_L_28	**4,357**	**10.06**	**4.59 E-37**	**0.84**	**0.81**	**ACT**
Q10_pH7.5_L_67	**13,762**	**19.38**	**1.05 E-22**	**0.88**	**0.81**	**TTR**
IMAC_L_38	**5,099**	**8.95**	**8.08 E-24**	**0.89**	**0.83**	*Under investigation*
CM10_pH4.0_L_62	**11,732**	**5.77**	**3.32 E-11**	**0.89**	**0.84**	**Beta-2-M**
TB vs OD, HIV-/+						
Protein	**m/z mean**	**Intensity diff**	**P-value**	**aucTrain**	**aucTest**	**Protein Identity**
Q10_pH9.5_L_28	**4,468**	**19.66**	**7.54 E-18**	**0.75**	**0.75**	**ACT**
CM10_pH6.0_L_42	**8,613**	**24.13**	**4.40 E-5**	**0.76**	**0.76**	**C4a desArg**
CM10_pH4.0_L_36	**13,802**	**6.74**	**6.37 E-5**	**0.81**	**0.78**	**TTR**
IMAC_L_36	**5,098**	**7.63**	**1.28 E-12**	**0.84**	**0.81**	*Under investigation*

Table 3 The EU-TB host protein signatures with performances as derived from the SELDI data shown as ROC AUC.

For each protein, the chip chromatographic conditions are shown first (left-hand column), followed by the mass to charge ratio (m/z) of the protein, the intensity difference between cases and controls, the individual P-value from Mann-Whitney tests and then the ROC AUC in the train and test sets. In the right-hand column, the identities of the protein biomarkers are shown, where known. The top row in each table represents the best individual biomarker, with subsequent rows reflecting the proteins that were added sequentially by the variable selection process (forward selection partial least squares linear regression). The ROC AUC values shown in each row are cumulative. ACT: alpha-1 antichymotrypsin; TTR: transthyretin; Beta-2-M: beta-2-microglobulin; C4a desArg: a de-arginated form of complement C4a.

An explanation of the protein identification experiments that were used in the case of each biomarker is beyond the scope of this thesis. The original aim for the serum proteomic arm of the EU-TB study, however, was that the biomarkers discovered by SELDI would then be re-tested by ELISA, in order to assess translatability to a multiplexed immunocapture-based POCT, e.g. lateral flow immunoassay.

5. Clinical variables and the Biology of TB

As mentioned above in section III. i), the other major benefit of studying host proteomes (or any “omes”) in disease states is that it gives insight into the biology of the disease process. Such an approach only gives half of the picture of the host-pathogen interaction, however the molecular state of the pathogen *in vivo* is extremely challenging to capture, particularly in diseases of low pathogen abundance such as TB. The host, however, can be studied accurately by this approach by virtue of access to biofluids containing high diversity of species that are abundant enough to be accurately quantified. In the case of serum proteins, the outputs of multiple organ systems can be studied simultaneously by virtue of their secretion (or release by tissue damage) into the circulation. Increasing knowledge about the biology of a disease is worthwhile from many perspectives as it could have immediate translational implications (e.g. biomarkers of severity or targets for host-directed therapy) or simply shed light on a complex pathogenesis that could impact future interventions.

As part of the EU-TB project, data was collected on multiple clinical and demographic variables for each patient. Variables for which data was gathered for at least 80% of the TB patients at both sites is shown in Table 4. Variables that were excluded due to missing data in excess of this rate included haemoglobin and white cell count.

Gender	BMI
Age	Sputum smear grade
Symptoms (Y/N)	HIV status
Symptom duration	CD4+ count (if HIV +ve)
Previous TB	ARV status (if HIV +ve)
Temperature	

Table 4 List of demographic and clinical variables in the EU-TB project for which information was gathered on participants.

Symptoms included night sweats, fever, malaise, cough, anorexia, weight loss, vomiting and diarrhoea. Symptoms were each scored as Yes/No and given a duration. Variables were excluded if the rate of missing data was > 20% at either study site. BMI: body mass index; ARV: antiretroviral.

Each of these variables could potentially be assessed for correlation with levels of host serum protein biomarkers. Any statistically significant correlations found might then shed light on the biology of TB or HIV-associated TB, or simply raise questions for further investigation. In addition, some of the variables are known markers of severity in TB: smear grade, BMI and (in the case of HIV co-infection) CD4+ count.[74-80]

6. Hypothesis

Host serum protein biomarkers in TB can form accurate diagnostic biosignatures for use in hospital settings by immunocapture, and reflect the underlying biology of the disease.

7. Primary Aims and Objectives

The primary aims of this thesis were two-fold. These are now given with their rationales plus explanations in bullet form.

i) To re-test the proteins from the AE-TBC study in a sample of the EU-TB cohort:

a) to assess how well the AE-TBC 7-protein diagnostic signature performs;

- this signature is now being taken forward for validation in the form of a multiplexed LFA for use in community settings.[81] If this signature were to perform well in secondary care settings as well it could be extremely beneficial for TB POC diagnostic testing in SSA as a whole;

b) to discover and assess the best new signature for the EU-TB cohort from all 22 proteins from the AE-TBC study;

c) to discover any promising individual biomarkers that could enhance the performance of the existing EU-TB signatures.

- the SELDI workflow was comprehensive, however accurate protein quantification by SELDI is only reliable below a molecular weight of approximately 30 kDa.

ii) To re-test the proteins from the EU-TB signatures by immuno-capture.

A diagnostic signature based on a small number (i.e. four) proteins could be highly impactful in low resource settings, however a mass spectrometry-based detection system is clearly impractical. The most desirable approach is likely immunocapture in the form of multiplexed LFA. Preservation of accurate protein quantification on switching to a different method, however, is not guaranteed. Different antibodies may need to be tried in order to achieve specificity (e.g. polyclonal vs. monoclonal) and proteins (or their fragments) will vary in their amenability to immunocapture. Careful validation of this technical switch in the laboratory is therefore required before translation to an LFA can be attempted.

8. Secondary Aim and Objectives

The secondary aim of this thesis is to discover new information about the biology of TB by looking for associations between levels of host serum proteins and clinical variables. Objectives within this are:

i) To seek associations of serum protein levels with markers of severity;

ii) To seek associations of serum protein levels with other clinical variables that illuminate understanding of the biology of TB.

9. Longitudinal cohort

In addition to the above Aims and Objectives for which serum from the EU-TB biobank will be used, we also had a small amount of plasma available for this thesis from a paediatric TB cohort that was recruited by Dr Elizabeth Whittaker (Clinical Lecturer, Paediatric Infectious Diseases and Immunology, Imperial College London). Patients in this cohort had samples collected both at the start and completion of TB treatment. Samples were offered for the purposes of

investigating whether levels of any specific proteins of interest returned to normal after treatment. Further details will be given in the Materials and Methods chapter (Chapter II) and the relevant Results chapter (Chapter IV).

CHAPTER II: MATERIALS AND METHODS

1. Introduction

This chapter describes the equipment, materials, study design, assay methodology and statistical methods that were used for the experiments that were performed for this thesis. Materials comprised sample biobanks and laboratory equipment. The Methods section then describes assay procedures and statistical methods.

2. Materials

i) Sample biobanks

In total, two sample biobanks were accessed: the serum biobank from the EU-TB (phase 1) study, and a smaller paediatric biobank for a "longitudinal study". Recruitment and sample collection in the EU-TB study was described in the Introductory Chapter.

Samples for the longitudinal study were taken from a biobank collected by Dr Elizabeth Whittaker, Academic Clinical Lecturer in Paediatric Infectious Diseases and Immunology, Imperial College London. These samples had been collected from children at Red Cross Hospital, Cape Town, South Africa during Dr Whittaker's PhD in 2011-2014. Dr Whittaker recruited children with active TB and collected whole blood in heparinised tubes before and after six months of TB treatment. Healthy controls were also recruited. These samples were also stored at St Mary's Hospital at -80°C. We utilised supernatants from unstimulated blood as a proxy for plasma. These samples were provided to investigate whether concentrations of any selected biomarkers of interest return to normal after treatment. Samples were available from 39 TB patients before and after treatment and 43 healthy controls. A description and explanation of which proteins were assayed in this cohort will be given in Chapter IV.

An explanation of which biobanks were used for which experiments, together with the assay method in each case, is shown below in Table 5.

Experiment	Biobank	Method
Re-test of protein biomarkers from the AE-TBC study in a sample of the adult EU-TB cohort	EU-TB study	Luminex
Re-test of protein biomarkers discovered by SELDI-TOF mass spectrometry in the adult EU-TB study	EU-TB study	ELISA
Comparison of levels of protein biomarkers in paediatric cohort before and after TB treatment, and with healthy controls	Longitudinal study	Luminex

Table 5 A description of which sample biobanks were used for which experiments in this thesis.

The assay method that was used is also shown. "Longitudinal study" refers to a paediatric TB cohort that was recruited by Dr Elizabeth Whittaker, Clinical Lecturer, Imperial College London and followed up at the end of treatment.

ii) Equipment and consumables used for Luminex assays

- -80°C. freezer
- Dry ice
- Ice machine (H_2O)
- Microcentrifuge with refrigeration
- Class II biosafety cabinet
- Variable volume pipettes, tips and filter tips
- 0.5ml, 1.5ml and 2.0 ml Eppendorf protein LoBind tubes (manufacturer numbers 022431064 (Sigma catalogue number Z666491), 022431081 (Sigma catalogue number Z666505) and 022431102 (Sigma catalogue number Z666513) respectively)
- Commercially available Luminex kits (see below)
- Cold room
- Sterilin polystyrene 7ml Bijou and 30ml Universal containers
- Double distilled water
- Vortex
- Aluminium foil
- Plate shaker
- Bio-Plex handheld Magnetic washer (Bio-Rad 171020100)
- 20% Surfanios
- Kimtech lint-free wipes
- Bio-Plex 200 Luminex plate reader (Bio-Rad)
- Sweetie jars
- 70% ethanol

iii) Additional equipment and consumables used for ELISA

- Commercially available ELISA kits (see below)
- VersaMax plate reader (Molecular Devices)

iv) Luminex panels/kits:

- Human Acute Phase Proteins (5-plex and 4-plex Panels; Bio-Rad 171A4S07M)
- Human Matrix Metalloproteinase Panel 2 (Merck Millipore HMMP2MAG-55K)
- Human Neurodegenerative Disease Panel 1 (Merck Millipore HNDG1MAG-36K)
- Human Cytokine/Chemokine Panel (Merck Millipore HCYTOMAG-60K)

A list of which proteins were assayed as part of each Luminex panel is shown below in Table 6. The list of proteins and of the kits that were used were taken from the AE-TBC study.(1)

Luminex panel	Proteins assayed
Acute Phase 5-plex	Procalcitonin, tissue plasminogen activator, fibrinogen, ferritin, serum amyloid A
Acute Phase 4-plex	Alpha-2-macroglobulin, CRP, haptoglobin, serum amyloid P
MMP	MMP-2, MMP-9
Neurodegenerative	Apolipoprotein (apo-)-AI, apo-CIII, transthyretin, complement factor H
Cytokine	TNF-alpha, IL-1 ra, IFN-alpha-2, TGF-alpha, VEGF, IFN-gamma, IP-10

Table 6 A list of which proteins were assayed as part of each Luminex panel.

The same panels were used as in the AE-TBC study.(1)

v) ELISA kits:

- Human transthyretin ELISA kit (Abcam ab108895)
- Human beta-2-microglobulin ELISA kit (Abcam ab108885)
- Human alpha-1-antichymotrypsin ELISA kit (Abcam ab157706)

vi) Software:

- SoftMax Pro v5 (for ELISA)
- Bio-Plex Manager v6.1 (for Luminex assays)
- Microsoft Office 2013 (Word and Excel)
- Prism v8 Graphpad Software Inc. La Jolla, USA
- EndNote X9

3. Methods

The methods employed for the Primary Aims of this thesis stated in the Introduction Chapter will be addressed first (Section 7 i) and ii)). These were:

i) <u>To re-test the proteins from the AE-TBC study in a sample of the EU-TB cohort;</u>

ii) <u>To re-test the proteins from the EU-TB signatures by immuno-capture.</u>

The methods employed for the analysis of protein levels in the longitudinal cohort, and for analyses of associations between protein levels and clinical variables will be explained second (X.3.4 – X.3.5).

i) Luminex assays of the 22 proteins in the AE-TBC study

a) Selection of serum samples from EU-TB biobank for Luminex assays

In the protein biomarker study performed by the AE-TBC, sera from 214 TB patients and 487 other disease (OD) patients were analysed (N = 701).(1) We aimed to have a sample size as close as possible to this in order to accurately retest the signature whilst still keeping numbers equal across our clinical groups. Owing to

slightly lower numbers of patients enrolled to the LTBI groups in the EU-TB study (especially at the Karonga site: 36 LTBI (HIV-) and 41 LTBI (HIV+)), we aimed to analyse 72 patients from each of our six clinical groups (36 from each site), i.e. 432 in total. It was estimated that six Luminex kits from each panel would be needed in order to fulfil this, which would then incorporate 438 patients. 438 patients were therefore selected randomly from the clinical databases using random number generation (Microsoft Excel 2013), with equal numbers from each of the three clinical groups.

b) Luminex assay procedure

Sera from patients (and LTBI controls) were randomised across the series of plates in each panel, resulting in approximately equal numbers from each clinical group in each plate. In each plate, the first three patients were run in duplicate to provide estimates of intra-assay variability. Quality controls were also run on each plate as per manufacturer's instructions. Before starting the assay series, original serum cryovials were thawed and fresh 45uL aliquots were taken. On the first day of the first series of assays ("Acute Phase 5-plex"), aliquots were re-thawed and diluted as shown in Table 7. Assays were run in the order shown in Table 7, in order to minimise the number of freeze-thaw cycles. Dilutions were performed according to manufacturer's instructions, except in the case of the MMP panel for which sera were diluted further following optimisation experiments (to 1 in 100).

Luminex panel	Serum dilution process	Final dilution
Acute Phase 5-plex	5uL serum added to 495uL of diluent	1 in 100
Acute Phase 4-plex	5uL of 5-plex dilutions added to 495uL diluent	1 in 10,000
MMP	1uL serum diluted in 99uL diluent	1 in 100
Neurodegenerative	5uL of MMP dilutions diluted in 1495uL diluent	1 in 30,000
Cytokine	25uL of serum used	Neat

Table 7 A description of the serum dilution process that was followed for the Luminex assays on the EU-TB sera.

Resulting final dilutions are also shown. Assays were performed in the order in the table, starting at the top row.

At the end of each assay, beads were analysed on a Bio-Plex 200 analyser. A minimum of 50 beads were read per analyte in each well, with doublet discriminator gate settings of 5,000 to 25,000. Bio-Plex Manager v6.1 was used for data acquisition (Bio-Rad Laboratories, Hercules, California, USA). Quality controls fell within the expected range in each plate. Intra-assay variability, calculated as the mean of the coefficients of variance of all duplicates for each analyte individually across all plates, was <12% for all 22 proteins. If a result for any patient serum protein concentration was below the lower limit of quantification, it was assigned a value of zero. If above the upper limit, the serum was re-run at a higher dilution. Extrapolations to outside of standard curves in both directions were accepted up to the limit determined by the software.

For the Acute Phase 4-plex assays, 181 patients had to be re-run at 1 in 500,000 dilution owing to fluorescence intensities for Haptoglobin being greater than the upper limit of quantification.

ii) ELISA of the three proteins in EU-TB diagnostic signatures derived from SELDI-TOF mass spectrometry

a) Selection of serum samples from EU-TB biobank for ELISA

In order to be adequately powered to detect significant differences between levels of markers in the different clinical groups in the EU-TB study, we selected 32 patients from each group, i.e. 192 in total. This number was used as a starting point based on previous experiments measuring human serum proteins with different underlying distributions and with α (one-tailed) = 0.05. The α was one-tailed because the aim was to look for replication of the SELDI-TOF findings (i.e. a marker should be higher in the TB group by ELISA if it was higher in the TB group by SELDI but lower by ELISA if it was lower by SELDI). Patients were selected randomly from each group as recorded in the clinical databases from each of the two recruitment sites (Karonga, Malawi and Cape Town, South Africa). The numbers of patients selected from each site were equal for each of the six clinical groups. The resulting combined list of 192 patients was then re-randomised in Excel and this list used for constructing templates for each assay for each protein.

b) ELISA procedure

At the time of commencement of this thesis, three of the proteins in the EU-TB signatures had been identified: transthyretin, beta-2-microglobulin (beta-2-M) and alpha-1-antichymotrypsin (ACT). For the first protein (transthyretin), the original cryovials were thawed and 25uL aliquots were taken. These were then frozen after each assay and re-thawed for analysis of the next protein. Thawing was done in ice on each occasion to minimise protein degradation. Beta-2-M was measured second and ACT third. All ELISA were performed as per manufacturer instructions. Standards, blanks and samples were all run in duplicate. For B2M and ACT, a control serum was also included in each plate, in duplicate.

At the end of each assay, stop solution was added as directed and absorbance at 450nm measured in each well with a VersaMax plate reader. SoftMax Pro v5 was then used for data acquisition.

iii) Statistical analyses

a) Analyses of individual markers for diagnostic performance

Data on each individual protein from both the ELISA and Luminex assays was analysed using Prism v8. For each protein (25 in total), data from both the HIV infected and uninfected clinical groups were combined to make three groups: Active TB (TB), LTBI and OD. Data were then analysed using the Mann-Whitney U test to assess whether levels in the TB group were different to in the relevant control group. In addition, areas under the receiver operating characteristic curve were calculated to estimate the discriminatory performance of each protein by ELISA, for comparison with that from the SELDI intensity. The control groups for each group of biomarkers were:

- LTBI and OD combined for all three proteins in the TB vs. LTBI/OD EU-TB signature (ACT, transthyretin and beta-2-M);
- OD alone for the two proteins in TB vs. OD alone EU-TB signature (ACT and transthyretin);
- OD alone for all 22 proteins in the AE-TBC study.

Sera from the healthy LTBI group were primarily included in the assays of the 22 proteins in the AE-TBC study to provide a reference point, however we did also calculate individual ROC AUCs of each protein to distinguish TB from LTBI alone.

b) Analyses of markers in combination for diagnostic performance

For the bio-signature discovery analyses, only TB and OD patients were included, and only those for whom data on the concentration of every protein was obtained. In total, this was 249 patients from the EU-TB cohort (122 TB and 127 OD). An illustration of the process followed from original recruitment to this point in this experiment is shown in Figure 9.

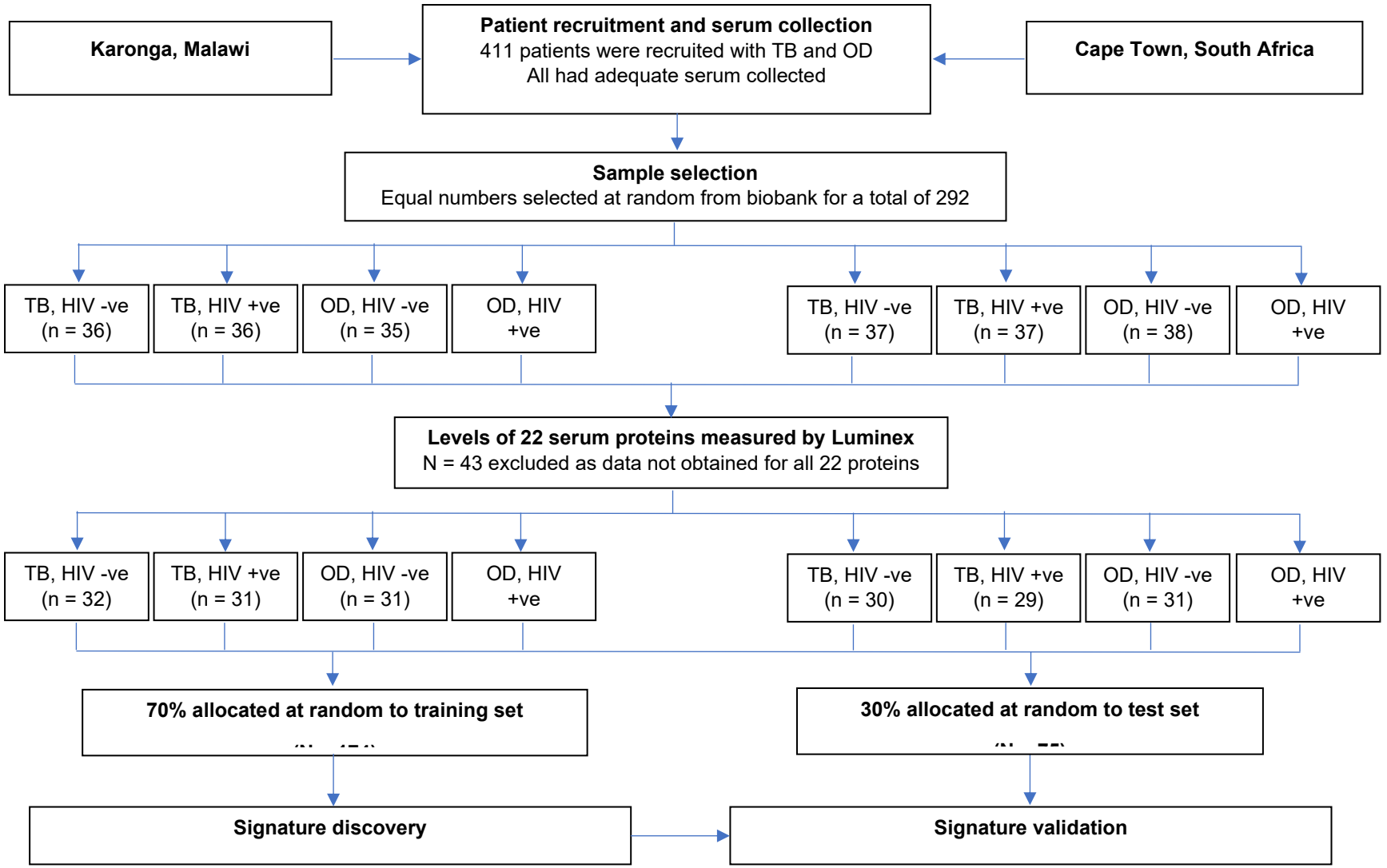


Figure 9 Flow diagram showing the process followed from patient recruitment and serum collection onwards for discovery of new bio-signatures in the EU-TB cohort.

For the application of the 7-protein AE-TBC signature to the EU-TB cohort, allocation to training and test sets was not required. TB: tuberculosis, OD: other diseases, HIV -ve: HIV uninfected, HIV +ve: HIV infected.

c) Analysis of performance of the seven-protein signature in the EU-TB cohort

Application of the seven-protein AE-TBC signature to the EU-TB cohort was performed by Professor Martin Kidd at the Centre for Statistical Consultation, Stellenbosch University, using Statistica (Statsoft, Ohio, USA). The same general discriminant analysis (GDA) method was used to assess the performance of the top 7-protein signature as in the AE-TBC study.(1) This method involves logarithmic transformation of the data on each protein. As advised by Professor Kidd, for the discovery cohort in this experiment, the entire AE-TBC cohort was used (n = 707). For the validation cohort, the whole of the sample described above from the EU-TB cohort (in section X.3.1.1) was used (n = 249).

d) Analysis of data on all 22 proteins for the best signature in our study

Efforts to discover the best new biosignature for discriminating TB from OD based on concentrations of the 22 proteins from the AE-TBC study in the EU-TB cohort only (n = 249) were made in parallel by Professor Kidd and Dr Clive Hoggart (Research Fellow in Statistical Analysis, Levin group, Imperial College London). Patients were allocated into training and test sets at a ratio of 70:30 using random number generation (Excel). Professor Kidd applied the same approaches as in the AE-TBC study, namely GDA and Random Forests.(1) Dr Hoggart performed a Parallel Regularised Regression Model Search method (PreMS) in "R", having normalised the data into deciles. The PReMS method has been published and is freely available online.(2) The same allocations to training and test sets were used for each method.

e) Optimisation of diagnostic cut-offs for increased sensitivity and specificity

For a screening test, albeit for use in the community, the WHO recommend a minimum sensitivity of 90%.(3) No criteria for a rule-in test are specified. After obtaining the best new signature from each of the above methods, we therefore re-tested them after adjusting the cut-off for defining a positive result to increase each of the sensitivity and specificity in turn to 90%. This was to assess the performance of each signature as either a rule-out or rule-in test for TB. There were no indeterminate test results.

f) Analysis of the performance of the EU-TB signatures using the ELISA data

Analysis of the performance of the proteins from the EU-TB study as a signature using immunocapture (i.e. ELISA) data cannot be completed until the remaining proteins have been identified and quantified using immunoassay. This process is being completed separately from this thesis under the leadership of Dr Shea Hamilton.

4. Analysis of serum protein levels before and after treatment in a paediatric cohort

As stated above, supernatants from unstimulated whole blood were available from 39 children with TB (before and after treatment) and 43 healthy controls. The demographic and clinical features of this cohort are provided in Chapter IV.

As explained in Chapter IV, levels of complement factor H were assayed in this cohort. This was performed using assay components that were available following the Luminex assays described above. Along with factor H, data on levels of apo-AI, apo-CIII, and transthyretin were also obtained (the "Neurodegenerative panel"). In order to obtain data on all of the children's samples, however, one further Luminex kit was purchased. In this extra kit, however, the assay for transthyretin had been replaced by complement C3. It was therefore not possible to obtain data on all of the proteins from every child, however the main aim of this experiment was to investigate complement factor H. Exact numbers of children on whom data was gathered for each protein is given in Chapter IV.

Prior to these assays, samples had been diluted 1 in 10 in RPMI medium. Samples were therefore diluted further by a factor of 1 in 3000 (Table 7, above). To compare protein levels before and after treatment in each child, Wilcoxon matched pairs signed-rank tests were used. One-sided tests were performed to investigate whether levels returned to, or moved towards, those of the healthy control group. To compare levels at the end of treatment with those in the healthy control group, Mann-Whitney tests were used.

5. Analyses of markers for biologically meaningful associations with clinical variables

After data on the levels of all AE-TBC 22 proteins in our cohort had been obtained, an untargeted survey was performed to look for associations between these and all clinical variables for which: a) sufficient data was available, and b) with which new associations would be of scientific interest. The definition of "sufficient" was if data was available on at least 80% of the patients at each study site (i.e. Karonga and Cape Town). To correct for multiple comparisons, resulting p-values were then adjusted using the Bonferroni-Sidak method. Multivariate analyses were also performed where indicated.

Statistical methods used to determine associations were predicated on the nature of the data on the clinical variable of interest: for continuous variables (e.g. age), Spearman's rank correlation analyses were performed. For categorical data, e.g. smear status, Mann-Whitney tests were used. All such analyses were performed using Prism v8. If multivariate analyses were deemed necessary, these were also performed in Prism v8.

CHAPTER III RESULTS: Re-test of 22 protein biomarkers from the AE-TBC in the EU-TB cohort

1. Introduction

This chapter presents the results of Primary Aim i) of this thesis which was to re-test the proteins from the AE-TBC study in a sample of the EU-TB cohort. There were three objectives within this Aim:

a) to assess how well the AE-TBC 7-protein diagnostic signature performs;

b) to discover and assess the best new signature for the EU-TB cohort from all 22 proteins from the AE-TBC study;

c) to discover any promising individual biomarkers that could enhance the performance of the existing EU-TB signatures.

As described in the Materials and Methods chapter, levels of the 22 proteins from the AE-TBC study were quantified in sera from 438 patients from the EU-TB cohort by Luminex. The demographic and clinical features of those patients are presented in Table 8, below. The list of diagnoses in the OD group is shown in Table 9.

Results are presented in the following stages. Firstly, levels of each protein in each clinical group are shown by virtue of fold change from the levels in the LTBI group (Heat Map 1). Next, they are presented as medians and interquartile ranges, regardless of HIV status (Table 10). As shown in Table 10, a different number of patients had results on each individual protein owing to variation in the number of samples that needed to be re-run in each Luminex panel. Following this, scatter-box plots will be shown to display individual patient results for each protein, both stratified by and regardless of HIV status (Figures 10 – 31). For each protein, levels in the TB group will be compared with levels in each of the OD and LTBI groups in turn. If there was a difference in levels between TB patients with and without HIV co-infection, this will also be stated. For each protein, the area under the receiver operating characteristic curve (ROC AUC) to distinguish TB from OD alone will then be given, regardless of HIV status.

After results for each protein have been presented, individual ROC AUCs are presented side-by-side, including after stratifying by each of HIV status and site (Figures 32 – 35). Individual ROC AUCs will then be compared with the corresponding ROC AUCs for each protein from the AE-TBC study (Table 11 and Figure 36). Following this, ROC AUCs of each protein individually to distinguish TB from LTBI are shown (Table 12). Finally, the results of the various biosignature analyses will be presented, including after adjusting the cut-off for defining a positive test for each of increased sensitivity and specificity in turn (Tables 13 – 19).

	TB, HIV-		TB, HIV+		LTBI, HIV-		LTBI, HIV+		OD, HIV-		OD, HIV+	
Site	**CPT**	**Karonga**	**CPT**	**Karonga**	**CPT**	**Karonga**	**CPT**	**Karonga**	**CPT**	**Karonga**	**CPT**	**Karonga**
Number	37	36	37	36	40	33	37	36	38	35	41	32
Age (IQR)	32·5 (26·5-41·9)	36 (25·5-53·4)	33·8 (29-37·9)	33·2 (28-39·7)	20·7 (19·3-23·4)	39 (32·4-51·4)	31·2 (27·9-35·1)	44·5 (35·5-49)	41·2 (29·2-51)	43·2 (27·5-53·6)	33·6 (28·6-36·7)	33·3 (29·4-41·2)
Male (n (%))	25 (69·4)	19 (52·8)	16 (43·2)	19 (52·8)	16 (40)	15 (45·5)	10 (27)	9 (25)	17 (44·7)	11 (31·4)	16 (40)	11 (34·4)
CD4+ count (IQR)	n/a	n/a	170 (69-293)	168 (45-276)[a]	n/a	n/a	345 (231-523)	312 (246-421)	n/a	n/a	183 (95-272)	182 (107-229)
On ART (n (%))	n/a	n/a	2 (5·4)	8 (22·2)	n/a	n/a	1 (2·7)	0 (0)	n/a	n/a	18 (45)	12 (37·5)
BMI (IQR)	20·4 (18·4-23·6)[b]	18·4 (16·6-19·3)	20·9 (18·6-23·4)	18·8 (18-20·8)	23·6 (21·5-28·7)	22·7 (20·6-23·7)	24·3 (20·9-27·9)	21·6 (18·7-23·7)	22·4 (20·1-23·8)[c]	20·8 (19·6-23)	21·2 (19·8-23·9)[d]	19·6 (18·1-21·5)

Table 8 Demographic and clinical features for the 438 participants randomly selected for this study from the EU-TB cohort.

Patients are shown by study site for each of the six clinical groups: active TB (TB), healthy controls with latent TB (LTBI), and unwell patients with other diseases (OD) who initially had TB in their differential diagnosis. The definition of a TB case was culture positivity for *Mycobacterium tuberculosis* (Mtb). The definition of an OD case was a confirmed other diagnosis plus exclusion of TB. HIV+: HIV infected; HIV-: HIV uninfected. CPT: Cape Town. IQR: interquartile range; ART: antiretroviral therapy; BMI: body mass index; n/a: not applicable. Numbers of missing values were less than five except where indicated by letters a – d (a = 6, b = 5, c = 19, d = 20).

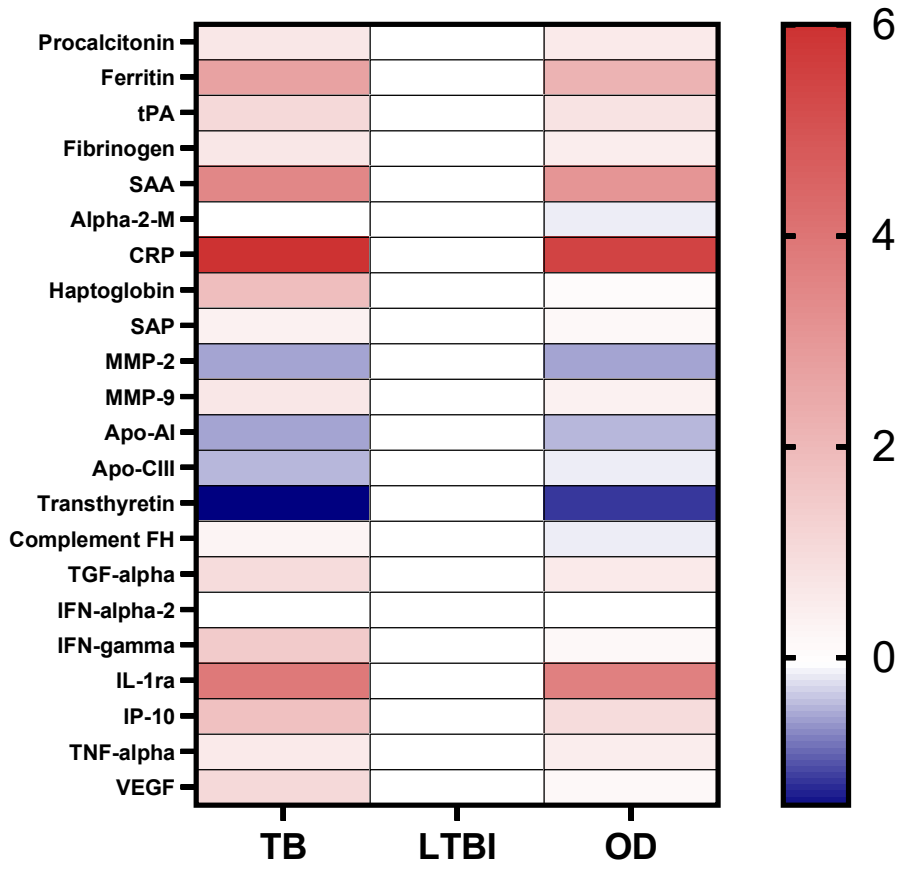
	HIV uninfected		HIV infected		Total (% of OD group)
	Karonga	Cape Town	Karonga	Cape Town	
Pneumonia/Bronchitis/PCP	12 (33%)	6 (16%)	15 (47%)	15 (38%)	**48 (33%)**
Malignancy or neoplasia other than KS*	3 (8%)	13 (35%)	1 (3%)	2 (5%)	**19 (13%)**
Genitourinary	6 (17%)	4 (11%)	2 (6%)	1 (3%)	**13 (9%)**
Meningitis (bacterial/viral/unspecified)	4 (11%)	0 (0%)	4 (13%)	2 (5%)	**10 (7%)**
Gastroenteritis/Hepatitis	0 (0%)	2 (5%)	0 (0%)	6 (15%)	**8 (6%)**
Kaposi's Sarcoma	0 (0%)	0 (0%)	1 (3%)	6 (15%)	**7 (5%)**
Pyelonephritis	0 (0%)	7 (19%)	0 (0%)	0 (0%)	**7 (5%)**
Cryptococcal meningitis	0 (0%)	0 (0%)	2 (6%)	3 (8%)	**5 (3%)**
Pleural effusion/empyema (non-TB)	0 (0%)	1 (3%)	0 (0%)	4 (10%)	**5 (3%)**
Bacteraemia (source not identified)	1 (3%)	0 (0%)	4 (13%)	0 (0%)	**5 (3%)**
Other**	5 (14%)	0 (0%)	0 (0%)	0 (0%)	**5 (3%)**
Hepatobiliary disease	0 (0%)	4 (11%)	0 (0%)	0 (0%)	**4 (3%)**
Peritonitis	3 (8%)	0 (0%)	1 (3%)	0 (0%)	**4 (3%)**
Malaria	1 (3%)	0 (0%)	1 (3%)	0 (0%)	**2 (1%)**
IBD	0 (0%)	0 (0%)	0 (0%)	1 (3%)	**1 (1%)**
Pyomyositis	1 (3%)	0 (0%)	0 (0%)	0 (0%)	**1 (1%)**
Persistent generalised lymphadenopathy	0 (0%)	0 (0%)	1 (3%)	0 (0%)	**1 (1%)**
TOTAL	**36**	**37**	**32**	**40**	**145**

Table 9 Major clinical diagnoses in the Other Diseases groups.

Data are stratified by HIV status and site. *: lung cancer (n = 7), lymphoma (n = 3), dermatological tumour (n = 2), unspecified (n = 2), mesothelioma (n = 1), hepatocellular carcinoma (n = 1), metastatic adenocarcinoma (n = 1), benign salivary gland tumour (n = 1). **: epilepsy (n = 3), headache (n = 1), pain unspecified (n = 1). One patient had no diagnosis listed, hence data is shown for 145 patients.

2. Individual serum protein levels and ROC AUCs (TB vs. OD)

Levels of each protein in each of the three clinical groups: TB, LTBI and OD are shown first, followed by data on individual ROC AUCs for distinguishing TB from OD. As shown below in Heat Map 1 on the basis of median protein levels, out of the proteins that were higher in TB than LTBI, levels of CRP showed the greatest fold change; and out of the proteins that were lower in TB, levels of transthyretin showed the greatest fold change. The only protein whose levels changed in a different direction in the TB and OD groups was complement factor H, which was higher in TB but lower in OD. Median and interquartile ranges are then shown in Table 10.



Heat Map 1. Differences in median levels of proteins between clinical groups. Differences were plotted as log-tranformed fold-changes (base 2) in comparison to the LTBI group. The legend on the right shows the colour scale in relation to log fold-change. Red: increased level; blue decreased.

	TB			LTBI			OD		
	No.	Median	IQR	No.	Median	IQR	No.	Median	IQR
Procalcitonin	131	1·90	(1·20 - 2·64)	133	1·20	(0·73 - 1·79)	132	1·79	(1·20 - 2·87)
Ferritin	131	74·45	(30·50 - 139·55)	133	11·78	(5·59 - 20·18)	132	54·32	(20·93 - 135·91)
tPA	131	2·72	(1·86 - 3·74)	133	1·29	(0·92 - 2·03)	132	2·22	(1·65 - 3·30)
Fibrinogen	131	2·71	(2·01 - 3·46)	133	1·65	(0·93 - 2·32)	132	2·34	(1·51 - 3·52)
SAA	131	6·36	(4·23 - 8·59)	133	0·55	(0·23 - 1·73)	132	4·58	(2·83 - 6·25)
Alpha-2-M	134	1·44	(1·22 - 1·77)	134	1·49	(1·19 - 1·76)	133	1·36	(1·09 - 1·62)
CRP	134	63·35	(29·20 - 100·12)	134	1·01	(0·54 - 3·28)	133	45·51	(9·55 - 163·47)
Haptoglobin	134	4·83	(0·35 - 12·75)	134	1·33	(0·20 - 3·43)	133	1·38	(0·28 - 5·70)
SAP	134	42·28	(34·19 - 52·28)	134	31·22	(25·17 - 38·05)	133	35·60	(24·82 - 47·74)
MMP-2	146	77·70	(63·56 - 100·63)	146	110·66	(85·30 - 134·00)	146	79·71	(59·99 - 110·24)
MMP-9	146	335·97	(195·54 - 516·25)	146	210·23	(106·45 - 317·56)	146	276·52	(149·94 - 571·16)
Apo-AI	146	230·81	(171·61 - 291·63)	146	324·44	(253·91 - 418·56)	146	241·75	(164·52 - 335·23)
Apo-CIII	146	85·49	(66·71 - 122·84)	146	113·88	(87·45 - 145·40)	146	105·62	(68·025 - 147·09)
Transthyretin	146	69·64	(48·80 - 104·55)	146	189·83	(129·33 - 260·92)	146	89·65	(59·68 - 158·53)
Complement FH	146	391·88	(324·43 - 455·99)	146	317·51	(287·54 - 359·88)	146	303·87	(257·64 - 376·53)
TGF-alpha	142	10·37	(6·01 - 19·37)	137	5·35	(2·56 - 11·17)	142	8·32	(2·94 - 18·09)
IFN-alpha-2	142	0·00	(0·00 - 33·50)	137	0·00	(0·00 - 14·51)	142	0·00	(0·00 - 35·88)
IFN-gamma	142	16·31	(7·70 - 40·85)	136	5·94	(2·33 - 18·92)	142	6·79	(2·35 - 20·65)
IL-1ra	142	16·71	(2·59 - 49·02)	137	1·10	(0·00 - 17·85)	141	14·56	(0·00 - 75·25)
IP-10	140	1·08	(0·66 - 2·04)	137	0·30	(0·18 - 0·65)	142	0·60	(0·25 - 1·20)
TNF-alpha	142	25·04	(15·90 - 39·73)	137	16·02	(10·75 - 24·02)	142	22·04	(13·25 - 31·95)
VEGF	139	291·97	(135·80 - 674·27)	137	134·09	(70·81 - 265·27)	139	157·41	(61·09 - 307·75)

Table 10 Serum concentrations of each protein in each of the three main clinical groups.

Data are shown regardless of HIV status and site. For each protein, the number of patients in each group for which data was gathered are shown. TB: active tuberculosis; OD: other diseases; LTBI: latent TB. IQR: interquartile range. Units of protein concentration: mg/ml: Alpha-2-M and Haptoglobin; ug/ml: fibrinogen, serum amyloid A, CRP, serum amyloid P, MMP-2, MMP-9, apo-AI, apo-CIII, transthyretin, and complement FH; ng/ml: procalcitonin, ferritin, tPA, and IP-10; and pg/ml: TGF-alpha, IFN-alpha-2, IFN-gamma, IL-1ra, TNF-alpha, and VEGF.

Levels of proteins in each patient in each of the clinical groups are now shown in turn.

Levels of procalcitonin were significantly higher in the active TB patients compared to LTBI. However, no difference was observed between TB and the OD groups (Figure 10, below). The area under the receiver operating characteristic curve (ROC AUC) to distinguish active TB from OD, regardless of HIV status, was 0.52 (95% confidence interval: 0.45 - 0.59).

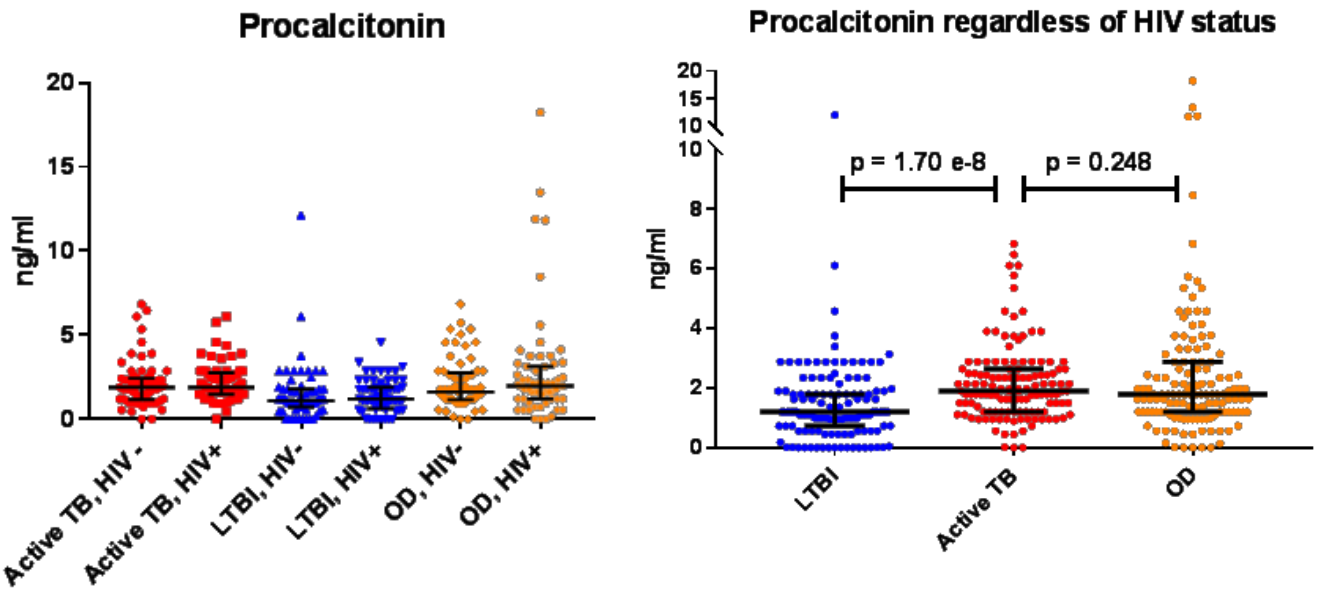


Figure 10 Serum levels of procalcitonin in each of the six clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of ferritin were significantly higher in the active TB groups than in both the LTBI and OD groups (Figure 11, below). The levels of this protein were also significantly higher in HIV infected individuals than HIV uninfected, both across all patient groups together (p = 9.70 e-3) and in the active TB patients (p = 0.024; two-sided p values from Mann-Whitney U tests). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.57 (95% confidence interval: 0.50 - 0.64).

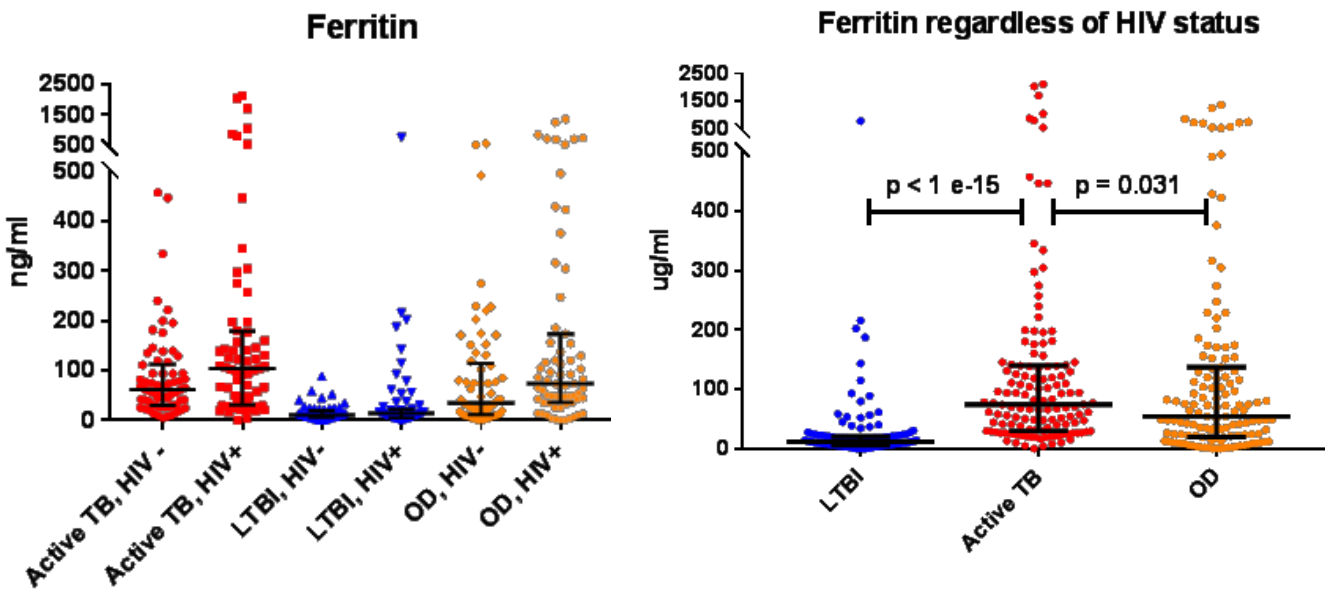


Figure 11 **Serum levels of ferritin in each of the clinical groups in the EU-TB study measured by Luminex.**

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of tissue plasminogen activator (tPA) were significantly higher in the active TB groups than in both LTBI and OD groups (Figure 12, below). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.57 (95% confidence interval: 0.49 - 0.64).

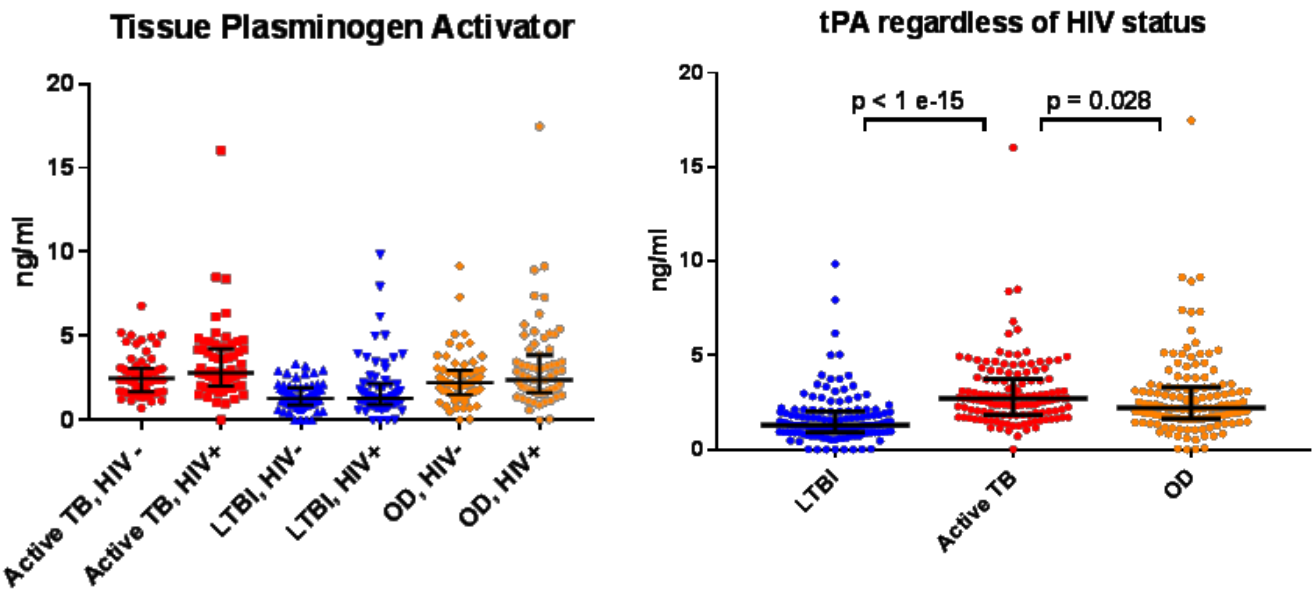


Figure 12 Serum levels of tissue plasminogen activator (tPA) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of fibrinogen were significantly higher in the active TB groups that in both LTBI and OD groups (Figure 13, below). The levels of fibrinogen were also significantly higher in HIV infected patients, both across all patient groups together (p = 0.01) and in the active TB patients (p = 0.046). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.56 (95% confidence interval: 0.49 - 0.63).

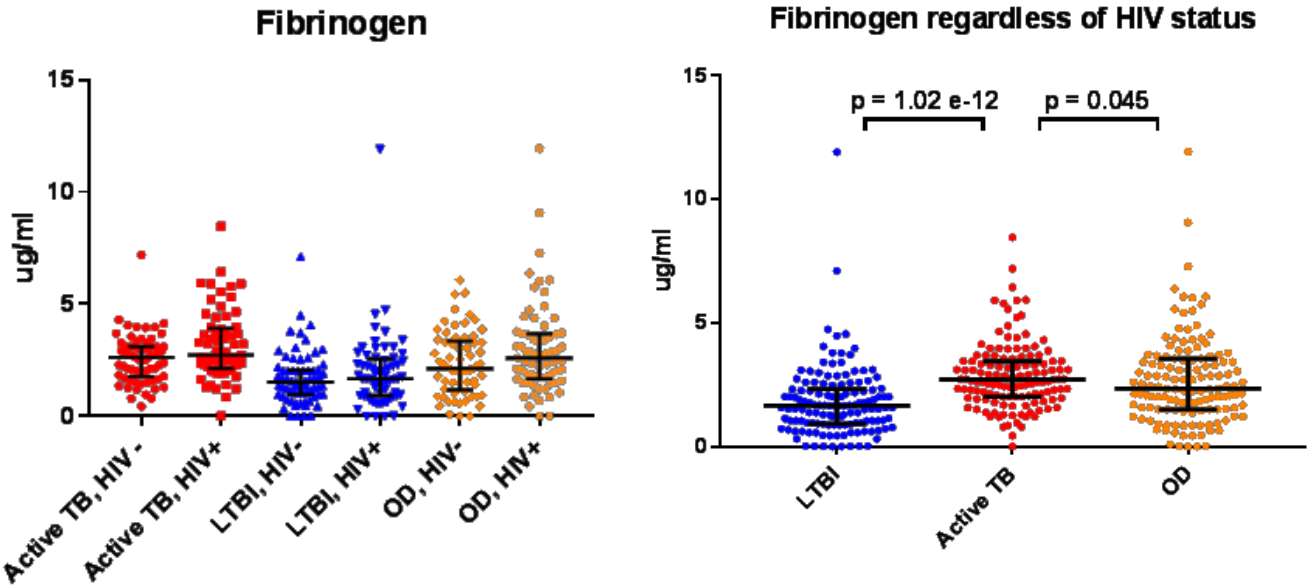


Figure 13 Serum levels of fibrinogen in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of serum amyloid A (SAA) were significantly higher in the active TB groups than in both LTBI and OD groups (Figure 14, below). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.65 (95% confidence interval: 0.58 - 0.71).

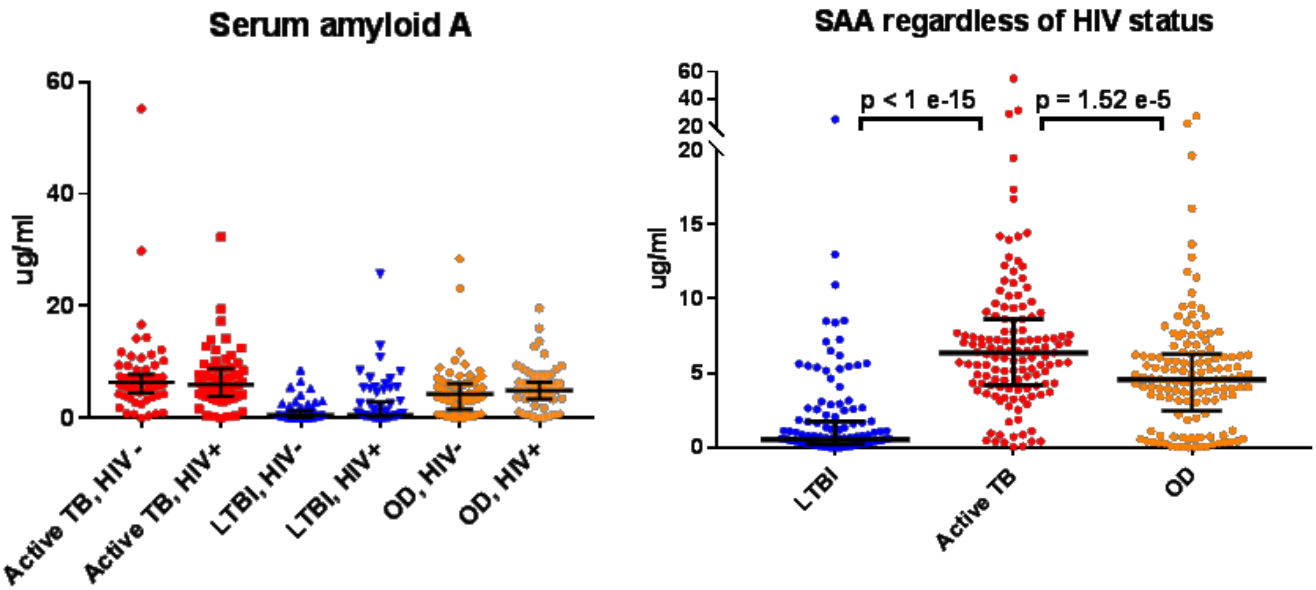


Figure 14 Serum levels of serum amyloid A (SAA) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of alpha-2-macroglobulin (alpha-2-M) were significantly higher in the active TB groups than in the OD groups, but not significantly different from levels observed in the LTBI groups (Figure 15, below). Levels of alpha-2-M were also significantly lower in patients infected with HIV, both across all groups (p = 1.00 e-7) and in the active TB patients (p = 1.48 e-4). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.57 (95% confidence interval: 0.50 - 0.64).

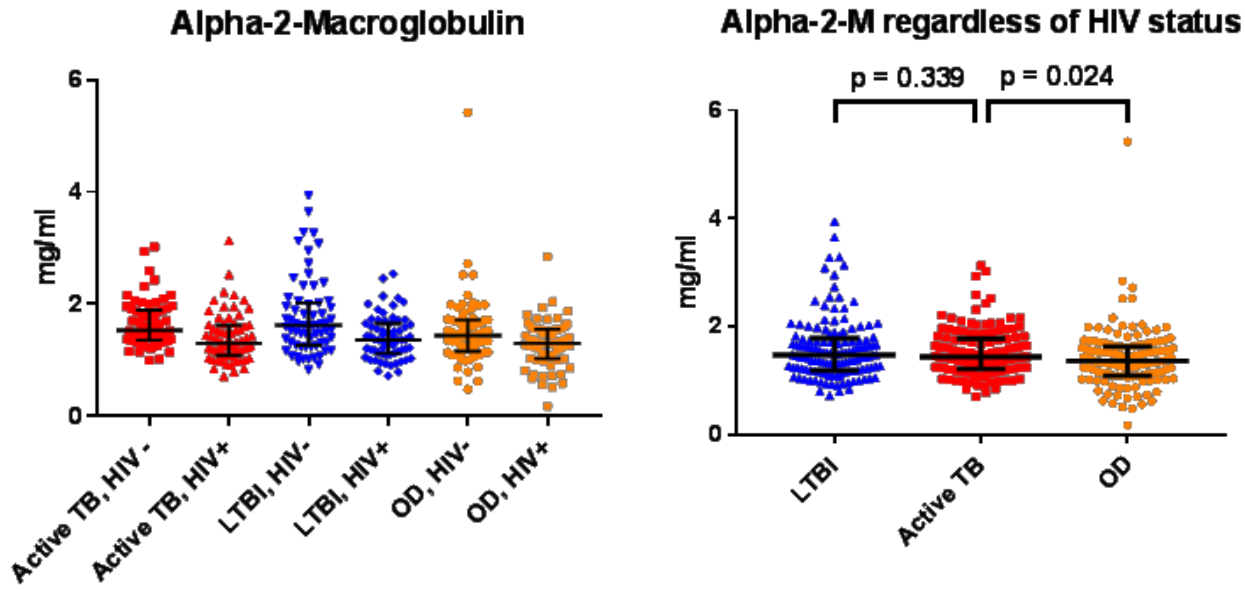


Figure 15 Serum levels of alpha-2-macroglobulin (alpha-2-M) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of serum amyloid P (SAP) were significantly higher in the active TB groups than in both the LTBI and OD groups (Figure 16, below). The levels in active TB patients that were co-infected with HIV (TB, HIV+) were significantly lower than in mono-infected TB patients (TB, HIV-, p = 8.62 e-3). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.64 (95% confidence interval: 0.57 - 0.71).

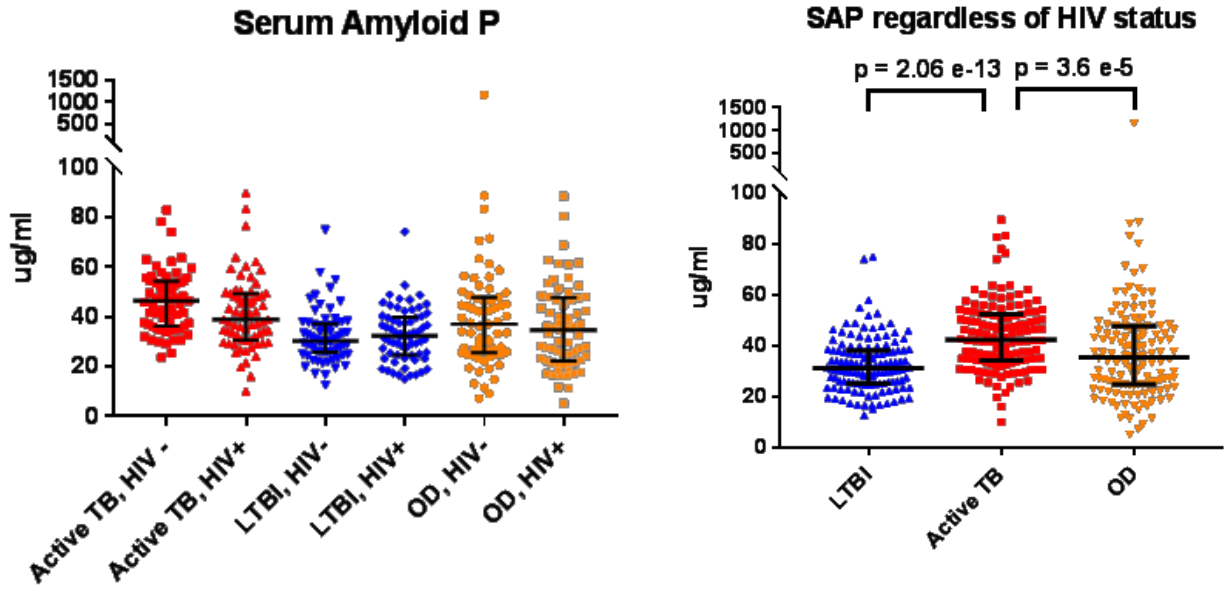


Figure 16 Serum levels of serum amyloid P (SAP) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of haptoglobin were significantly higher in the active TB groups than in both the LTBI and OD groups (Figure 17, below). The levels of this protein appeared higher in TB, HIV- than TB, HIV+, but this difference did not reach statistical significance (p = 0.057). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.64 (95% confidence interval: 0.58 - 0.71).

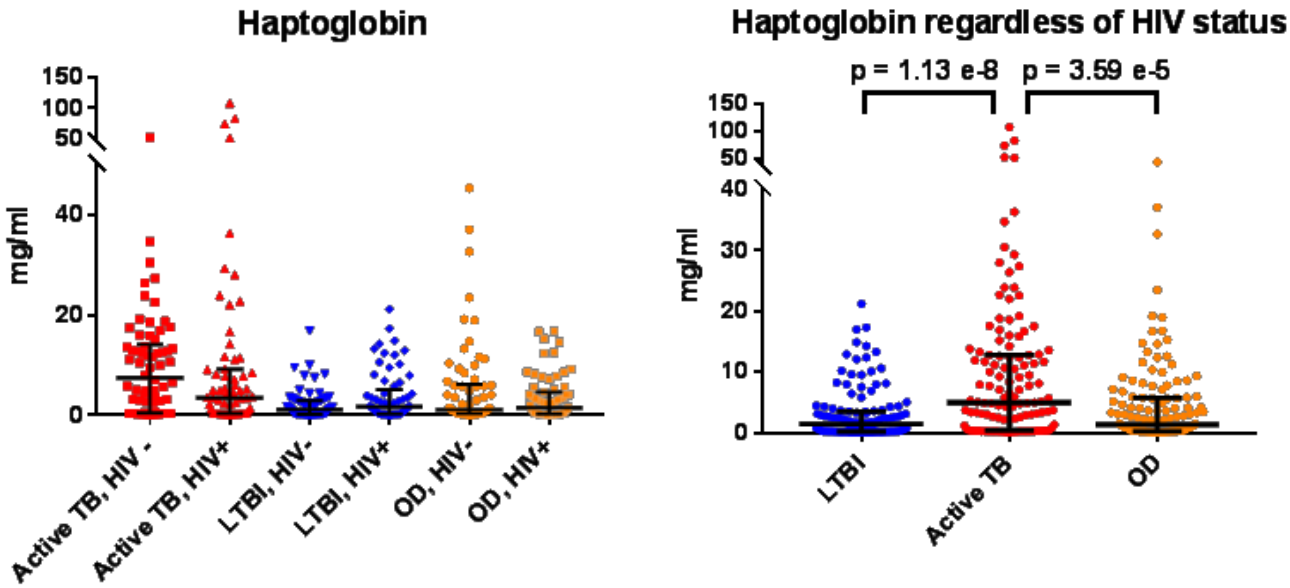


Figure 17 **Serum levels of haptoglobin in each of the clinical groups in the EU-TB study measured by Luminex.**

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of C-reactive protein (CRP) were significantly higher in the active TB groups than the LTBI groups, but not significantly higher than the OD groups (Figure 18, below). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.51 (95% confidence interval: 0.43 - 0.58).

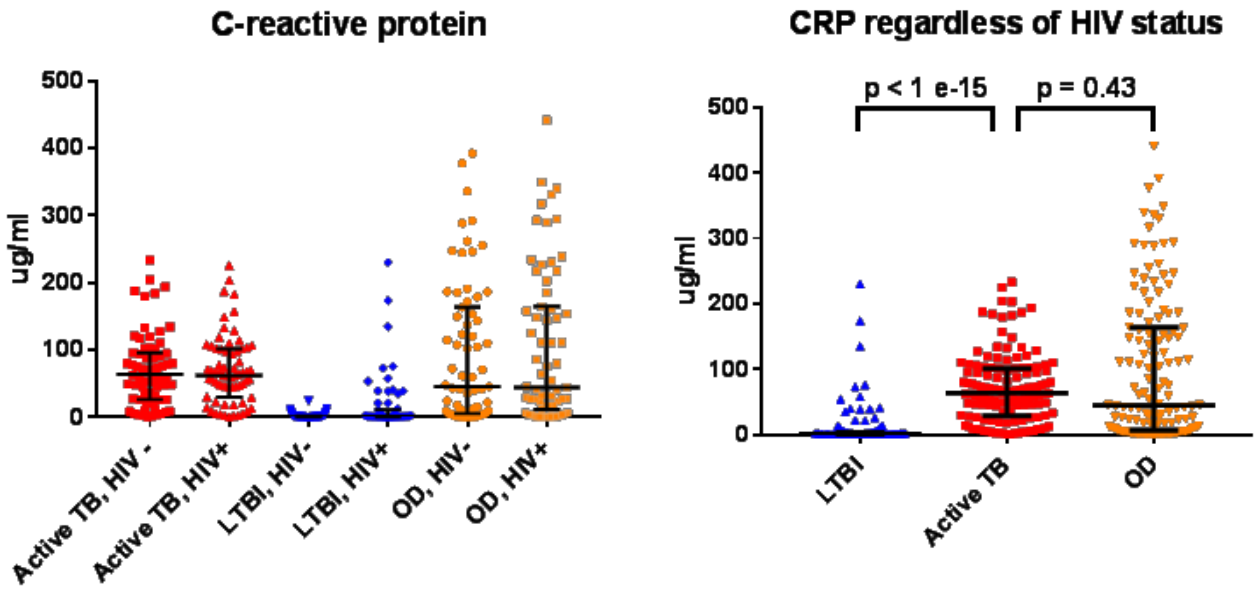


Figure 18 Serum levels of C-reactive protein (CRP) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of matrix metalloproteinase-2 (MMP-2) were significantly lower in the active TB groups than the LTBI groups, but not significantly different from the OD groups (Figure 19, below). Within the active TB and LTBI groups, levels were also significantly lower in patients with HIV infection than in those without (p = 0.01 and 3.68 e-7, respectively). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.52 (95% confidence interval: 0.43 - 0.62).

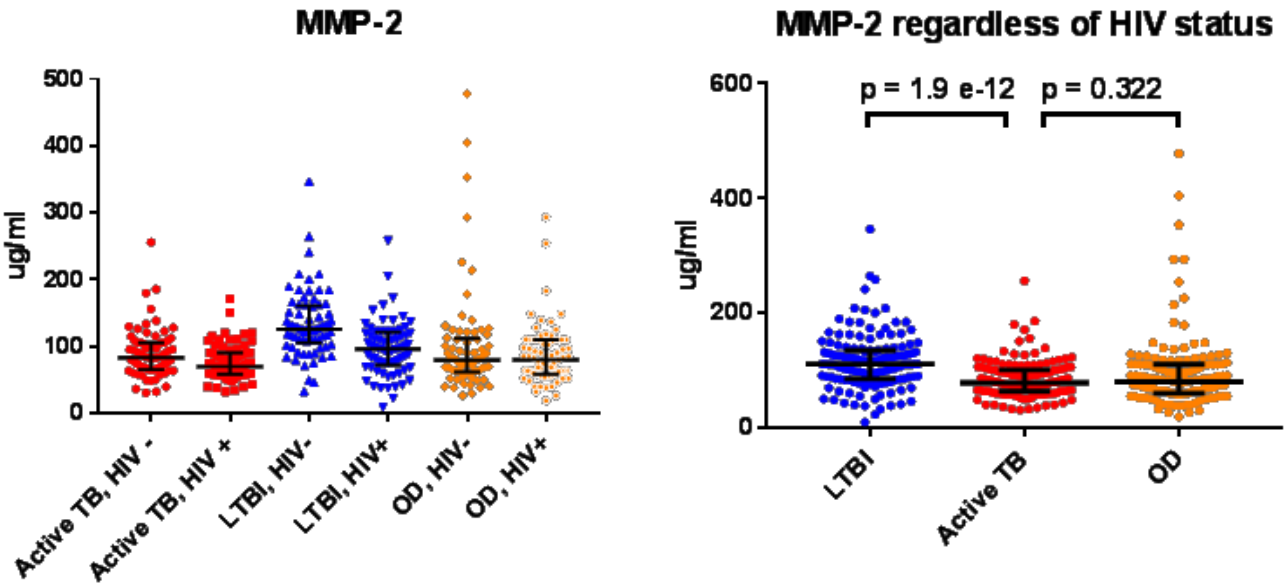


Figure 19 Serum levels of matrix metalloproteinase-2 (MMP-2) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of matrix metalloproteinase-9 (MMP-9) were significantly higher in the active TB groups than the LTBI groups, but not significantly higher than in the OD groups (Figure 20, below). Levels were significantly lower in HIV infected individuals, both across all groups (p = 1.50 e-13) and within the active TB group (p = 9.18 e-10). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.53 (95% confidence interval: 0.46 - 0.60).

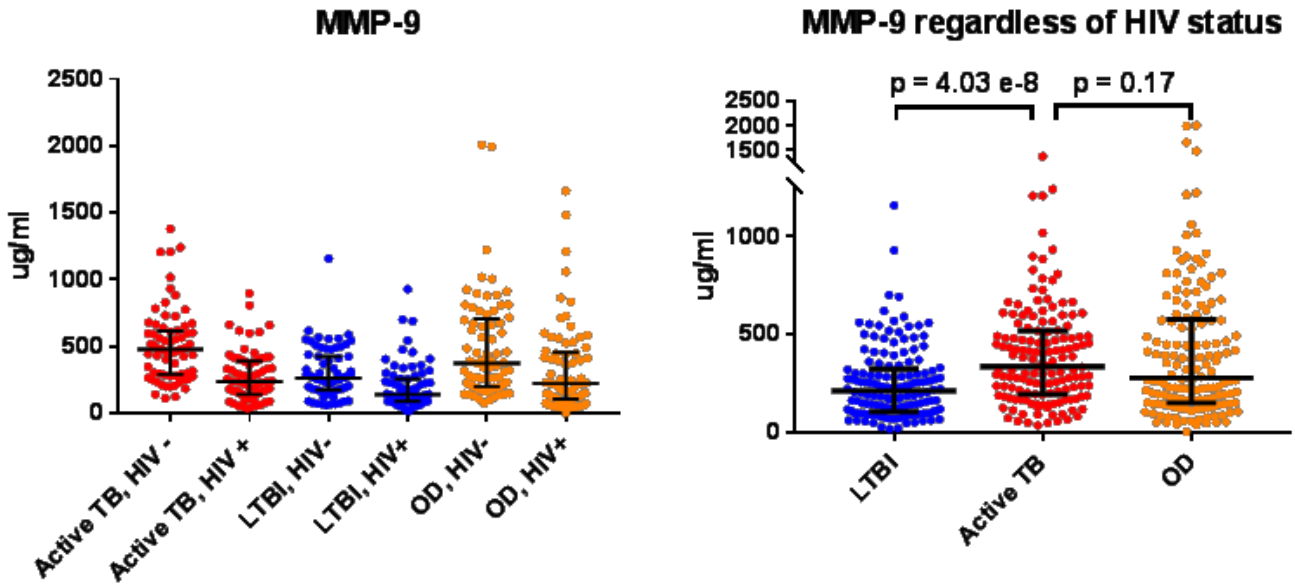


Figure 20 Serum levels of matrix metalloproteinase-9 (MMP-9) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of apolipoprotein A-I (apo-AI) in the active TB groups were significantly lower than in the LTBI groups but were not significantly different from the OD groups (Figure 21, below). The levels of this protein were also significantly lower in the HIV infected patients in the LTBI and OD groups (p = 2.41 e-4), but not in the active TB groups (p = 0.079). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.52 (95% confidence interval: 0.45 - 0.58).

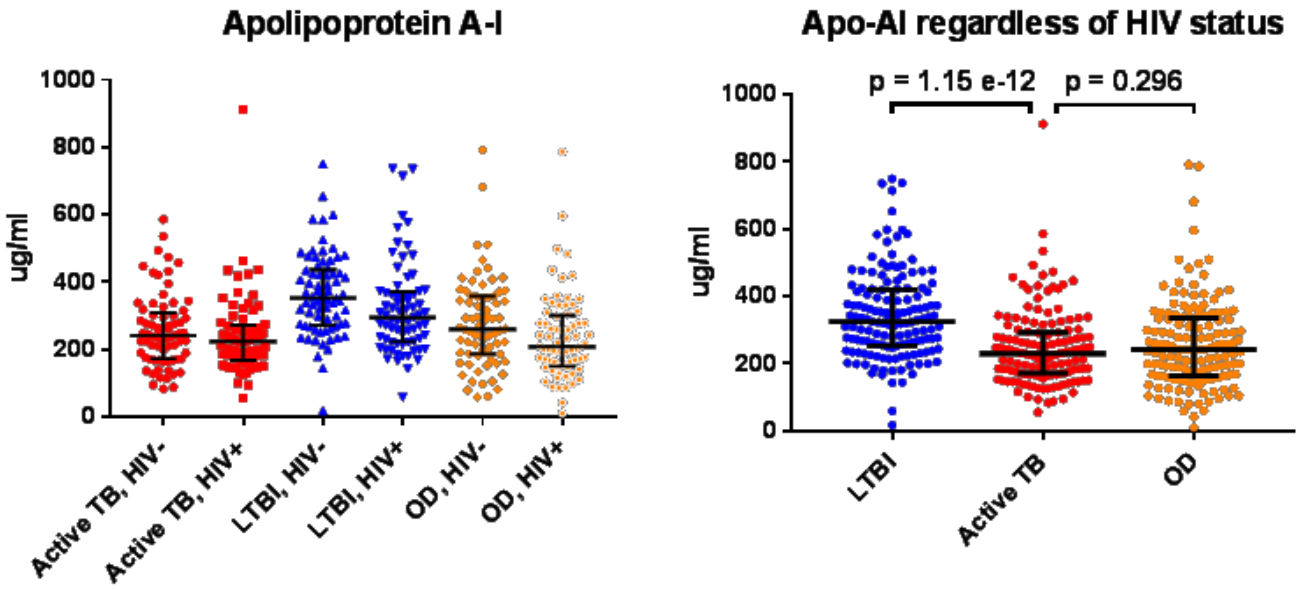


Figure 21 Serum levels of apolipoprotein A-I (apo-AI) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of apolipoprotein C-III (apo-CIII) were significantly lower in the active TB groups than in both the LTBI and OD groups (Figure 22, below). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.58 (95% confidence interval: 0.51 - 0.64).

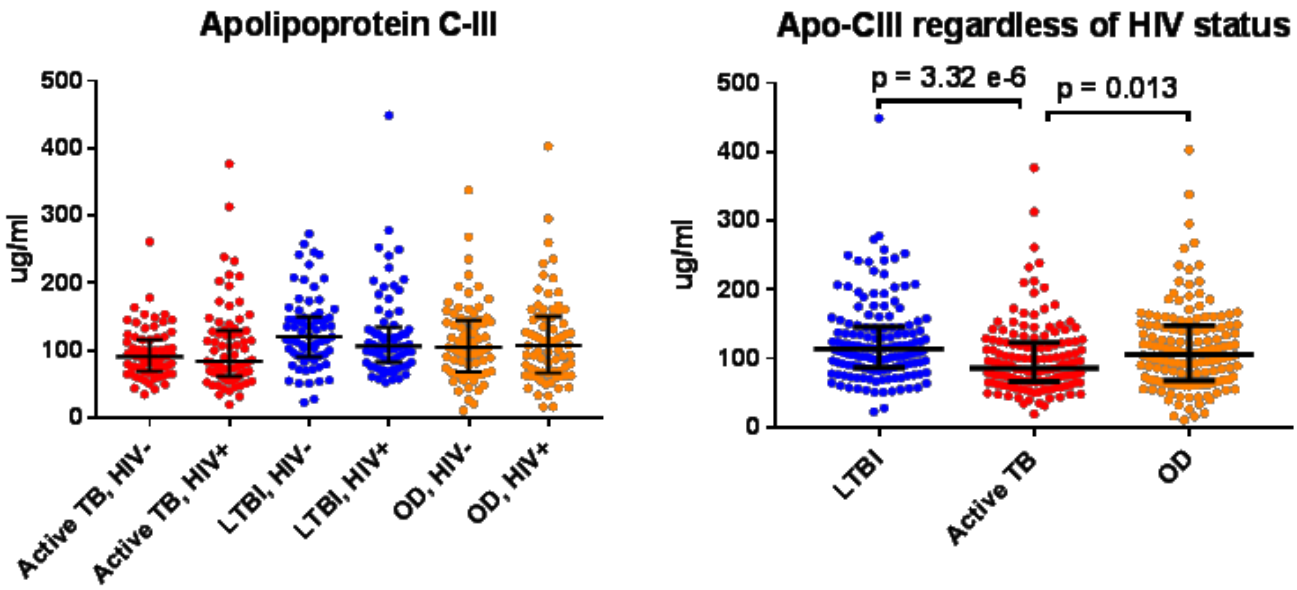


Figure 22 Serum levels of apolipoprotein C-III (apo-CIII) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of transthyretin (TTR) were lower in the active TB groups than in both the LTBI and OD groups (Figure 23, below). The levels of TTR were also significantly lower in individuals infected with HIV in the LTBI and OD groups ($p = 4.21$ e-4), although not significantly so within the active TB group ($p = 0.11$). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.61 (95% confidence interval: 0.55 - 0.68).

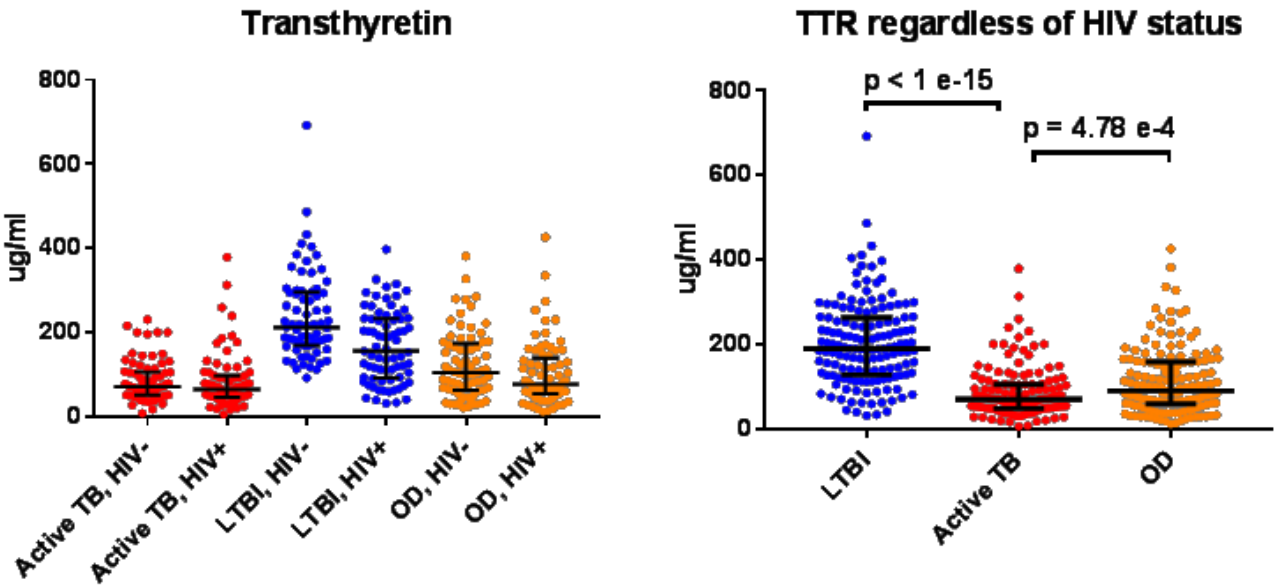


Figure 23 Serum levels of transthyretin (TTR) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of complement factor H (FH) were significantly higher in the active TB groups than in both the LTBI and OD groups (Figure 24). Levels in the OD group trended towards being lower than in the LTBI group (p = 0.08). In individuals infected with HIV, FH levels were significantly lower, both across all groups (p = 7.64 e-4) and within the active TB group (p = 1.98 e-3). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.70 (95% confidence interval: 0.64 - 0.76).

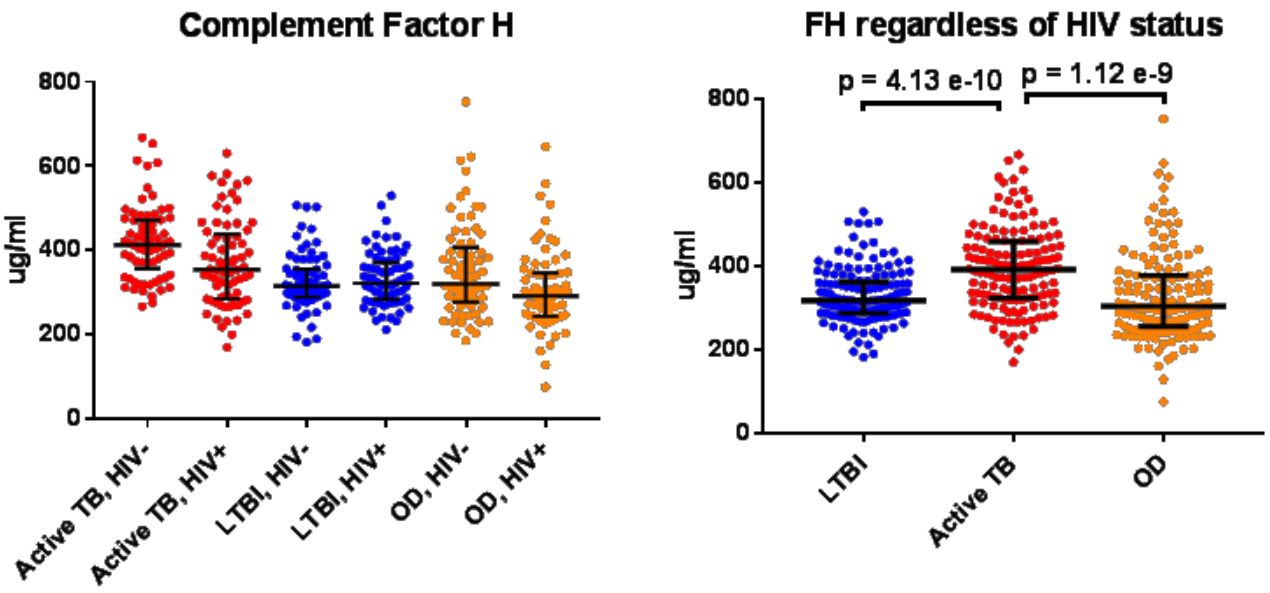


Figure 24 Serum levels of complement factor H (FH) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of interferon gamma (IFN-gamma) were significantly higher in the active TB groups than in both the LTBI and OD groups (Figure 25). In addition, levels were higher in HIV infected patients across the LTBI and OD groups (p = 0.03) and within the active TB group (p = 4.70 e-3). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.60 (95% confidence interval: 0.51 - 0.69).

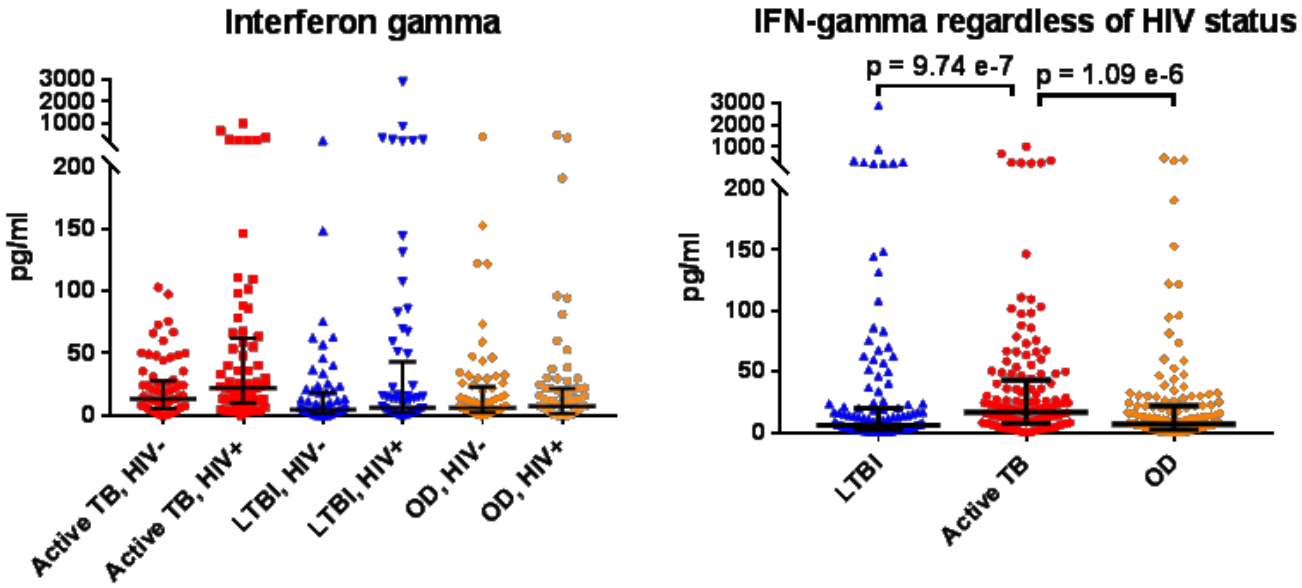


Figure 25 Serum levels of interferon gamma (IFN-gamma) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of interferon gamma-inducible protein-10 (IP-10) were significantly higher in the active TB groups than in both the LTBI and OD groups (Figure 26). In addition, levels were significantly higher in HIV infected individuals with both LTBI and OD ($p < 1$ e-15) and within the active TB group ($p = 2.6$ e-12). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.66 (95% confidence interval: 0.6 - 0.73).

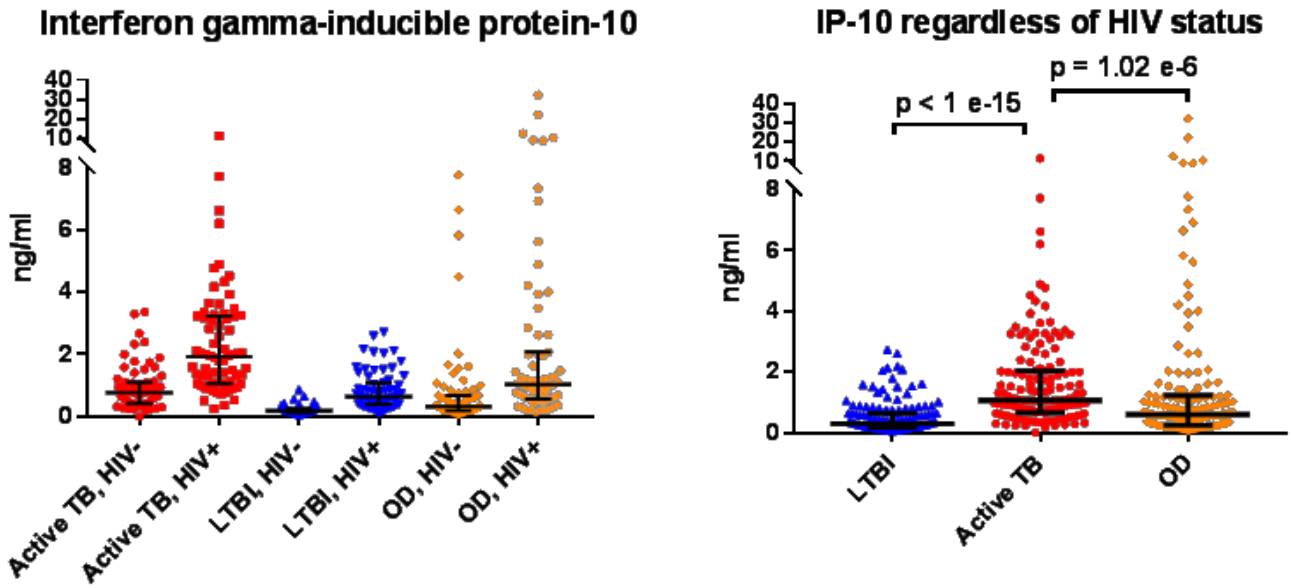


Figure 26 Serum levels of interferon gamma-inducible protein-10 (IP-10) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of interferon alpha-2 (IFN-alpha-2) were significantly higher in the active TB groups than the LTBI groups only (Figure 27). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.52 (95% confidence interval: 0.45 - 0.58). For this analyte, 252 patients in total were below the lower limit of detection (LLD) of the assay. For these patients a value of zero was assigned.

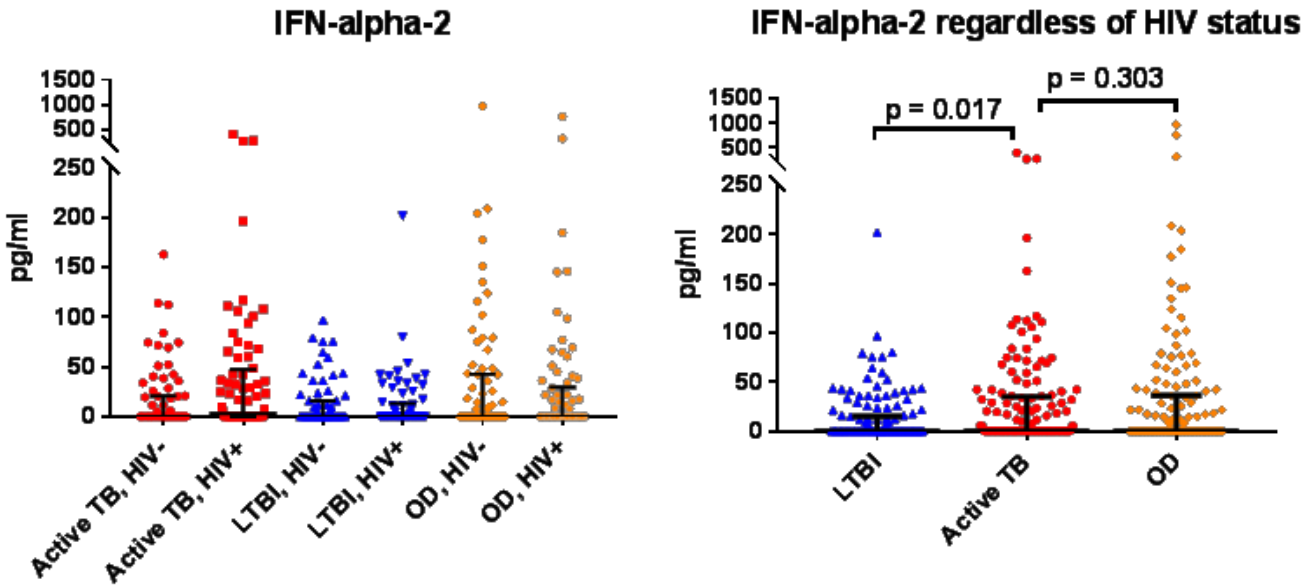


Figure 27 Serum levels of interferon alpha-2 (IFN-alpha-2) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of vascular endothelial growth factor (VEGF) were significantly higher in the active TB groups than in both the LTBI and OD groups (Figure 28). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.64 (95% confidence interval: 0.58 - 0.71).

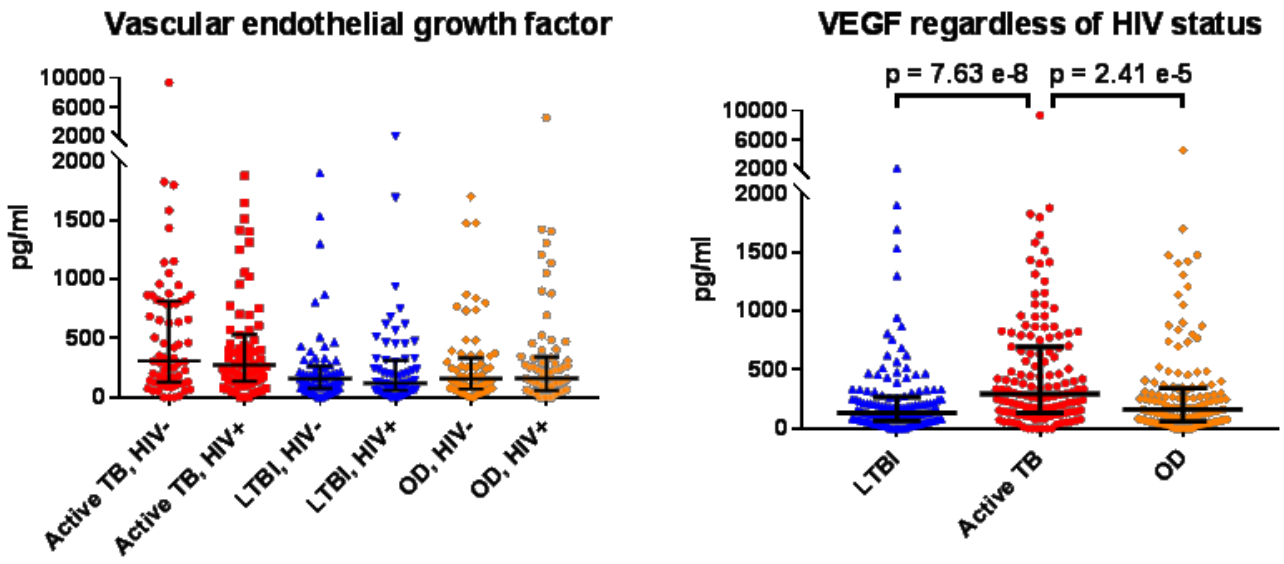


Figure 28 Serum levels of vascular endothelial growth factor (VEGF) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of interleukin-1 receptor antagonist (IL-1 RA) were significantly higher in the active TB groups than in the LTBI groups only (Figure 29). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.51 (95% confidence interval: 0.44 - 0.58). For this analyte, 126 patients in total were below the LLD of the assay: 26 with Active TB; 61 with LTBI; and 39 with OD. For these patients a value of zero was assigned.

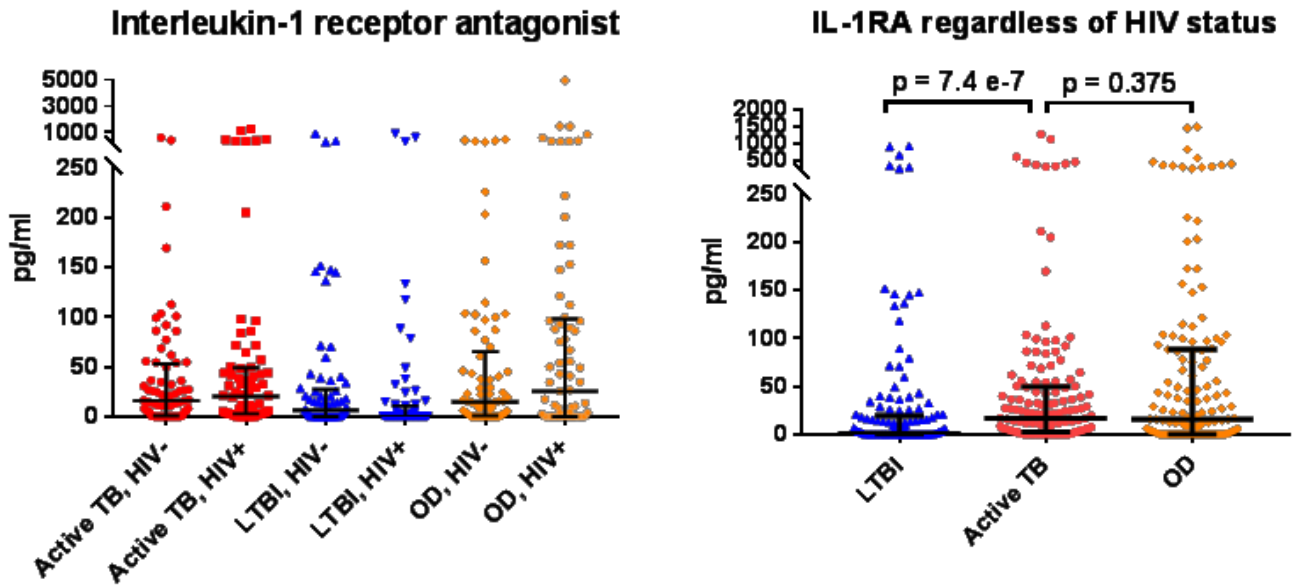


Figure 29 Serum levels of interleukin-1 receptor antagonist (IL-1 RA) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of transforming growth factor alpha (TGF-alpha) were significantly higher in the active TB groups than in both the LTBI and OD groups (Figure 30). In addition, levels were significantly higher in individuals infected with HIV with LTBI and OD (p = 7.36 e-5) and within the active TB group (p = 1.76 e-4). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.56 (95% confidence interval: 0.49 - 0.63).

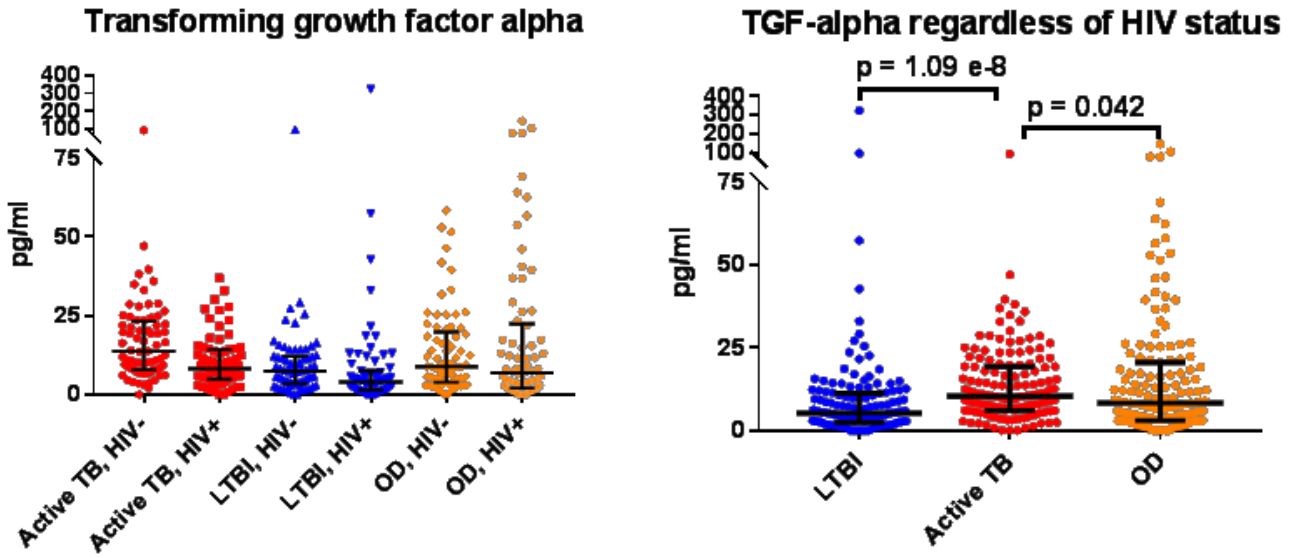


Figure 30 Serum levels of transforming growth factor alpha (TGF-alpha) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

Levels of tumour necrosis factor alpha (TNF-alpha) were significantly higher in the active TB groups than in the LTBI groups only (Figure 31). In addition, levels were significantly higher in individuals infected with HIV, both across LTBI and OD groups (p = 3.82 e-12) and within the active TB group (p = 3.72 e-5). The ROC AUC to distinguish active TB from OD, regardless of HIV status, was 0.53 (95% confidence interval: 0.47 - 0.6).

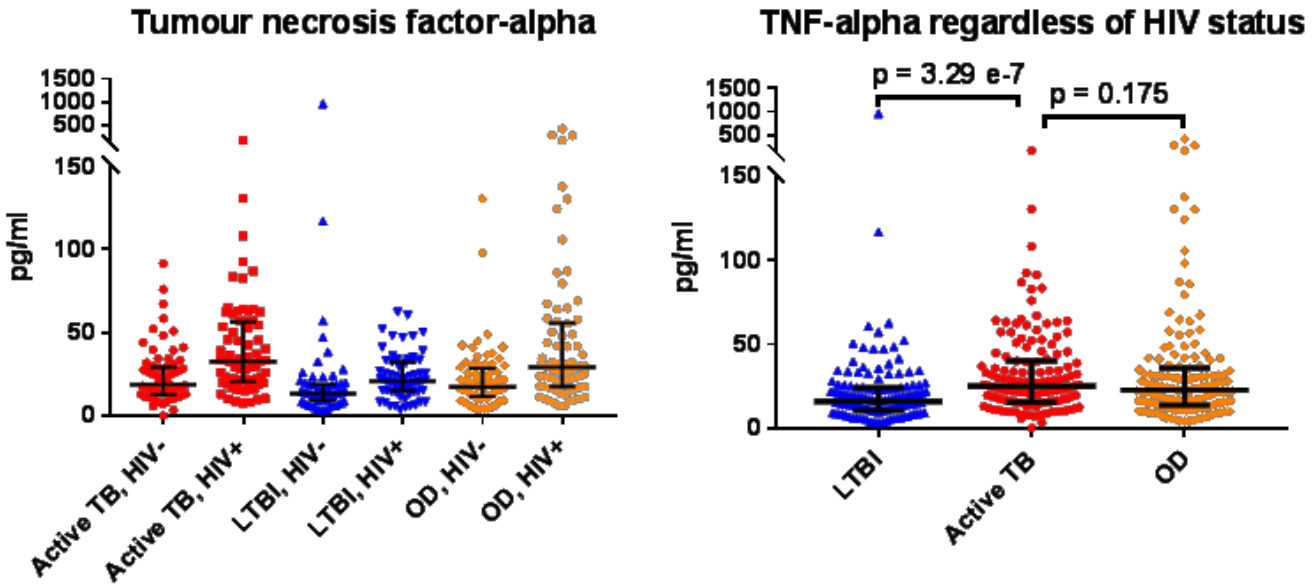


Figure 31 Serum levels of tumour necrosis factor alpha (TNF-alpha) in each of the clinical groups in the EU-TB study measured by Luminex.

The left panel shows protein levels from all six clinical groups. The right panel reports combined clinical groups regardless of HIV status. Error bars represent medians and interquartile ranges (IQRs). p-values are one-sided and derived from Mann-Whitney U tests.

The medians and interquartile ranges of each protein in each of the TB, OD and LTBI groups are shown below in Table 10.

The ROC AUCs of each protein individually to distinguish TB from OD are shown in Figure 32, below. Twelve proteins had ROC AUCs with 95% confidence intervals greater than 50%.

The best performing biomarker individually was complement factor H (ROC AUC 70% (64% - 76%). The performance of this protein was preserved across both HIV status (71% in HIV uninfected, 69% in HIV infected) and study site (70% in Cape Town, 71% in Karonga). The corresponding ROC curves are shown below in Figure 33.

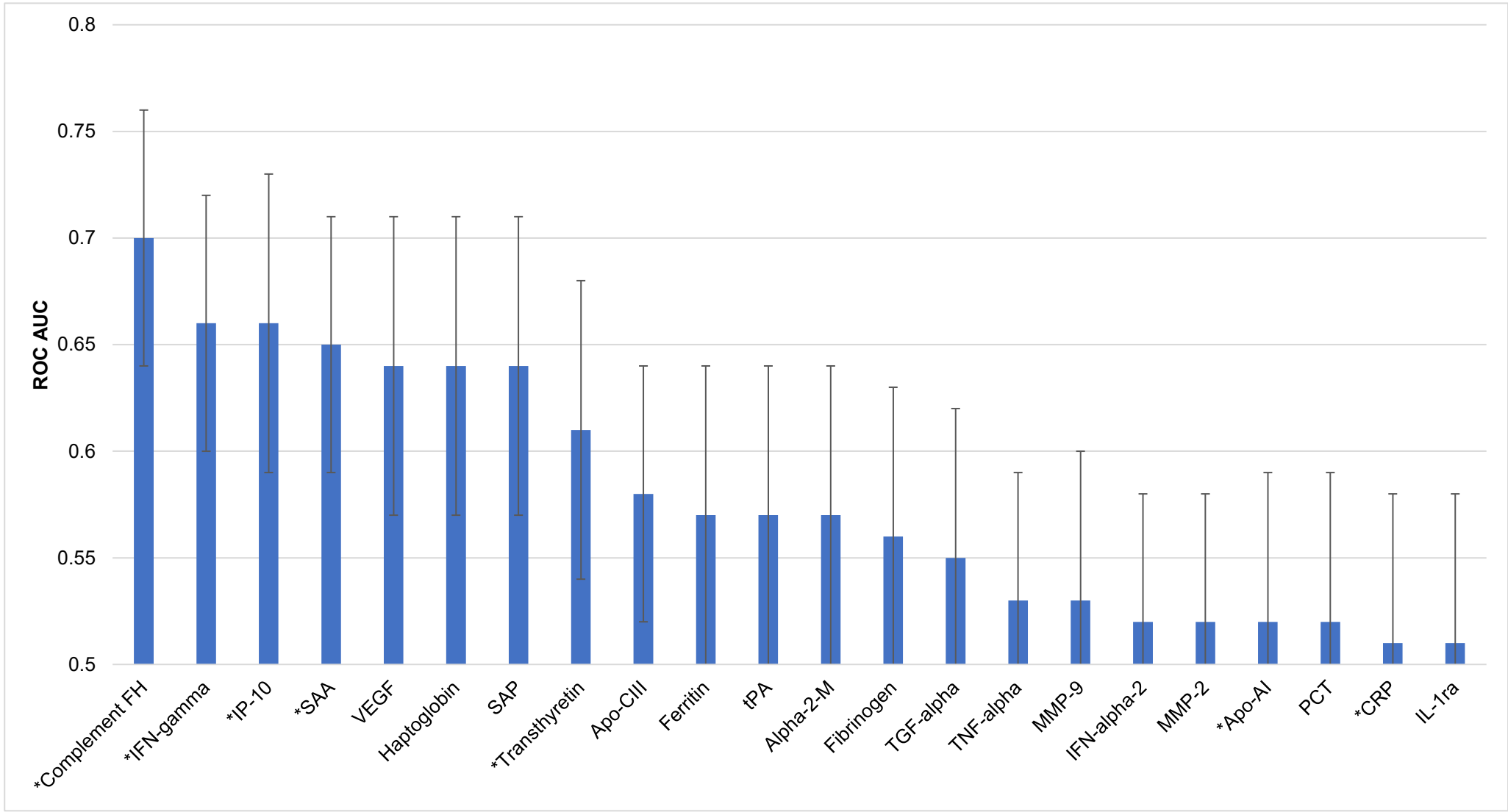


Figure 32 Areas under the receiver operating characteristic curves (ROC AUCs) of each protein to distinguish TB from OD in the EU-TB cohort.

Error bars represent 95% confidence intervals. Asterisks indicate proteins belonging to the seven-protein diagnostic signature from the AE-TBC study.(1)

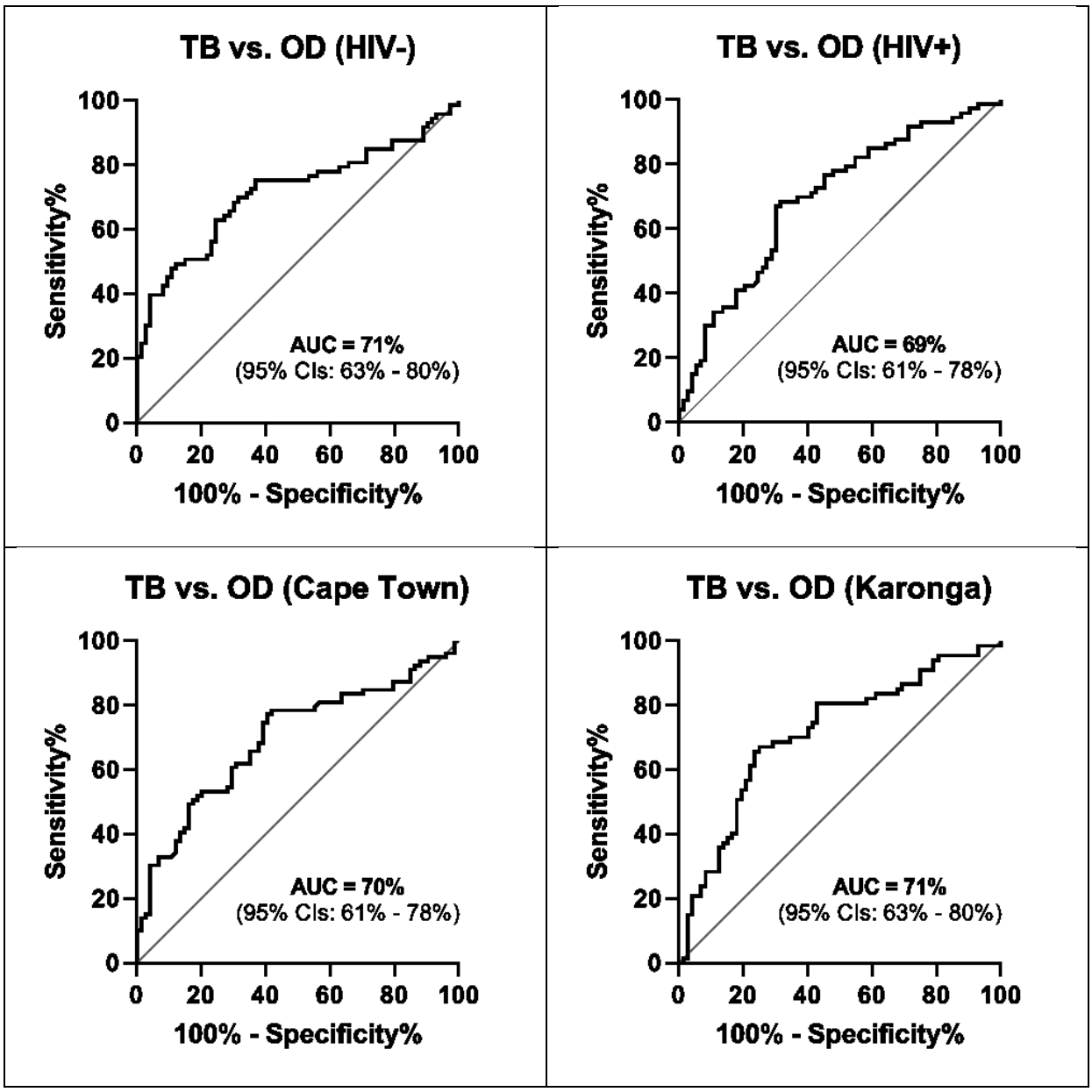


Figure 33 ROC curves for the accuracy of complement factor H after stratification by each of HIV status and site.

All patients with results for complement FH were included: 146 with TB and 146 with OD. HIV-: HIV uninfected; HIV+: HIV infected.

The ROC AUCs for each protein after stratifying for each of HIV status and site are now shown (Figures 34 and 35). With regards to ROC AUCs by HIV status, 16 proteins performed better in HIV uninfected (HIV-) patients: complement FH, IP-10, SAA, VEGF, haptoglobin, SAP, transthyretin, apo-CIII, ferritin, alpha-2-M, TGF-alpha, TNF-alpha, MMP-9, apo-AI, PCT, and CRP. Five proteins performed better in HIV co-infected (HIV+) patients: IFN-gamma, fibrinogen, IFN-alpha-2, MMP-2, and IL-1ra. However, confidence intervals overlapped for every protein (Figure 34). With regards to ROC AUCs by site, seven proteins performed better at Karonga (complement factor H, serum amyloid A, transthyretin, apo-CIII, ferritin, apo-AI and IL-1ra (Figure 35). 15 proteins performed better at Cape Town (IFN-gamma, IP-10, VEGF, haptoglobin, serum amyloid P, tPA, alpha-2-macroglobulin, fibrinogen, TGF-alpha, TNF-alpha, MMP-9, IFN-alpha-2, MMP-2, procalcitonin and CRP). As with HIV status, however, 95% confidence intervals overlapped for every protein (Figure 35).

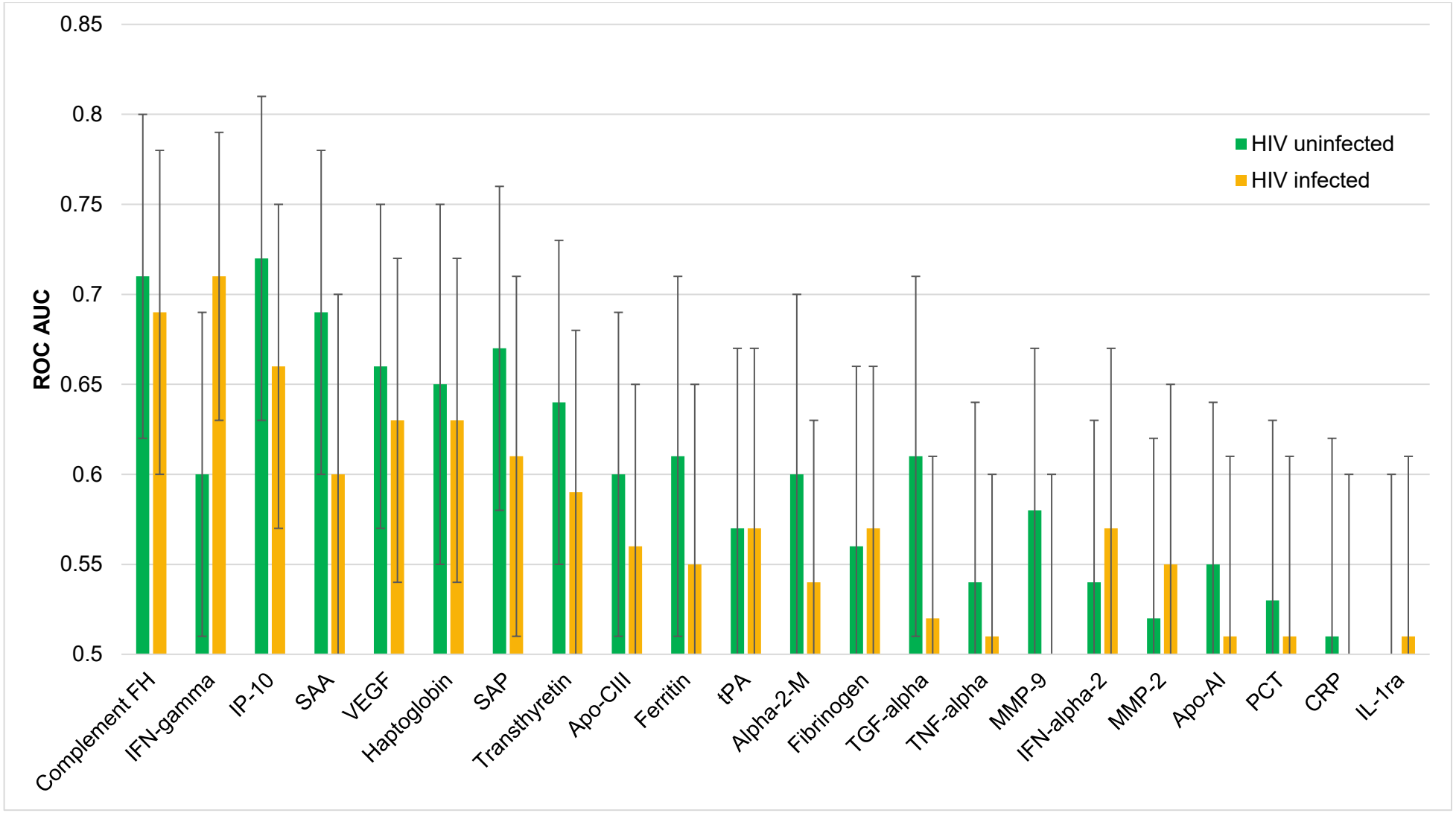


Figure 34 **Individual ROC AUCs for each of HIV uninfected and infected halves of the cohort.** Green bars: distinction of TB, HIV- from OD, HIV-; yellow bars: TB, HIV+ from OD, HIV+.

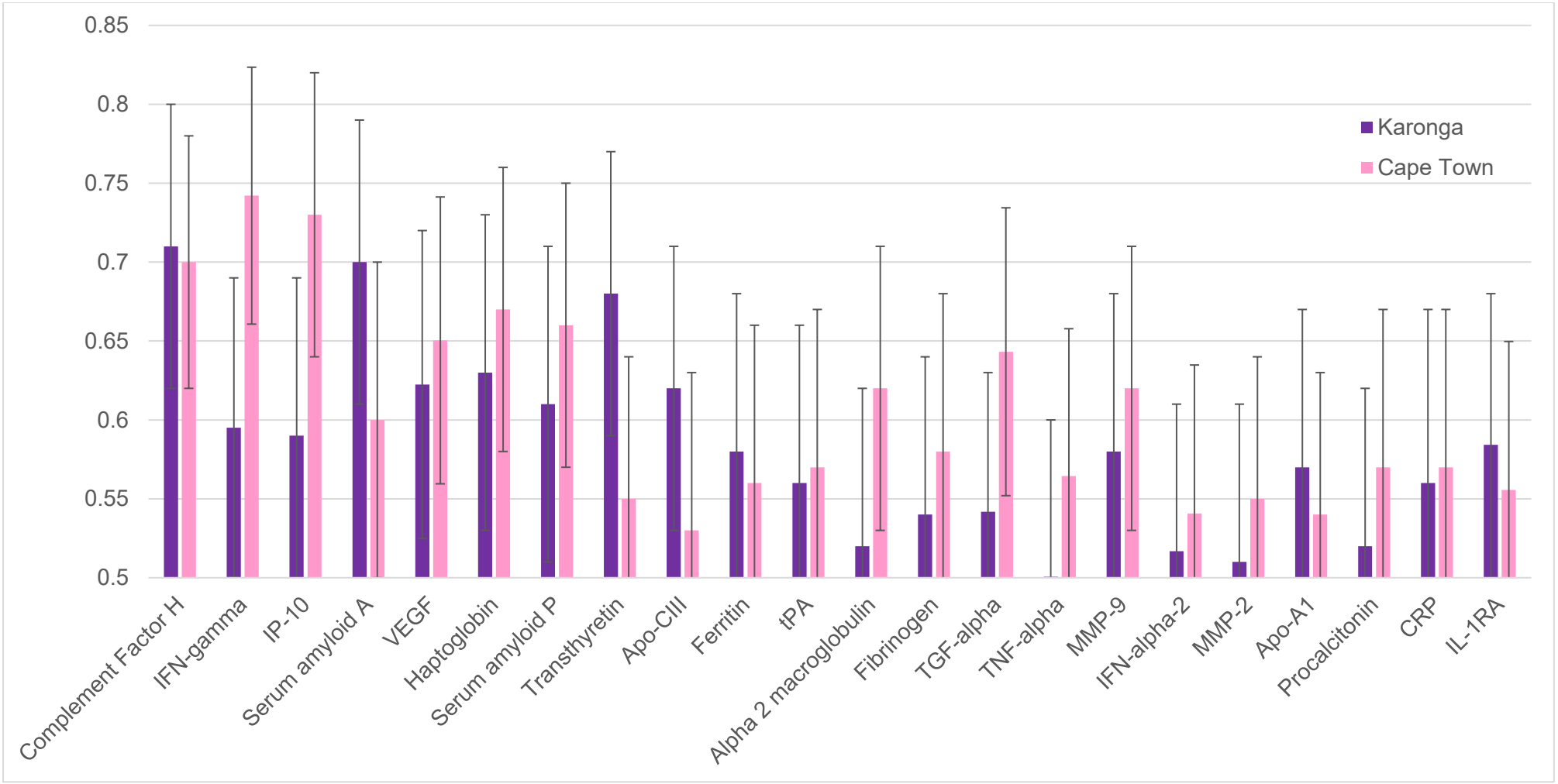


Figure 35 Individual ROC AUCs for each protein at each of the Karonga and Cape Town sites. Purple bars: distinction of TB, Karonga from OD, Karonga; Pink bars: distinction of TB, Cape Town from OD, Cape Town.

ROC AUCs for each protein tested in the EU-TB cohort were compared to the corresponding results from the AE-TBC study in Table 11 and Figure 35. In Figure 35, proteins are listed in order of the size of the difference in performance between the two studies. As shown in Table 10, a different number of patients had results on each individual protein owing to variation in the number of samples that needed to be re-run in each Luminex panel. The numbers of TB and OD patients with results for each protein are shown in Table 11.

Overall, 14 proteins performed significantly less well than in the AE-TBC study, as judged by non-overlapping confidence intervals (Figure 35). Four appeared to perform better, however, with one protein (complement factor H) performing significantly better (confidence intervals non-overlapping). Of the seven proteins in the AE-TBC signature, five had ROC AUCs with 95% confidence intervals greater than 50% (complement FH, IP-10, IFN-gamma, serum amyloid A and transthyretin), however two did not (apo-AI and CRP).

Protein	Number tested from EU-TB cohort	ROC AUC in EU-TB study (%)	ROC AUC in AE-TBC study(1) (%)
Complement FH	292	70 (64 – 76)	58 (53 – 62)
IP-10	282	66 (60 – 73)	82 (79 – 86)
IFN-gamma	284	66 (60 – 72)	80 (76 – 84)
SAA	263	65 (58 – 71)	83 (80 – 86)
VEGF	278	64 (57 – 71)	70 (65 – 74)
Haptoglobin	263	64 (58 – 71)	62 (57 – 66)
SAP	267	64 (57 – 71)	58 (53 – 63)
Transthyretin	292	61 (55 – 68)	78 (74 – 82)
Apo-CIII	292	58 (51 – 64)	65 (61 – 70)
Ferritin	263	57 (50 – 64)	78 (75 – 82)
tPA	263	57 (50 – 64)	72 (68 – 76)
Alpha-2-M	267	57 (50 – 64)	54 (49 – 58)
Fibrinogen	263	56 (49 – 63)	73 (69 – 77)
TGF-alpha	284	55 (49 – 62)	73 (69 – 77)
TNF-alpha	284	53 (46 – 59)	69 (65 – 74)
MMP-9	292	53 (47 – 60)	59 (53 – 64)
Apo-AI	292	52 (45 – 59)	69 (65 – 73)
Procalcitonin	263	52 (45 – 59)	68 (63 – 72)
IFN-alpha-2	284	52 (45 – 58)	67 (62 – 71)
MMP-2	292	52 (45 – 58)	54 (49 – 59)
CRP	267	51 (43 – 58)	84 (81 – 87)
IL-1ra	283	51 (44 – 58)	63 (58 – 68)

Table 11 Diagnostic accuracy of protein biomarkers individually.

ROC AUCs for the performance of each protein to distinguish TB (n = 146) from OD (n = 146) are shown, regardless of HIV status or site. Proteins are listed in descending order of performance in the EU-TB cohort with the numbers of patients for which results were obtained for that protein. ROC AUCs are shown as percentages. Results from the AE-TBC study are shown to the right for comparison.(1) Bracketed values indicate 95% confidence intervals.

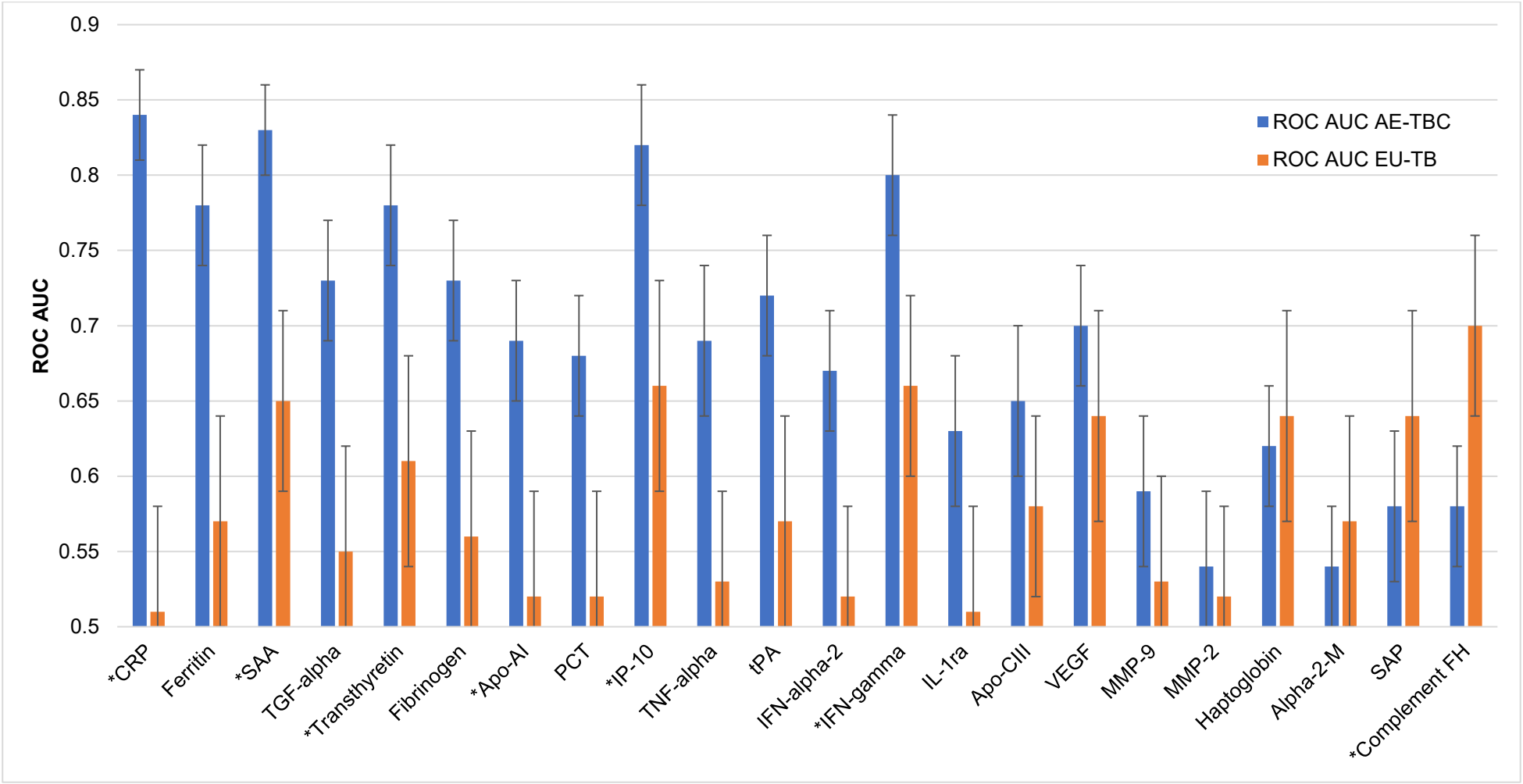


Figure 36 ROC AUCs of each protein to distinguish TB from OD in order of the size of the difference in performance between the two studies.

Orange bars: performance in ILULU-TB cohort; blue bars: performance in AE-TBC cohort. Error bars represent 95% confidence intervals. Asterisks indicate proteins belonging to the AE-TBC 7-protein signature.

3. ROC AUCs of individual proteins to distinguish TB from LTBI

As shown below in Table 12, the best performing biomarker to distinguish TB from LTBI (irrespective of HIV status or site) was CRP (ROC AUC 92% (95% CI 88% - 95%). In total, 20 of the 22 proteins had ROC AUCs for this purpose whose 95% confidence intervals were greater than 50%. Those that did not were IFN-alpha-2 and alpha-2-M.

	ROC AUC (95% CI)	Number of patients tested
CRP	92 (88-95)	267
Ferritin	88 (84-92)	265
SAA	87 (83-92)	265
Transthyretin	86 (82-90)	292
IP-10	83 (78-87)	279
tPA	79 (74-85)	265
SAP	75 (69-81)	267
Fibrinogen	75 (69-80)	264
Apo-AI	73 (68-79)	292
MMP-2	73 (67-79)	292
Complement FH	70 (64-76)	292
Haptoglobin	69 (63-76)	262
PCT	69 (63-75)	265
TGF-alpha	69 (63-75)	281
VEGF	68 (61-74)	277
MMP-9	68 (62-74)	292
TNF-alpha	67 (60-73)	281
IFN-gamma	66 (59-72)	281
IL-1ra	66 (59-72)	281
Apo-CIII	65 (58-71)	292
IFN-alpha-2	56 (49-63)	282
Alpha-2-M	51 (44-58)	267

Table 12 ROC AUCs for each protein to distinguish active TB from LTBI.

Results are shown regardless of HIV status or site. Proteins are listed in descending order of diagnostic performance. The number of patients with both TB and LTBI that had results for each protein are shown in the right-hand column.

4. Results of analyses of biosignatures to distinguish TB from OD

For the biosignature discovery analyses, only TB and OD patients were included, and only those for which data on all 22 proteins were collected (n = 249). If protein levels fell above the upper limit of detection of the Luminex assay, they were re-run at a higher dilution, however the number of Luminex kits (plates) purchased for each panel of proteins was planned at the start of the study. Further plates were not purchased at the end of assay series because this could have led to lot-lot variation and would've exceeded the study budget. This number was therefore less than the number of TB and OD patients from which individual ROC AUCs were calculated for each protein (n 263 - 292). Assuming the AE-TBC signature had the same accuracy in our data, however, with 249 patients (122 TB and 127 OD), we had 95% power to show a sensitivity of greater than 90% and specificity of greater than 66.5% with any of the new signatures. The number of TB and OD patients that were included is shown in Table 13, stratified by site and HIV status.

	Cape Town	Karonga	Total
TB, HIV-	30	32	**62**
TB, HIV+	29	31	**60**
OD, HIV-	31	31	**62**
OD, HIV+	37	28	**65**
Total	**127**	**122**	**249**

Table 13 Numbers of patients with TB and OD that were included in the biosignature analyses, stratified by HIV status and site.

Only patients for which data on all 22 proteins was gathered were included.

Results from each of the biosignature discovery approaches are now shown in turn.

i) Application of the AE-TBC 7-protein signature to the EU-TB cohort

As described in Materials and Methods, for this experiment, the whole of the AE-TBC cohort was used as the discovery cohort (n = 701), and the EU-TB cohort as the validation set (n = 249). This was on the advice of Professor Martin Kidd, Centre for Statistical Consultation at Stellenbosch University, who performed this analysis as well as the original analyses in the AE-TBC study.(1) The signature that emerged from this larger discovery cohort comprised the same seven proteins (apo-AI, transthyretin, CRP, IFN-gamma, IP-10, serum amyloid A and complement factor H) and had a ROC AUC of 91% (95% CI 89% - 93%). This was the same as in the AE-TBC study itself, when the training set comprised 491 patients, although with a slightly tighter confidence interval. As shown below in Table 14, in the EU-TB cohort (validation set), sensitivity was higher than in the AE-TBC study, but specificity was markedly reduced (98% (95% CI: 94 – 100%) and 12% (7% - 19%) respectively).

	AE-TBC cohort (discovery set)	**ILULU-TB cohort (validation set)**
ROC AUC	**91 (89 – 93)**	
Sensitivity	85 (80 – 90)	98 (94 – 100)
Specificity	85 (81 – 88)	12 (7 – 19)
PPV	71 (65 - 76)	52 (45 - 58)
NPV	93 (90 - 95)	88 (64 - 99)

Table 14 Performance of the seven-protein signature from the AE-TBC study in the ILULU-TB cohort.

The entire AE-TBC cohort was used as the discovery set (n = 701), and the whole sample of the ILULU-TB cohort as the validation set (n = 249). Results are shown with the threshold for defining a case of TB at the default setting (p > 0.5). A ROC AUC in the validation set was not generated. All results are given as percentages. PPV: positive predictive value. NPV: negative predictive value. Bracketed values represent 95% confidence intervals.

To investigate possible reasons for this poor performance, we then compared the levels of each protein in the TB groups of the two cohorts (AE-TBC vs EU-TB). Levels were plotted as medians and interquartile ranges, which were available from the AE-TBC publication.(1) Results are shown below in Figure 37. The interquartile ranges (IQRs) of most proteins overlapped between studies, however in the case of five proteins, they did not. These were apo-AI, complement factor H, apo-CIII, transthyretin and procalcitonin. Levels of these five proteins were lower in the EU-TB dataset in every case. In the case of apo-AI, levels were approximately ten-fold lower in the EU-TB dataset (median 230.81 vs. 2000 ug/ml).(1)

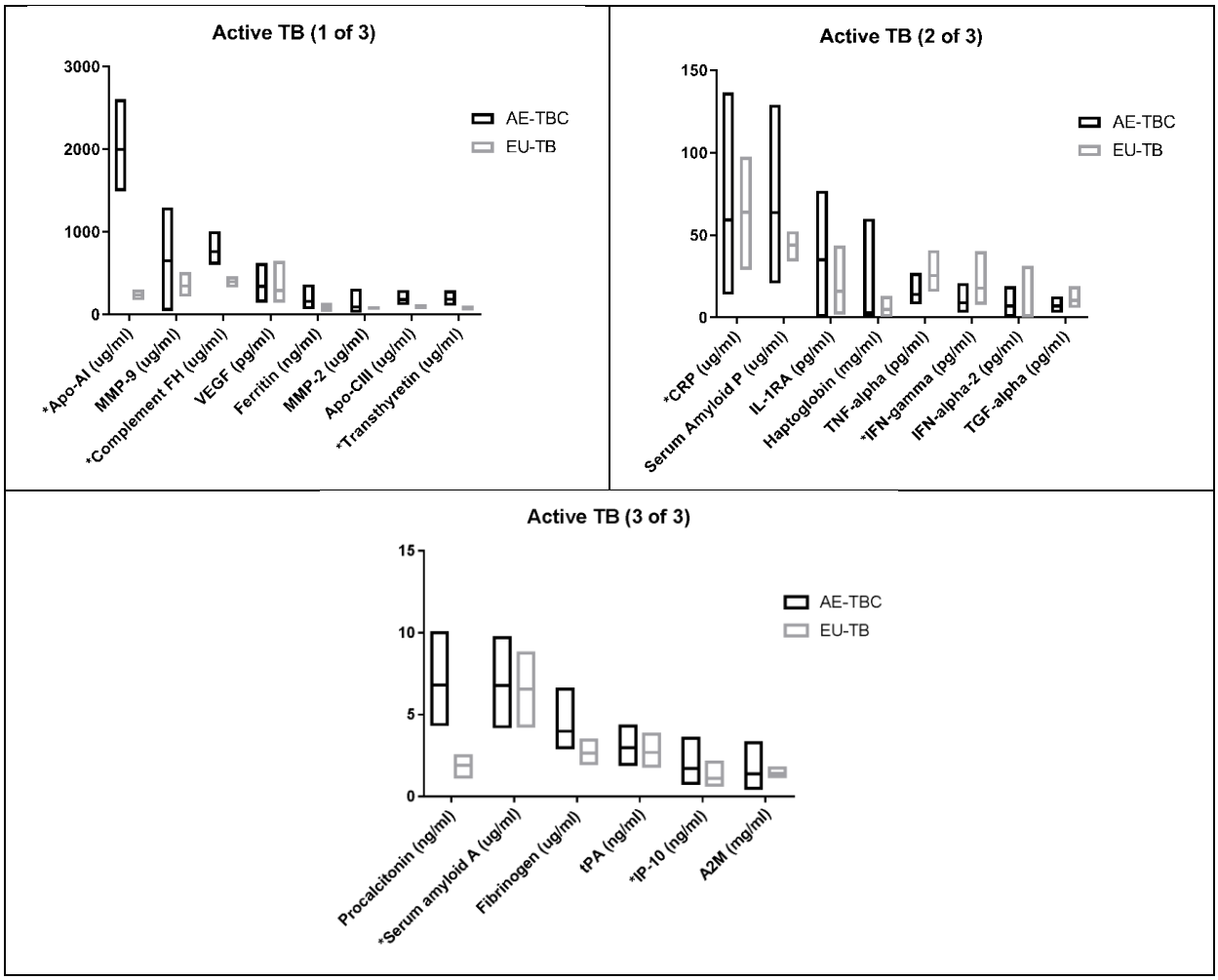


Figure 37 Levels of each of the 22 proteins from the AE-TBC study in the TB groups of the two cohorts (AE-TBC and EU-TB). Units are shown after each protein name. Results are split over three graphs to accommodate protein ranges. Lines represent medians and boxes interquartile ranges Asterisks indicate proteins that belong to the AE-TBC seven-protein signature.

The Luminex panels that these proteins were part of were as follows: apo-AI, complement factor H, apo-CIII and transthyretin were the four proteins that comprised the “Neurodegenerative” panel; procalcitonin was one of the five proteins that comprised the “5-plex Acute Phase Protein” panel.

To explore the reasons for these differences, levels of each of these five proteins in our healthy LTBI control group were compared with normal ranges in the literature.(2-6) As shown below in Table 15, levels of apo-CIII in our LTBI group matched those in the literature,(4) however levels of apo-AI, complement FH and transthyretin were lower than in the literature.(2, 3, 5) Of those, the difference was greatest for apo-AI, which was three or four-fold lower than in the literature. The median level of apo-AI in the TB group of the AE-TBC study, however, was 2000 ug/ml (IQR 1494 – 2604), which is almost two-fold higher than the normal range in the literature, even though this protein is a “negative

acute phase reactant" whose levels tend to drop in inflammation and disease (including TB, in our cohort).(7) Levels of procalcitonin (5-plex Acute Phase panel) were higher in our LTBI group than in the normal range in the literature.(6)

	Median (IQR) in LTBI group; units	**Normal levels (in equivalent units)**
Apo-AI (ug/ml)	324.4 (253.9 – 418.6)	Mean 1180 (95% CI 1170 - 1180)(2)
Complement factor H (ug/ml)	317.5 (287.5 – 359.9)	Mean 400 +/- 62 (SE)(3)
Apo-CIII (ug/ml)	113.9 (87.5 – 145.4)	Median 114 (IQR 79 – 153)(4)
Transthyretin (ug/ml)	189.8 (129.3 – 260.9)	Mean 292 +/- 10.8 (SE)(5)
Procalcitonin (ng/ml)	1.2 (0.73 – 1.79)	Mean 0.15 +/- 0.016 (SE)(6)

Table 15 Comparison of levels of apo-AI, complement FH, apo-CIII, transthyretin and procalcitonin in the LTBI group of the EU-TB cohort with normal ranges taken from the literature. Levels of thee five proteins were significantly lower in the TB group of the EU-TB cohort compared to the TB group of the AE-TBC cohort. IQR: interquartile range. 95% CI: 95% confidence interval. SE: standard error.

Because of these differences between the two datasets, and especially since apo-AI, complement factor H and transthyretin were all part of the AE-TBC seven-protein signature, we decided not to proceed further with this approach to re-testing the AE-TBC biomarkers as set out in Primary Aim 1 part i).

ii) *Discovery of the best new signature in the EU-TB cohort;*

For this aim, as described in the Methods chapter, section 3.3.2.2, parallel discovery efforts were made by Professor Kidd at Stellenbosch University, and Dr Hoggart at Imperial College London. First, patients (n = 249) were randomised to training and test sets (ratio 70:30). Following this, the same General Discriminate Analysis (GDA) and Random Forests approaches to signature discovery were performed, having log-transformed the data as in the AE-TBC study. Using the same allocations to training and test sets, Dr Hoggart performed an independent variable selection experiment using Parallel Regularised Regression Model Search method (PReMS), having normalised the data into deciles.

Results from each of these approaches are now shown in Tables 16 – 19, below. In each case, results are shown with the cut-off for defining a positive test result at the default setting ($p > 0.5$), and after adjusting this cut-off to increase each of the sensitivity and specificity in turn to 90%. This was to assess test performance as each of a rule-out and rule-in test, respectively. The target of 90% sensitivity was chosen as this is the minimum criterion for a screening test specified by the WHO.(8) No specific criteria for a rule-in test have been set, hence the same target of 90% was made for specificity also. At each of these three cut-offs, results are expressed in terms of sensitivity and specificity, and positive and negative predictive value (PPV and NPV). The signature results are shown in the following order: GDA, Random Forests and PReMS.

As shown below in Table 16, the GDA method yielded a 5-protein signature comprising complement FH, IP-10, CRP, serum amyloid A and transthyretin. The ROC AUC in the training set was 84% (95% CI 78% - 90%). Sensitivities and specificities in the test set were 81% and 63% initially, 79% and 41% after increasing sensitivity, and 58% and 89% after increasing specificity.

	Training set	Test set
ROC AUC	**84 (78 – 90)**	
Sensitivity	82 (73 – 90)	81 (65 – 92)
Specificity	74 (64 – 83)	63 (46 – 78)
PPV	75 (65 – 84)	68 (52 – 81)
NPV	81 (71 - 89)	77 (59 – 90)
After adjustment of cut-off for increased sensitivity		
Sensitivity	89 (80 - 94)	79 (63 - 90)
Specificity	48 (37 - 59)	41 (25 - 58)
PPV	64 (55 - 73)	58 (43 – 71)
NPV	80 (67 – 90)	65 (43 – 84)
After adjustment of cut-off for increased specificity		
Sensitivity	61 (50 - 71)	58 (41 - 74)
Specificity	91 (82 - 96)	89 (75 - 97)
PPV	87 (76 – 94)	85 (65 – 96)
NPV	69 (59 – 77)	67 (52 – 80)

Table 16 Performance of a new five-protein signature derived from Generalised Discriminant Analyses (GDA).

This analysis was performed using data on the ILULU-TB cohort only (n = 249). Patients were randomly assigned to training and test sets at a ratio of 70:30. The signature comprised Complement FH, IP-10, CRP, SAA, and Transthyretin. A ROC AUC in the test set was not generated by this method. Results are shown both before and after adjusting the probability threshold in the training set for diagnosing TB to increase each of sensitivity and specificity in turn to 90%.

As shown in Table 17 below, the Random Forests method, which utilised all 22 proteins, yielded a signature with sensitivities and specificities in the test set of 73% and 71% initially, 92% and 58% after increasing sensitivity, and 95% and 43% after increasing specificity.

	Training set	Test set
Sensitivity	72 (61 - 81)	73 (56 - 86)
Specificity	80 (70 - 88)	71 (54 - 85)
PPV	77 (66 – 86)	71 (54 – 85)
NPV	75 (65 – 83)	73 (56 – 86)
After adjustment of cut-off for increased sensitivity		
Sensitivity	82 (73 - 90)	92 (78 - 98)
Specificity	64 (53 - 74)	58 (41 - 74)
PPV	69 (59 – 77)	68 (53 – 80)
NPV	79 (68 – 88)	88 (69 – 97)
After adjustment of cut-off for increased specificity		
Sensitivity	65 (54 - 75)	43 (27 - 61)
Specificity	89 (80 - 94)	95 (82 - 99)
PPV	85 (74 – 92)	89 (65 – 99)
NPV	72 (63 – 81)	63 (49 – 76)

Table 17 Performance of a new 22-protein signature derived from Random Forests analyses.

The same patients with the same allocations to training and test sets were used as for the GDA analyses (Table 9). ROC AUCs were not generated by this method.

The PReMS method yielded a nine-protein signature comprising fibrinogen, alpha-2-M, CRP, MMP-9, transthyretin, complement FH, IFN-gamma, IP-10, and TNF-alpha. As shown below in Table 18, this had a ROC AUC of 90% in the training set and 84% in the test set. Sensitivities and specificities in the test set were 86% and 74% initially, 92% and 71% after increasing sensitivity, and 75% and 81% after increasing specificity. At the cut-off for increased sensitivity, PPV and NPV in the test set were 75% and 90%, respectively.

	Training set	Test set
ROC AUC	**90 (86 – 95)**	**84 (73 – 94)**
Sensitivity	83 (75 – 90)	86 (73 – 95)
Specificity	82 (73 – 89)	74 (58 – 86)
PPV	84 (75 - 91)	85 (70 - 94)
NPV	82 (73 - 89)	76 (62 - 87)
After adjustment of cut-off for increased sensitivity		
Sensitivity	90 (83 – 95)	92 (80 – 98)
Specificity	71 (61 – 80)	71 (56 – 84)
PPV	75 (66 - 82)	75 (62 - 86)
NPV	89 (80 - 95)	90 (76 - 97)
After adjustment of cut-off for increased specificity		
Sensitivity	73 (63 – 82)	75 (60 – 87)
Specificity	90 (82 – 95)	81 (67 – 91)
PPV	87 (78 - 94)	80 (65 - 91)
NPV	78 (69 - 85)	77 (63 - 88)

Table 18 Performance of a nine-protein signature derived from Parallel Regularised Regression Model Search.

The same patients and allocations were used as for the GDA and Random Forests analyses (Tables 16 and 17). This signature comprised fibrinogen, alpha-2-M, CRP, MMP-9, transthyretin, complement FH, IFN-gamma, IP-10, and TNF-alpha.

The performance of the 9-protein signature was then examined after stratifying the test set by each of HIV status and site. As shown in Table 19 below, the ROC AUC in the test set was 87% for HIV uninfected patients and 84% for co-infected patients, regardless of site. At each of the two sites, the ROC AUC in the test set was 95% at Cape Town and 78% at Karonga, regardless of HIV status.

	ROC AUC (95% CIs)
HIV-	87 (72 - 100)
HIV+	84 (72 - 97)
Karonga	78 (63 - 93)
Cape Town	94 (86 - 100)

Table 19 ROC AUCs of the nine-protein signature after stratifying for each of HIV status and site.

Results shown relate to performance in the test set only. Bracketed values represent 95% confidence intervals. For the HIV- vs. HIV+ stratification, patients from both sites were kept together. For the Karonga vs. Cape Town stratification, patients with both HIV statuses were kept together.

5. Discussion

In regard to the first of two Primary Aims of this thesis, we quantified the serum concentrations (levels) of the 22 proteins from the AE-TBC study in a large sample of the EU-TB cohort. In the field of serum proteomics-based TB diagnostics research, this stands out in that firstly no study, to our knowledge, has ever tested a host serum protein-derived signature entirely by immunocapture in patients recruited to secondary care in Africa. As described in the Introductory Chapter, various studies have sought similar aims, but have either incorporated less rigorous recruitment criteria,(9) not incorporated patients co-infected with HIV,(10, 11) or not completed the process of protein identification and re-test by immunocapture.(9, 11) Other studies have met the above criteria, but, due to their own particular Aims, restricted recruitment to patients seeking attention at primary healthcare facilities.(1, 12) Secondly, this is the first time in the literature to our knowledge that biomarkers of TB discovered and tested at one level of a healthcare system have been re-tested at another (i.e. primary to secondary or vice versa). This is of paramount importance to global health, since testing kits could otherwise be erroneously applied to the wrong population with adverse consequences. The study presented in this chapter sheds light on the transferability of host protein biomarkers, probably in any body fluid, across such levels, and reveals that independent biomarker discovery efforts are warranted in each of primary and secondary care settings respectively.

Limitations of this study were that a) the EU-TB and AE-TBC studies had different designs: AE-TBC was a cohort study, with proportions of TB : OD cases, and HIV infected : uninfected reflecting the local epidemiology, whereas the EU-TB study was case-control, with the resulting group sizes pre-determined; b) biomarkers were not derived from an untargeted survey of sera collected from patients attending at secondary care. Discovery was therefore limited to a shortlist of 22 that had been selected from a different population, albeit from equivalent geographical areas.

Overall, the performance of the biomarkers individually was less good than in the AE-TBC study. Only 12 of the 22 biomarkers had a ROC AUC with a 95% confidence interval > 50%, and were thus significant to distinguish TB from OD in our cohort, whereas 20 proteins had met this criterion in the AE-TBC study.(1) In direct comparison of ROC AUCs, 14 of the proteins performed significantly less well in our cohort, with only one, complement FH, performing significantly better. The protein with the biggest drop in performance between studies was CRP, which had been the top biomarker individually in the AE-TBC study, and part of their 7-protein signature, but was entirely non-discriminatory in our dataset. The reason for this is very likely due to the different patient populations, and primarily the OD groups: those in the EU-TB cohort were all recruited at hospital attendance and therefore likely more unwell and with a greater disturbance to their serum proteome. By contrast, OD (or “non-TB”) patients recruited to the AE-TBC study did not have diagnoses confirmed, but were felt to be suffering largely from infections of the upper and lower respiratory tract, including exacerbations of COPD or asthma.(1) The resulting proteomes would likely have been much less deranged than in the EU-TB OD group, particularly with regard to the acute phase proteins. The markedly poorer performance of CRP in the EU-TB cohort is in keeping with this and has been reported before.(13, 14)

The top-performing biomarker in our cohort was complement factor H. Complement factor H was a novel biomarker of TB when published by AE-TBC, and as such is a novel TB biomarker in the setting of secondary care also.(1) Of particular promise was firstly that diagnostic performance was sustained across HIV status and geographical site, and secondly that the levels of factor H differed from those in the LTBI group in a different direction in TB as compared with OD: higher in TB, lower in OD. This pattern was not seen in any of the other protein concentrations and implies firstly that rising levels of factor H are a TB-specific effect, not just an acute phase response, and secondly that factor H is a particularly robust biomarker in more unwell populations.

The attempt to re-test the 7-protein AE-TBC signature in our cohort by direct transfer of a mulitvariate linear regression equation based on their dataset to ours was not successful. On comparison of protein levels between studies, focussing on the TB groups in order to remove variation arising from the difference in the control groups, it appeared that this was most likely due to marked variation in levels of 5 of the proteins, which were lower in our dataset in every case. In comparison to normal levels published in the literature, however, levels of 3 of the 5 proteins were also lower in our healthy LTBI control group, however levels of one (apo-CIII) were the same as in the literature and of another (procalcitonin) were higher. In addition, levels of apo-AI, which was the protein that was most different between TB groups, were unusually low in our study compared to the literature, but unusually high in the AE-TBC study. It seems most likely that the differences in levels between studies were therefore due to lot-lot variation in quantification, for example following some over-correction of calibration of quantification by the manufacturer in the time between the two sets of assays being performed. We did not discuss this with the manufacturers, however, and it is also possible that inter-operator variability contributed to some degree. This problem is informative for future research, however, since it implies that host serum protein signatures derived from immunocapture cannot easily be tested between datasets from different cohorts, and that combining serum/plasma samples is required for subsequent analysis at one site by one operator (or team of operators).

Re-application of the GDA method to the data from our cohort only was interesting in that it yielded a 5-protein signature, the constituent proteins of which were a subset of the AE-TBC 7-protein signature. Overall performance was poorer, however still comparable with 95% confidence intervals of the ROC AUC in the training set overlapping with that in the AE-TBC study. This result also provides verification that these 5 proteins were among the best for inclusion in the AE-TBC signature, and as such is an important finding.

The Random Forests method discovered a signature with sensitivity and specificity in the test that were both greater than 70%, which is a reasonable result, but not sufficient to satisfy any of the WHO targets for TB diagnostic tests as described in the Materials and Methods chapter. These results were also lower than in the AE-TBC, which found sensitivity and specificity in the test set to be 83% and 89% respectively.(1) This method does not generate ROC curves, hence it is more difficult to compare accuracy across studies, however these differences are in keeping with the overall poorer performance of the analytes as biomarkers individually.

The best results came from a 9-protein signature which was discovered by the PReMS method.(15) This method was also based on multivariate regression but selected a slightly different combination of proteins to the GDA method. Specifically, the GDA signature included serum amyloid A, whereas the PReMS signature did not. Instead, the PreMS method found that addition of fibrinogen, alpha-2-M, MMP-9, IFN-gamma and TNF-alpha further enhanced the diagnostic performance. These differences are most likely due to the mathematic differences in types of regression used, although there may also be some degree of inter-changeability between proteins. Four proteins were common to both signatures, however: complement FH, IP-10, CRP and transthyretin, suggesting that these biomarkers make clear contributions independently to diagnosis. It is interesting also, however, that CRP was included in both multivariate regression signatures, since it did not appear to have definite discriminatory potential on its own (ROC AUC 51% (95% CI 43% – 58%)). This was also the case for fibrinogen, TNF-alpha and MMP-9, which were included in the PreMS signature (ROC AUCs 56% (49% – 63%), 53% (46% – 59%) and 53% (47% – 60%) respectively). The reason for this must simply be that their inclusion adds a very small but significant amount to the overall diagnostic performance. In a heterogenous population of both TB and OD cases, one can imagine that some patients will be correctly diagnosed on the basis of the levels of one biomarker (e.g. TNF-alpha), whereas for another the level of MMP-9 will be important. One

further factor within this may be HIV status, such that overall a protein may appear to be unreliable as a biomarker individually but may be necessary in either HIV- or + populations. In the case of these four proteins (CRP, fibrinogen, TNF-alpha and MMP-9), this was not obviously the case from comparing the 95% confidence intervals of the ROC AUCs in each of HIV- and HIV+ individuals (they overlapped), however the differences did appear quite marked in the cases of alpha-2-M and MMP-9 (Figure 33).

The performance of the 9-protein signature was particularly promising on adjusting the cut-off for increased sensitivity. At this threshold, the sensitivity and specificity in the test set met the criteria for the WHO triage test for TB.(16) These criteria were set for use of a triage test in community settings, however, and the same criteria may not be universally felt to be sufficient for secondary care settings. Whilst we did not have definite categories of severity in which to test signature performance, however, it is likely that the cases of TB that are missed by the signature would be less severe than those detected. This was the case in the AE-TBC study where the performance was least good in smear-negative and culture-negative cases,(1) and it is generally seen in host-based diagnostic tests that the more pronounced the illness, the greater the "molecular distance to health", and the more accurate the test becomes.(17) This would suggest that any TB cases missed by a test such as our 9-protein signature at presentation to secondary care would be less unwell than those detected, and that with appropriate safety-netting advice the test would be safe. Verification in future studies would be needed, however, to ensure that use of such a test as a screening tool in secondary care is safe, and in addition to saving money and time, positively affects health outcomes overall. Our findings that individual proteins vary in their performance between primary and secondary care sites, and that there are biomarkers such as complement factor H that perform significantly better in secondary care, suggest that further studies incorporating additional biomarkers may yield signatures that perform even better than our 9-protein signature. As a result, weight is added to the importance of the proteomic arm of the EU-TB study, in which an untargeted survey of serum proteomes was performed of patients presenting to such locations. Lastly, complement FH itself represents a biomarker that could well add value to the signatures discovered by the EU-TB SELDI study, since at a molecular weight of 155 kDa is it above the limit of accurate quantification of intact proteins detectable by SELDI-TOF mass spectrometry.(18) This finding therefore fulfils Primary Aim1 iii), which was "to discover any promising individual biomarkers that could enhance the performance of the existing EU-TB signatures".

In summary, we quantified the levels of the 22 proteins from the AE-TBC study in sera from a large sample of the EU-TB cohort. We identified that whilst data from two separate studies cannot easily be combined to study the performance of a signature across cohorts, promising signatures can be discovered based on proteins selected from other cohorts, including those recruited from a different level of healthcare systems. In addition, individual serum protein biomarkers perform differently across primary and secondary care cohorts, with the majority of proteins that were discovered in primary care performing less well in secondary care. This is most likely due to the greater degree of derangement in the serum proteome of patients with other diseases that present to secondary care. However, some proteins may perform better in secondary care, and we confirmed this in in the EU-TB cohort in the case of complement factor H. This protein is now a candidate for inclusion in the validation steps of the EU-TB proteomic project, as to whether it may add to the diagnostic performance of the EU-TB signatures in the context of multiplexed detection by immunocapture.

CHAPTER IV RESULTS: Levels of AE-TBC proteins before and after treatment in a paediatric TB cohort

1. Background

Of the 22 proteins in the AE-TBC, the majority differed in level in the TB and OD groups in the same direction in comparison to the healthy LTBI controls. This was likely because most of the proteins were either acute phase proteins or cytokines, and the OD group comprised mainly inflammatory or other infectious conditions which would be expected to influence levels of such proteins in a similar way to TB, albeit to differing degrees. The one protein that didn't fit this pattern was complement factor H (complement FH): in comparison to LTBI, levels were higher in the TB group ($p = 4.1$ e-10) but trended towards being lower in the OD group ($p = 0.08$, Chapter III, Figure 33). Elevated levels of complement FH have previously been shown to predispose to meningococcal and pneumococcal infection, and to be genetically determined.(1-3) Complement FH is a major inhibitor of the alternative complement pathway and is bound by the surface of various pathogens, as it is by host cells, to resist complement-mediated attack as reviewed by Parente *et al.*(4) As a result, a form of genetic susceptibility to meningococcal disease exists. *Mycobacterium bovis* BCG has also been shown to avidly bind complement FH resulting in reduced phagocytosis by macrophages *in vitro*.(5, 6) In addition, the presence of complement FH led to a marked increase in the pro-inflammatory cytokines TNF-alpha, IL-1 and IL-6 secretion by macrophages upon incubation with *M. bovis* BCG, but a decrease in secretion of IL-10, TGF-beta and IL-12.(6) It could be conceived that this might predispose to acute inflammation and dissemination, since there would be a pro-inflammatory cytokine balance and less Th1-mediated intracellular killing of Mtb by macrophages due to less IL-12. We therefore hypothesised that a similar pattern might occur in TB, namely that higher complement FH levels predispose to more severe infection, albeit possibly by a different mechanism to in meningococcal disease. If this was the case, we would expect the higher levels of serum FH pre-treatment to persist post-treatment.

To investigate this, unstimulated heparinised whole blood supernatants were used from a paediatric cohort that had previously been recruited by Dr Elizabeth Whittaker (Clinical Lecturer, Paediatric Infectious Diseases and Immunology, Imperial College) from Red Cross Children's Hospital, Cape Town, South Africa. As described in Chapter II (Materials and Methods), samples were analysed from 39 children with active TB (pulmonary and/or extrapulmonary) at baseline and after 6 months of TB treatment, and from 43 healthy control children. All children were HIV uninfected. Our hypothesis was that levels of complement FH in TB patients would be the same after treatment as before, which would fit with our suggestion that they were higher because of genetic factors. The contingency was that if levels returned to normal after treatment, then we would gain information on the utility of complement FH as a biomarker of treatment response.

The samples used for this thesis were selected randomly from Dr Whittaker's biobank. The demographics of the TB patients and healthy controls, as well as the culture status and disease site for the TB patients is shown in Table 20 below. With regards to disease site, for this cohort, the term "pulmonary TB" was used to refer to pulmonary disease with or without dissemination. Patients with disease solely outside of the lung were termed "extrapulmonary" cases. These were as per Dr Whittaker's original case definitions.

	TB patients (n = 39)	Healthy Controls (n = 43)
Age in months (median (IQR))	53 (21 – 103)	17 (6.5 – 33.5)
Male (n (%))	21 (54)	33 (77)
Culture positive (n (%))	23 (59)	n/a
Pulmonary (n (%))	28 (72)	n/a

Table 20 Demographic and clinical features of the patients and healthy controls randomly selected for this study from the paediatric cohort of Dr Elizabeth Whittaker.

The definition of pulmonary TB was disease in the lung with or without dissemination. Non-pulmonary (or extrapulmonary) cases were those with disease solely outside of the lung.

The same Luminex kits were used as for the adult cohort. For the second of the two plates that were run, the assay for transthyretin had been replaced by complement C3 by the manufacturer. Data were therefore gathered on levels of complement FH, apo-AI and apo-CIII for the number of children stated above (including TB patients before and after treatment), on transthyretin for 15 TB patients and 22 healthy controls, and on complement C3 for 24 patients and 21 healthy controls. Samples had previously been diluted 1 in 10 in RPMI media and were then diluted further for these assays in Luminex assay buffer according to manufacturer instructions (Merck Millipore). For ease, these will be referred to as "plasma" herein. Results from these assays are shown in Figures 38 – 42 with an explanation of results given after each figure. In each case, the effect size is given as the median of the differences between pre- and post-treatment levels.

2. Results

Contrary to our hypothesis, levels of complement FH in active TB patients were significantly lower after treatment than before (Figure 38 below, left panel). In comparison with the healthy controls, levels at month 0 were significantly higher but by month 6 this was no longer the case. In fact, the median at month 6 appeared lower than that of the healthy controls. The right panel of Figure 38 confirms that FH levels were lower after treatment in most TB cases, however in 12 of 39, levels were higher after treatment. It is evident, however, that the patients with the highest levels before treatment all had lower levels after treatment. The median of the differences in FH levels before and after treatment was -49.8 ug/ml.

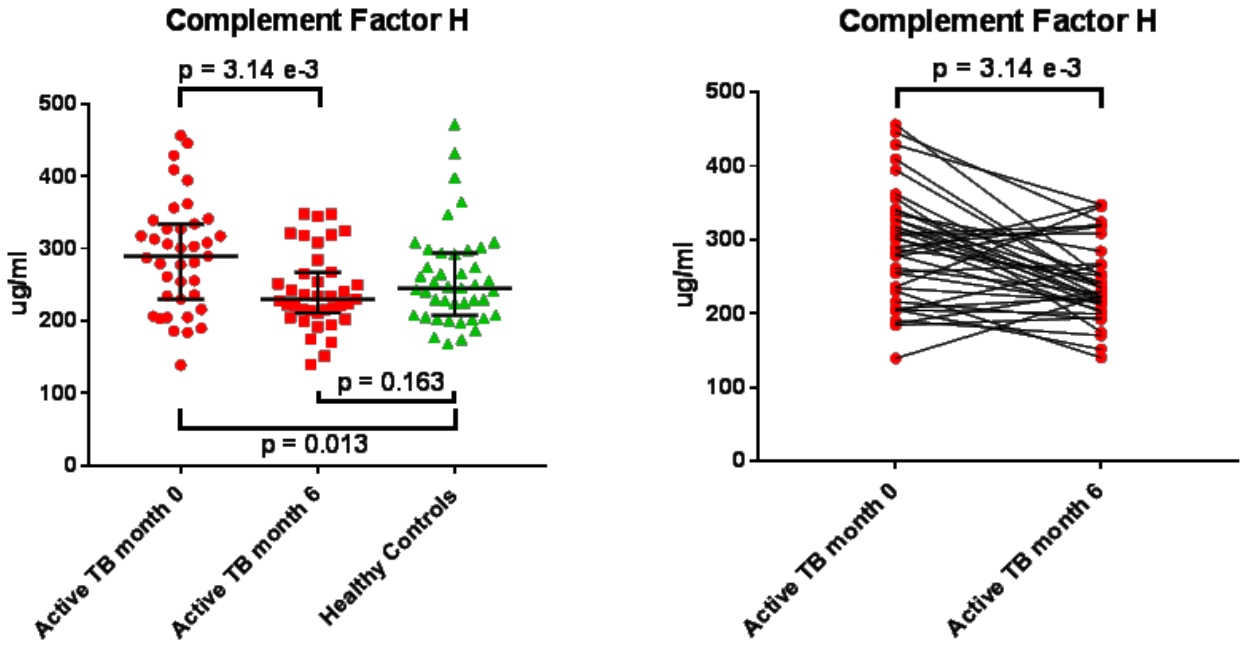


Figure 38 Plasma levels of complement factor H (FH) in a South African paediatric TB cohort (n = 39) before and after TB treatment, as measured by Luminex.

Left panel: FH levels at baseline and after treatment (months 0 and 6 respectively) and in 43 healthy control children. Right panel: FH levels with connecting lines for each child before and after treatment. P-values are one-sided and derived from a Wilcoxon matched pairs signed-rank test for the comparison of levels before and after treatment and from Mann-Whitney tests for comparisons with healthy controls.

Levels of apolipoprotein A-I (apo-AI) in TB patients were significantly higher after treatment than before (Figure 39, below). In comparison with healthy controls, levels at month 0 were significantly lower, but by month 6 this was no longer the case (left panel). The right panel of Figure 39 confirms that apo-AI levels were higher in most patients after treatment, although in 10 of 39, they were lower after treatment. The median of the differences in levels before and after treatment was 64.6 ug/ml.

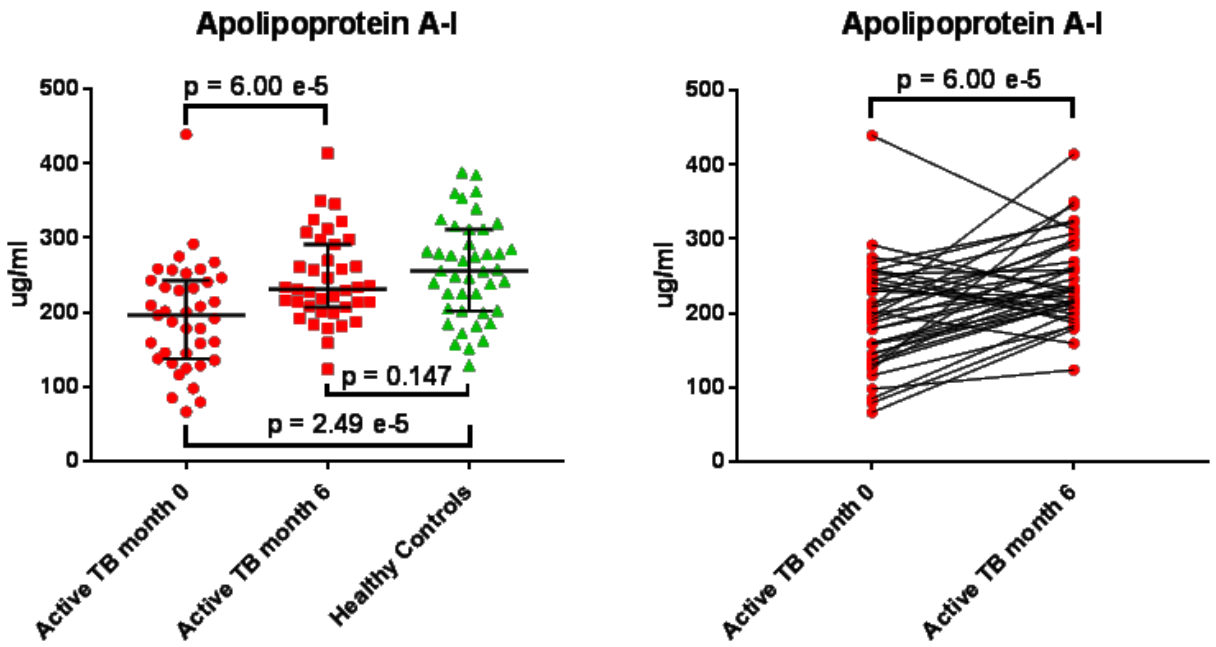


Figure 39 Plasma levels of apolipoprotein A-I (apo A-I) in a South African paediatric TB cohort (n = 39) before and after TB treatment, as measured by Luminex.

Left panel: apo A-I levels at baseline and after treatment (months 0 and 6 respectively) and in 43 healthy control children. Right panel: apo-AI levels with connecting lines for each child before and after treatment. P-values are one-sided and derived from a Wilcoxon matched pairs signed-rank test for the comparison of levels before and after treatment and from Mann-Whitney tests for comparisons with healthy controls.

Levels of apo C-III were not significantly higher after treatment than before (Figure 40, below). In comparison with healthy controls, levels were significantly lower in TB cases both before and after treatment (left panel). The right panel shows that levels were higher after treatment in 17 of 39 patients, but lower in 22. In most patients, however, levels appear very similar before and after treatment. The median of the differences in levels before and after treatment was -12.0 ug/ml.

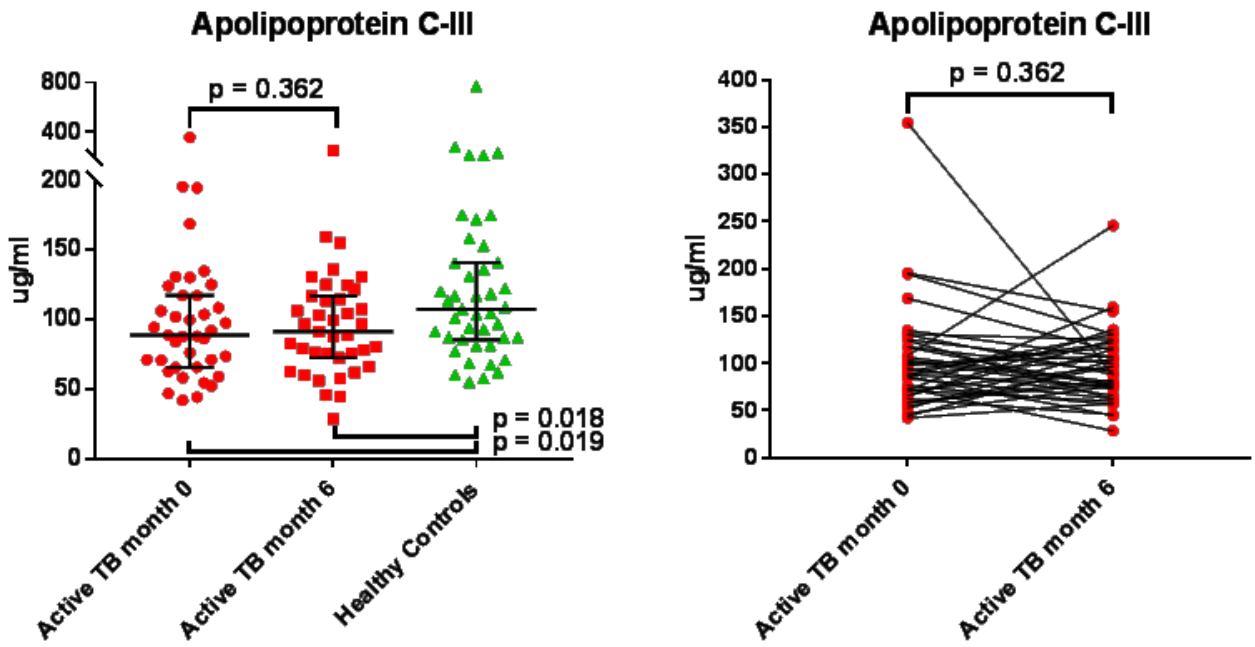


Figure 40 Plasma levels of apolipoprotein C-III (apo-CIII) in a South African paediatric TB cohort (n = 39) before and after TB treatment, as measured by Luminex.

Left panel: apo C-III levels at baseline and after treatment (months 0 and 6 respectively) and in 43 healthy control children. Right panel: apo C-III levels with connecting lines for each child before and after treatment. P-values are one-sided and derived from a Wilcoxon matched pairs signed-rank test for the comparison of levels before and after treatment and from Mann-Whitney tests for comparisons with healthy controls.

Levels of transthyretin were significantly higher in the TB cases after treatment than before (Figure 41, below). In comparison with healthy controls, however, levels were still significantly lower at month 6 (left panel). The right panel shows that levels increased in 10 of 15 patients but decreased in 5. The median of the differences in levels before and after treatment was 22.6 ug/ml.

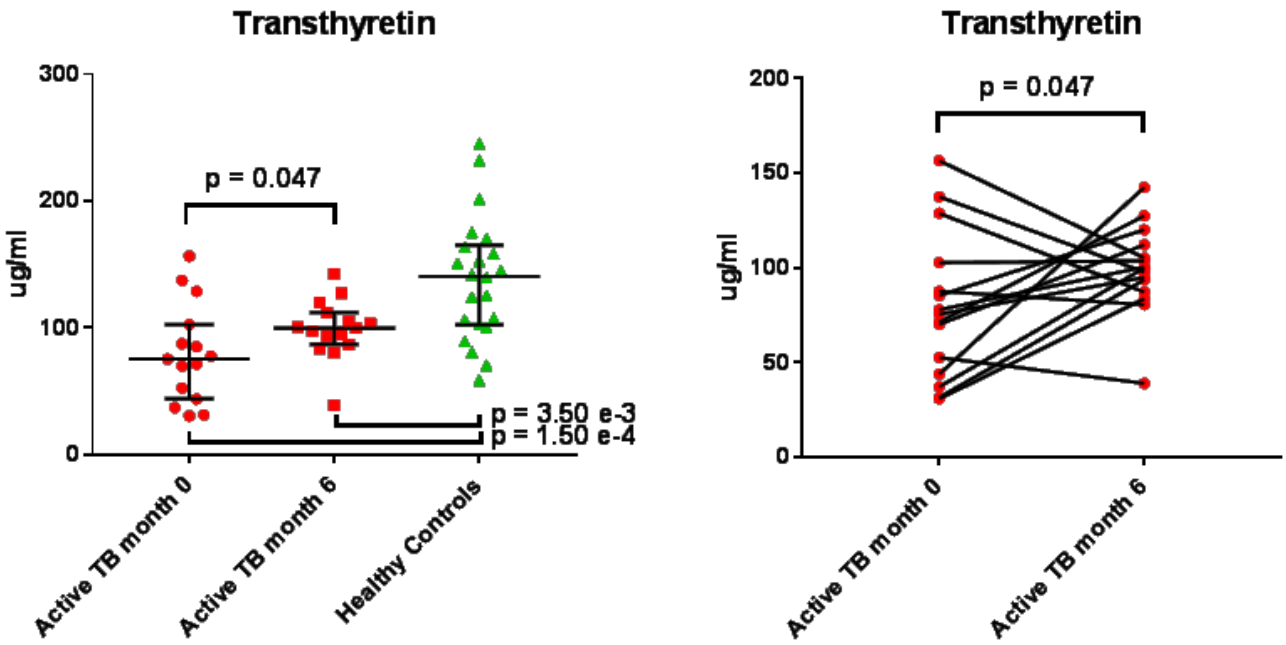


Figure 41 Plasma levels of transthyretin in a South African paediatric TB cohort (n = 15) before and after TB treatment, as measured by Luminex.

Left panel: transthyretin levels at baseline and after treatment (months 0 and 6 respectively) and in 22 healthy control children. Right panel: transthyretin levels with connecting lines for each child before and after treatment. P-values are one-sided and derived from a Wilcoxon matched pairs signed-rank test for the comparison of levels before and after treatment and from Mann-Whitney tests for comparisons with healthy controls.

Levels of complement C3 were not significantly higher after treatment (Figure 42, below). In comparison with healthy controls, however, levels in TB cases were significantly lower before treatment but not after treatment (left panel). The right panel shows that complement C3 levels were higher after treatment in 12 of 24 patients but lower in the other 12. The median of the differences in levels before and after treatment was 8.6 ug/ml.

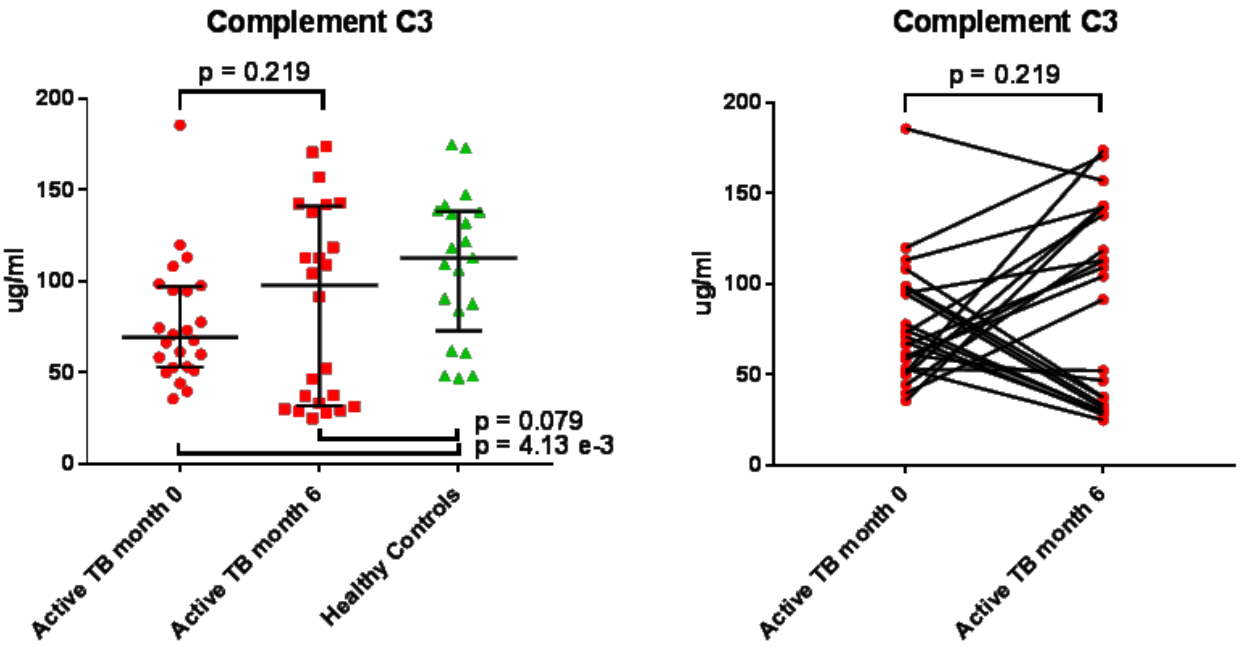


Figure 42 Plasma levels of complement C3 in a South African paediatric TB cohort (n = 24) before and after TB treatment, as measured by Luminex.

Left panel: complement C3 levels at baseline and after treatment (months 0 and 6 respectively) and in 22 healthy control children. Right panel: complement C3 levels with connecting lines for each child before and after treatment. P-values are one-sided and derived from a Wilcoxon matched pairs signed-rank test for the comparison of levels before and after treatment and from Mann-Whitney tests for comparisons with healthy controls.

In summary, levels of three of the five proteins that were studied returned to or towards the level of the healthy controls after TB treatment (i.e. complement FH, apo-AI and transthyretin). To assess whether it was the same patients whose protein levels returned towards the levels of the healthy controls after treatment (i.e. as expected) or moved further away, results for each patient were tabulated and cells coloured to indicate direction of change (Table 21, below). Only one patient had levels of all three proteins move further away from control levels (n = 15). When only complement FH and apo-AI are considered, which were measured in all 39 children, only 3 patients had levels of both proteins move further away from control levels. Numbers of patients in whom protein levels failed to normalise are also shown by protein as a Euler diagram below (Figure 43).

Patient ID	Apo-A1	Transthyretin	Complement FH	No. proteins whose levels moved away from those of the healthy controls
3	-161.9		24.3	0
4	-83.4	-34.8	206.6	0
5	-149.3		36.1	0
10	-48.6		81.2	0
14	7.7		53.8	1
15	-73.8	-42.1	-31.5	1
22	-67.8		18.1	0
24	-65.4	-62.8	33.9	0
25	25.7	-1.1	181.9	1
29	-76.5	-52.9	101.1	0
32	-132.5	-62.3	-66.7	1
34	-26.0	13.7	40.3	1
35	42.1	51.4	95.4	2
38	-68.2		5.2	0
39	44.9	-19.5	-10.4	2
41	-42.2		160.1	0
44	-64.6		-40.5	1
46	-72.7		123.2	0
55	47.1	-22.6	-62.7	2
56	-6.0		-10.1	1
57	48.5		90.1	1
62	41.9	39.7	84.2	2
64	-86.0	-98.7	111.7	0
66	-106.2		139.2	0
67	-29.4	-56.2	80.9	0
68	-37.1	41.5	-11.1	2
69	178.0		291.2	1
71	-82.4		85.3	0
78	71.0	7.1	-108.7	3
79	-74.4		60.3	0
80	127.1		121.5	1
90	-112.2		-84.6	1
91	-55.7		-43.8	1
93	-204.4		-2.1	1
94	-382.0		-280.9	1
96	64.8		102.1	1
98	-54.4		175.1	0
103	-104.0		-15.1	1
105	-225.0		19.7	0
106	-115.9		74.9	0
107	-57.4		49.8	2

Table 21 Difference in levels of apo-AI, transthyretin and complement factor H in a South African paediatric TB cohort before and after TB treatment, as measured by Luminex.

Cells are coloured to indicate whether protein level returned to or towards the normal level or not. The number of patients with apo-AI and complement FH levels measured was 39. The number of patients with transthyretin levels measured was 15. Green fill: returned to or towards the level of healthy controls; red fill: moved further away. All values are in ug/ml. Right hand column shows how many proteins moved further away from that of the healthy controls per patient.

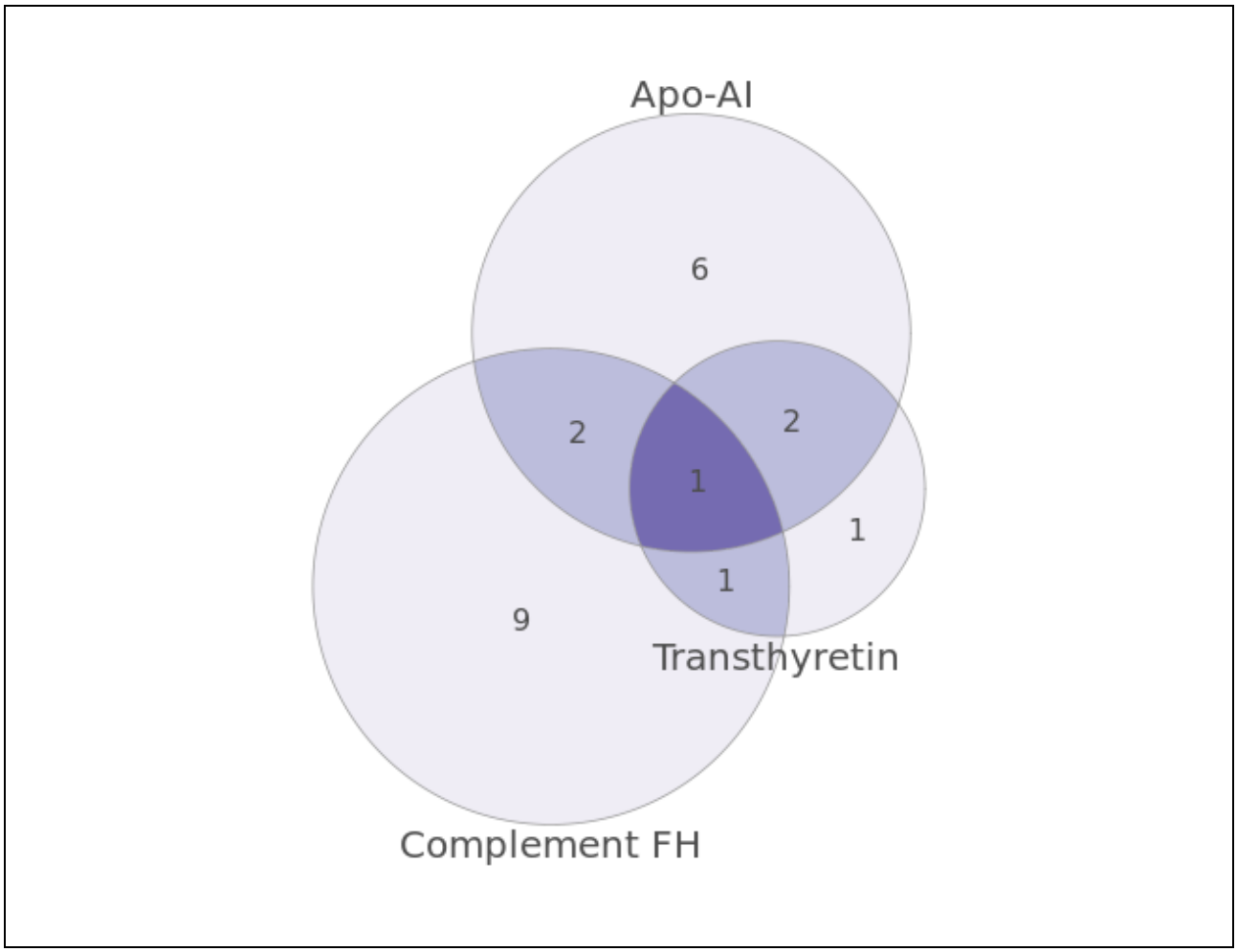


Figure 43 Euler diagram showing numbers of patients with protein levels that failed to normalise after treatment, by protein.

Only those patients that had protein levels that failed to normalise after treatment are shown. The diagram was created at www.venndiagrams.net.

3. Discussion

The driver for this study was the prior knowledge that elevated levels of complement FH predispose to other bacterial infectious diseases in the context of the facts that a) complement FH levels were higher in our adult TB group relative to the LTBI groups, but trended towards being lower in the OD group, and b) complement FH levels are genetically determined. In this dataset, levels of complement FH returned to or towards normal after treatment in 27 of 39 patients with active TB, and the average effect size was also negative (median of differences: -49.8 ug/ml). There was also no significant difference between end of treatment with those of the healthy controls, suggesting that, in this paediatric cohort, levels of complement FH in untreated TB patients compared to the healthy controls were not higher because of genetic polymorphisms, but rather represented a host response. The biology of active TB in adults differs from that in children in that most adult TB cases (especially those not co-infected with HIV) arise from reactivation of LTBI, whereas in paediatric TB, clinical disease arises from continuous progression of primary infection. To our knowledge, serum levels of complement FH have not previously been measured in paediatric TB, however if increased susceptibility to TB was conferred by genetic polymorphisms producing higher serum levels, we would expect this effect to manifest in

childhood as well as adulthood. It is still possible that higher serum complement FH levels contribute towards TB susceptibility to a minor extent, since the relevant mutations are present in African populations and the range of serum levels in the healthy controls was wide in this sample (168 – 472 ug/ml, IQR 208 – 293 ug/ml).(7) In addition, genome-wide association studies of TB susceptibility have used generic genotyping chips that would not necessarily be sensitive for these mutations in African populations.(8, 9) To investigate this further, however, a large case-control study would be required to compare the number of incident cases of TB per person with and without the relevant mutations, ideally with paired sera to confirm the three-way association in the study population. Such an undertaking is not supported by this data.

The results for complement FH are still a novel finding, however, since levels of this protein have not previously been measured before and after treatment in TB, in adults or children. Indeed, a return to or towards normal after successful TB treatment is one source of verification of host-derived diagnostic biomarkers. A comparison of levels before and after treatment also allows an initial assessment of performance of proteins as biomarkers of treatment response. Complement FH may therefore have some utility in this regard.

The other protein whose levels normalised after treatment was apo-AI (Figure 40). As with complement FH, levels of apo-AI before and after TB treatment have not previously been reported. The fact that they normalise is promising since both proteins were in the 7-protein SUN community-based TB diagnostic signature in adults. This also indicates a degree of overlap between biomarkers of TB in adults and children.

Transthyretin was measured in a smaller number of patients (15 with TB and 22 healthy controls), but the level of this protein significantly increased on treatment in two thirds of cases. On average, however, it appeared that levels returned towards, but not completely to, normal (Figure 41). This is interesting and may reflect the fact that whilst the inflammatory drive to suppress transcription of the transthyretin gene might have lifted by the end of treatment, lean body mass may not have fully recovered. There may therefore be a need to ensure TB patients go on to fully recover their nutritional status post-TB cure. The finding that transthyretin levels rise during recovery on TB treatment has been previously demonstrated, but not in children.(10)

Levels of apo-CIII did not appear to normalise at all after treatment. This is surprising since levels were lower in the active TB group compared with the LTBI group in the adult cohort (albeit only marginally, Chapter III, Figure 22). Apo-CIII serum levels have never previously been reported in paediatric TB, however it is biologically interesting since apo-CIII is a component of lipoproteins, just as is apo-AI. Apo-AI is solely present in high-density lipoproteins (HDLs) where it is the primary structural protein, present at one molecule per HDL particle. By contrast, apo-CIII is an accessory protein present in all density classes of plasma lipoproteins. It is the most abundant of the apo-C class and is involved in lipoprotein particle assembly and inhibition of degradation by lipoprotein lipase.(11, 12) It may be that the failure (or delay) of serum apo-CIII levels to normalise post-treatment reflects the fact that hepatic triglyceride stores remain low and, like lean body mass, need longer to recover.

Levels of complement C3 trended towards that of the healthy controls post-treatment but were not significantly higher than before treatment (Figure 42). Complement C3 is constitutively expressed and secreted by the liver and is the common target of the three pathways of complement activation (classical, alternative and lectin-binding), as reviewed by Norris and Remuzzi.(13) Upon its lysis by any of the respective C3 convertase enzymes, a chain reaction ensues which results in the formation of the membrane attack complex. Levels are therefore decreased in any condition that results in activation of any of the three pathways and have previously been found to be significantly higher post-TB

treatment in adults.(14) It seems likely that this is also the case in children, however the range of concentrations post-treatment was wide, and the number of children whose sera were assayed for this protein was lower. Sera from a larger numbers of patients would be needed to answer this question.

Numerous studies have measured host serum protein biomarkers before and after treatment for the purposes of finding markers of cure or impending treatment failure or relapse, whether with a targeted or untargeted approach. These have largely been restricted to acute phase proteins, however, with the most frequently studied protein that has been consistently found to normalise post-treatment being CRP.(15-18) At the time of writing, no signatures of serum proteins have been developed to predict relapse-free cure either during, or at the end of, TB treatment. A large study is underway, however, led by the University of California, San Francisco, to collect and study samples from 1000 patients with pulmonary TB at multiple timepoints during treatment and up to and including 12 months post-treatment (the TB SMART study: http://grantome.com/grant/NIH/R01-AI104589-05). The aim of this is to discover surrogate markers of treatment effect, including predictive ones that can be measured before the end of treatment, so that drug trials of novel compounds can be conducted in an expedited manner and without reliance on sputum culture. Our finding that levels of apo-AI, transthyretin and complement FH move away from those of healthy controls in a significant proportion of patients, but very seldom in unison, validates the need for such a study as it would appear that no single serum protein biomarker will be reliable in isolation for such a purpose.

The concept of mass-producing a multiplexed serum protein-based test for the purposes of treatment monitoring in clinical trials is feasible. Use of such a tool in individual patient care in resource limited settings, however, seems less feasible, since it would mandate an extra cost for healthcare systems on top of whatever technology is being purchased for diagnosis. If some of or all the biomarkers that are included in a diagnostic testing kit also functioned as markers of treatment response, however, this would likely be of great help to clinicians since pre-treatment readings would have already been taken and the test could then be repeated as needed. The data presented in this Chapter suggest that complement FH, apo-AI and possibly transthyretin, which were all included either in our 9-protein diagnostic signature or the AE-TBC 7-protein signature, may have utility to this end. A limitation of the study in this chapter, however, is that outcome data was not prospectively collected. These three proteins would therefore need to be re-tested in larger paediatric and adult cohorts that are recruited specifically for this purpose and followed up beyond treatment completion to document cure vs. failure/relapse.

CHAPTER V RESULTS: Evaluation of performance of protein biomarkers from Surface-enhanced laser desorption ionisation time-of-flight mass spectrometry (SELDI) signatures by immunoassay

1. Introduction

As described in the Introductory Chapter (Chapter I, Section 4), from the serum proteomic arm of the EU-TB study, two diagnostic signatures were discovered by surface-enhanced laser desorption ionisation time-of-flight (SELDI) mass spectroscopy: one to distinguish TB from LTBI and OD combined (LTBI/OD) and the other to distinguish TB from OD alone. Both signatures consisted of four proteins which had been identified using forward selection partial least squares (FSPLS) regression variable selection of data from 588 patients.(1) The performances of these signatures were previously shown in the Introductory Chapter (Table 3).

To advance translatability of these proteins to point-of-care (POC) diagnostic tests using lateral flow immunoassay (LFA), levels of each protein were measured individually in serum from a subset of patients from each clinical group (TB, LTBI and OD) by ELISA. As stated in the Materials and Methods Chapter (Chapter II, Section 3, part 2 a)), 30 patients were selected at random from each of the six clinical groups in the EU-TB study, i.e. 180 patients in total. From each site, 15 were selected for each of the six clinical groups. The exact number of patients for whom data were obtained on each protein (e.g. after sera had to be re-run at a different dilution) is given below along with the results for each protein.

For this thesis, levels of the following three proteins were quantified by single-plex ELISA: transthyretin, beta-2-microglobulin (beta-2-M) and alpha-1-antichymotrypsin (ACT). Transthyretin and ACT were part of both signatures (i.e. TB vs. LTBI/OD and TB vs. OD alone), whereas beta-2-M was only part of the TB vs. LTBI/OD signature. With regards to ACT, in each of the two signatures, it was represented by a different molecular weight peak: 4,357 m/z for the TB vs. LTBI/OD signature and 4,468 m/z for the TB vs. OD signature. These were both later identified by Dr Shea Hamilton as being C-terminal fragments of ACT, but as having different isoelectric points , confirming that they represented different fragments of the 47,651 Da parent protein.(2)

Results of these ELISAs are shown in this chapter. For each protein, data are presented in the following five ways:

i) correlation plots to compare level by ELISA with intensity by SELDI;
ii) scatter box-plots to show and compare protein levels between clinical groups;
iii) individual receiver operating characteristic (ROC) curves;
iv) side-by-side comparison of individual ROC AUCs by ELISA vs. by SELDI;
v) individual ROC curves by each of HIV status and study site.

To grade strength of correlation, the Evans classification for Pearson's correlation coefficients was used: < 0.20 is very weak, 0.20 to 0.39 is weak, 0.40 to 0.59 is moderate, 0.60 to 0.79 is strong and 0.80 or greater is very strong.(3)

Regarding expected levels of each biomarker by immunoassay in each group, in the SELDI data, intensities in the TB group compared to the respective control group (whether LTBI/OD or OD alone) were lower for transthyretin, but higher for beta-2-M and ACT.

With regards to transthyretin, since levels were also measured by Luminex on a larger number of patients for the purposes of re-evaluating the AE-TBC signature, a correlation plot and ROC curve derived from the Luminex data are shown for comparison alongside the ELISA data.

With regards to ACT, correlation plots and ROC curves are shown for each of the two signatures, using data on intensity of each of the corresponding SELDI peaks (4,357 m/z and 4,468 m/z).

2. Results

i) Transthyretin

Serum levels of transthyretin were obtained by ELISA for a total of 167 patients. There was a moderately strong correlation between levels of transthyretin when measured by ELISA and the corresponding SELDI intensities (rho = 0.55 (95% confidence intervals 0.42 – 0.65)) as shown in Figure 44A, below. In the ELISA data, levels of transthyretin were lower in the active TB group than in the control group (i.e. LTBI/OD, Figure 44C). This mirrors the findings for this biomarker by SELDI. This difference was largely due to the difference in level between the TB group and the LTBI group (Figure 44D), although levels in TB trended towards being lower than in OD at this sample size.

Figures 44E and 44F shows the ROC curves for transthyretin (TTR) for each of the two diagnostic aims (TB vs. LTBI/OD and TB vs. OD) using the ELISA data. To distinguish TB from LTBI/OD, the ROC AUC was significant at 0.68 (95% confidence interval 0.60 – 0.76), although this was slightly lower than in the SELDI data when the same 167 patients were considered (0.73 (0.65 – 0.81)). To distinguish TB from OD alone, the ROC AUC by ELISA was 0.55, however the 95% confidence intervals crossed 0.50 (0.44 – 0.66).

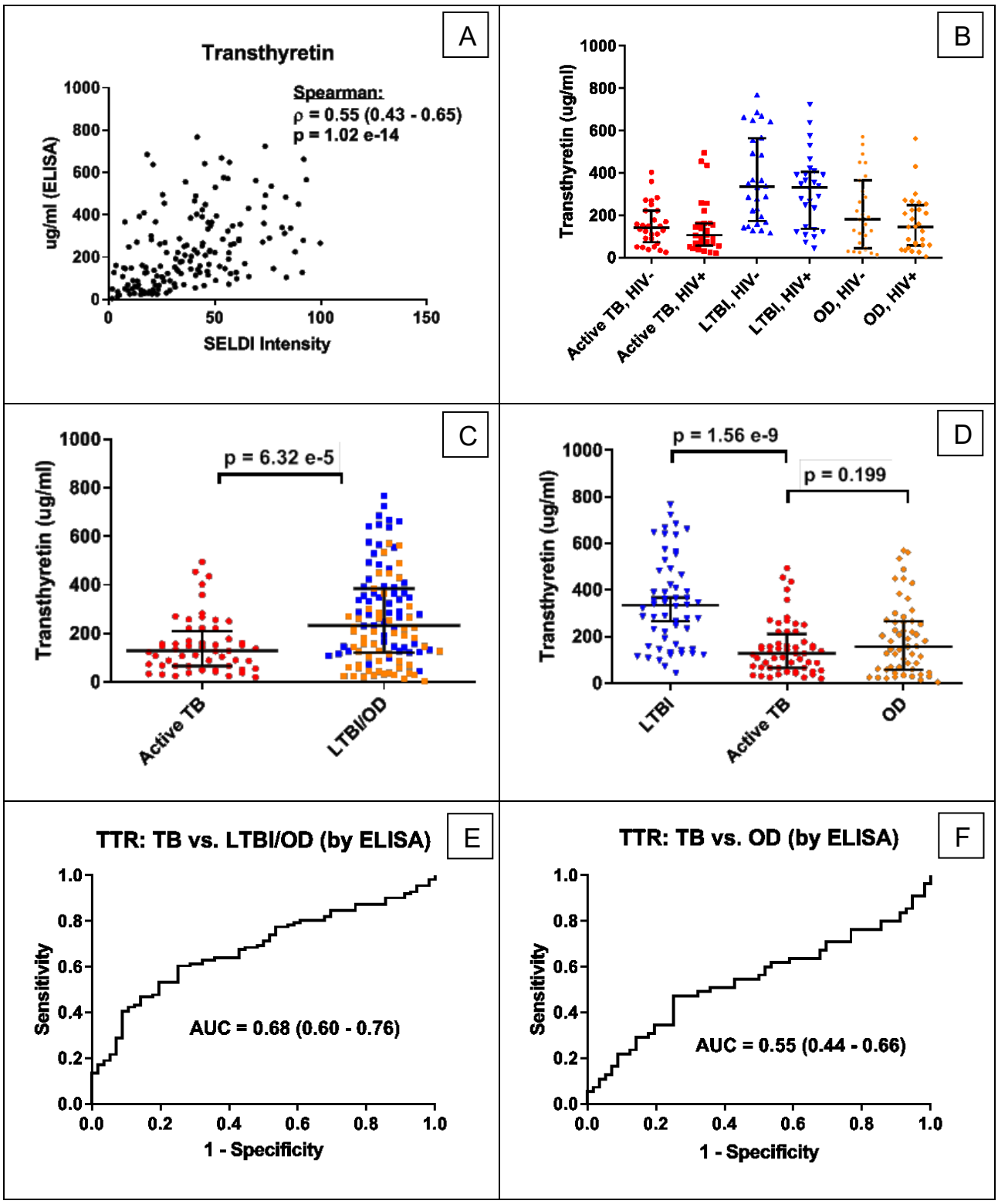


Figure 44 Serum levels and diagnostic performance of transthyretin (TTR) as measured by ELISA (N = 167).

A: correlation plot of TTR levels by ELISA with TTR intensities by SELDI. Bracketed values represent 95% confidence intervals of Spearman's rank correlation coefficient. B: TTR in the subgroups. C: TB compared with LTBI and OD together, regardless of HIV status. D: levels in LTBI, active TB and OD groups, regardless of HIV status. E: receiver operating characteristic (ROC) curve for use of TTR level to distinguish active TB from LTBI and OD combined, regardless of HIV status. F: ROC curve to distinguish TB from OD alone, regardless of HIV. p-values in the box plots were derived from Mann-Whitney tests and are 1-sided. AUC: area under the curve.

For comparison, the correlation of the Luminex data with the SELDI data for transthyretin is shown below in Figure 45A, along with the two ROC curves derived from the Luminex data for each of TB vs. LTBI/OD and TB vs. OD (Figures 45B and 45C). A side-by-side comparison of the ROC AUCs of transthyretin from each of the SELDI data, ELISA data and Luminex data is then shown in Table 44. Figure 45A shows that the correlation between the Luminex data and SELDI data for transthyretin was significantly stronger than between the ELISA and SELDI data (Spearman's rho 0.77 vs. 0.55, confidence intervals not overlapping). This translated into higher ROC AUCs with the Luminex data for both TB vs. LTBI/OD and TB vs. OD (Table 22, below). The ROC AUCs by Luminex were also closer to those of the SELDI data on the corresponding patients than was the case with the ELISA data (Table 22). The confidence intervals of the ROC AUCs by ELISA overlapped those of the ROC AUCs by SELDI in the corresponding patients, however.

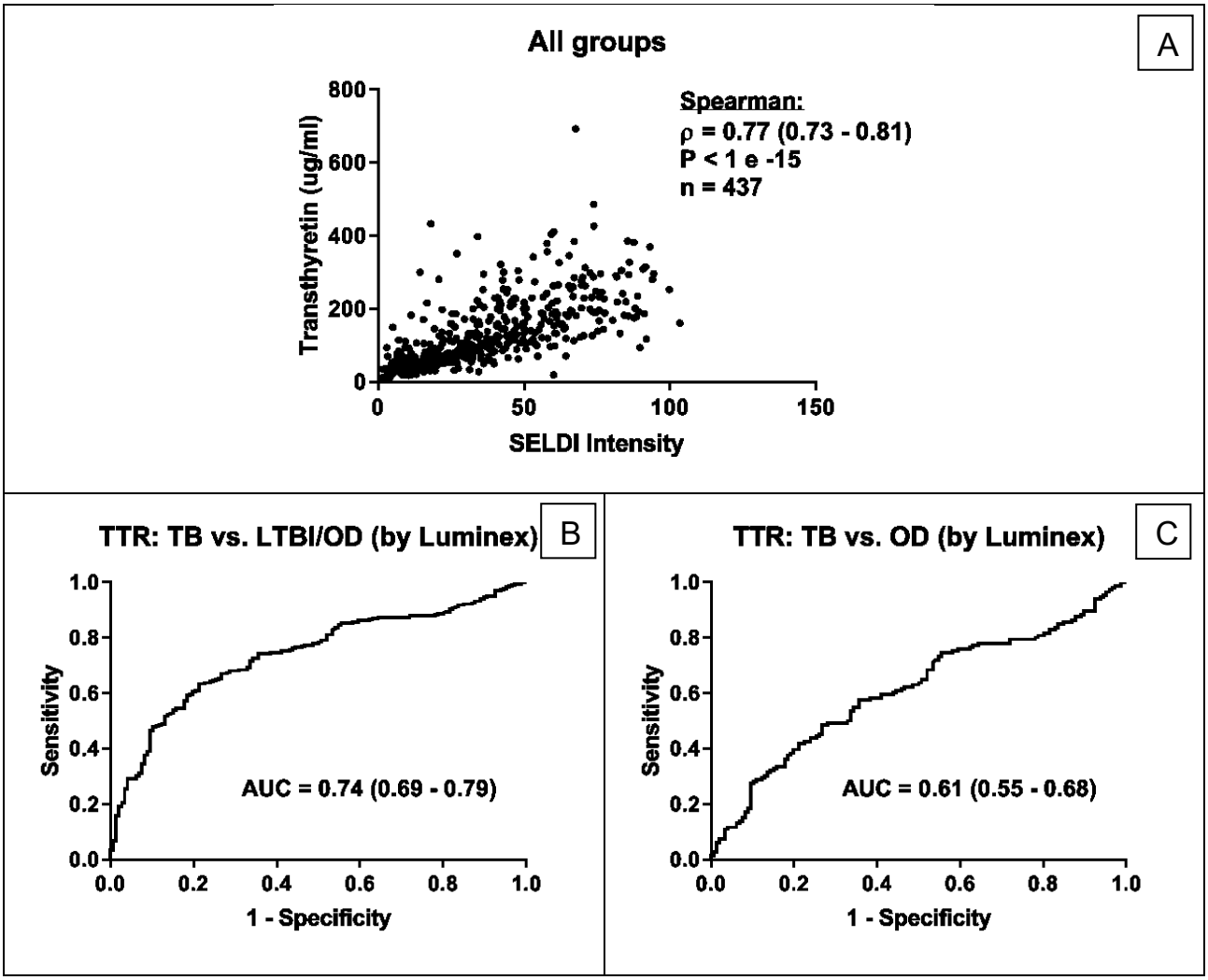


Figure 45 Correlation with SELDI intensity and diagnostic performance of transthyretin as measured by Luminex (N = 437).

A: correlation with SELDI intensities. B: ROC curve for distinguishing TB from LTBI and OD combined. C: ROC curve for distinguishing TB from OD alone.

Diagnostic Aim	Method	N	ROC AUC by Immunoassay	ROC AUC by SELDI
TB vs. LTBI/OD	**ELISA**	167	0.68 (0.60 – 0.76)	0.73 (0.65 – 0.81)
	Luminex	437	0.74 (0.69 – 0.79)	0.77 (0.73 – 0.82)
TB vs. OD	**ELISA**	111	0.55 (0.44 – 0.66)	0.65 (0.54 – 0.75)
	Luminex	292	0.61 (0.55 – 0.68)	0.68 (0.62 – 0.74)

Table 22 Areas under the ROC curves (ROC AUCs) for transthyretin for distinguishing TB from each of LTBI/OD and OD alone when measured by ELISA or Luminex.

In each case, the ROC AUC from the SELDI data on the corresponding sample of patients is shown for comparison (right hand column).

When the patients were stratified by either HIV status and study site, the performances of transthyretin by ELISA to distinguish TB from either LTBI/OD or OD were comparable, as shown in Tables 23 and 24. There was a trend towards transthyretin performing better in HIV uninfected individuals, and at Karonga as opposed to Cape Town, but 95% confidence intervals were overlapping in all cases.

Diagnostic Aim	HIV uninfected	HIV infected
TB vs. LTBI/OD	0.70 (0.59 – 0.81)	0.67 (0.54 – 0.79)
TB vs. OD	0.56 (0.40 – 0.72)	0.54 (0.39 – 0.70)

Table 23 ROC AUCS for transthyretin when measured by ELISA for distinguishing TB from each of LTBI/OD and OD in each of HIV uninfected and infected patients.

Bracketed values represent 95% confidence intervals.

Diagnostic Aim	Karonga	Cape Town
TB vs. LTBI/OD	0.71 (0.60 – 0.83)	0.65 (0.53 – 0.77)
TB vs. OD	0.61 (0.46 – 0.76)	0.51 (0.36 – 0.67)

Table 24 ROC AUCS for transthyretin when measured by ELISA for distinguishing TB from each of LTBI/OD and OD in patients from Karonga and Cape Town.

Bracketed values represent 95% confidence intervals.

ii) Beta-2-Microglobulin

Serum levels of beta-2-microglobulin (beta-2-M) were obtained by ELISA for a total of 172 patients. The correlation between beta-2-M levels by ELISA and SELDI intensity was weak (Figure 46A, below). Despite this, levels in TB patients were significantly higher than levels in the LTBI and OD groups combined (Figure 46C). As shown in Figure 46D, this difference was most significant in comparison with the LTBI group, which is in keeping with the fact that this biomarker was selected for the TB vs. LTBI/OD signature and not for the TB vs. OD signature.

The ROC AUC to distinguish TB from LTBI/OD was significant at 0.6 (0.51 – 0.69) (Figure 46E), however, as with transthyretin, this was lower than when calculated from the SELDI data on the same 172 patients (0.70 (0.63 – 0.78)). In contrast to the transthyretin ELISA data, however, the point estimate for the ROC AUC of beta-2-M from the ELISA data fell below the 95% confidence intervals for that from the SELDI data. Furthermore, the correlation with the SELDI intensities for beta-2-M by ELISA was significantly weaker than for transthyretin (95% confidence intervals not overlapping).

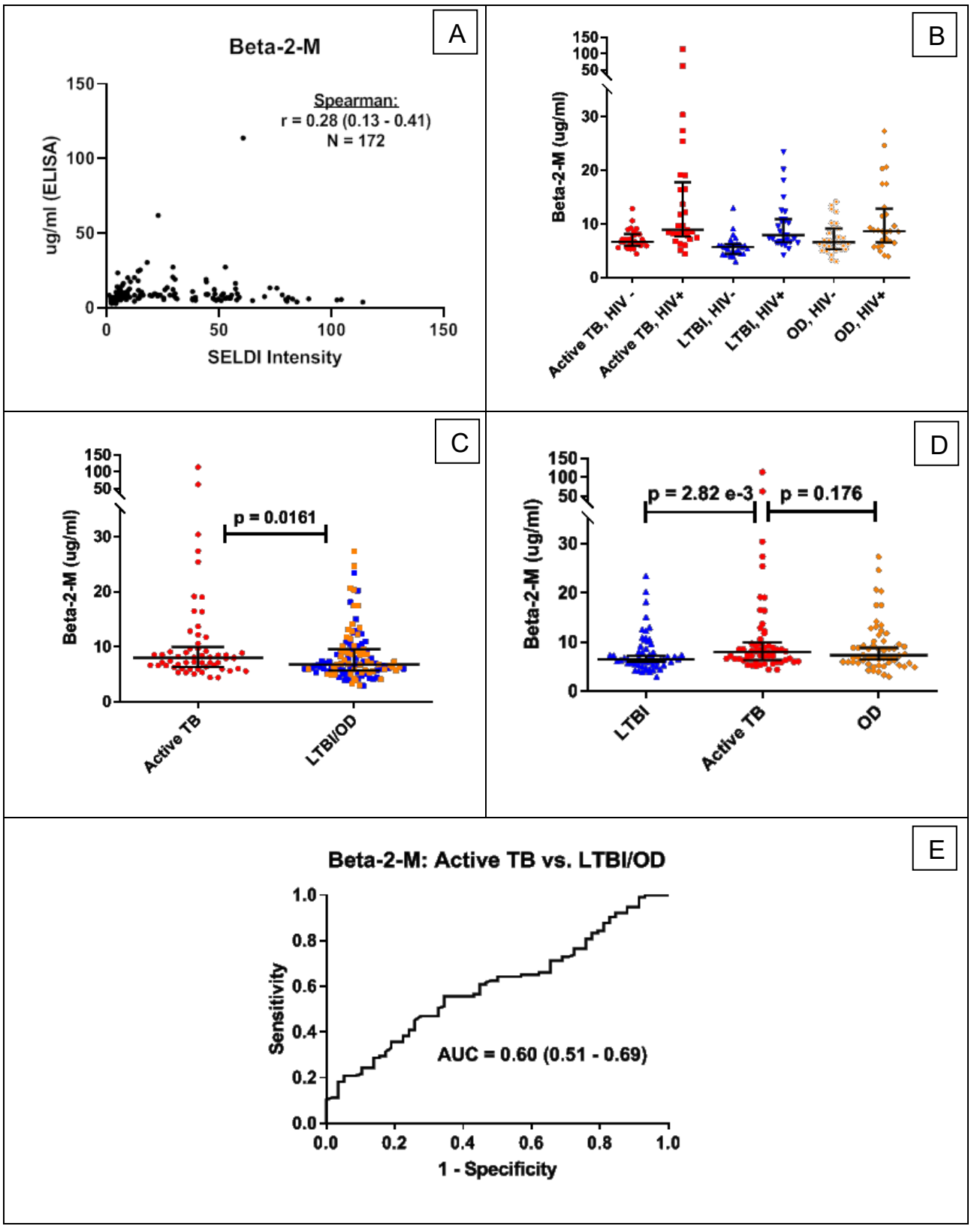


Figure 46 **Serum levels of beta-2-microglobulin (beta-2-M) as measured by ELISA (N = 172).**

A: correlation with SELDI intensity. B: levels in each of the six clinical subgroups. C: levels in active TB compared with LTBI and OD groups together, regardless of HIV status. D: levels in LTBI, active TB and OD groups, regardless of HIV status. E: ROC curve to distinguish active TB from LTBI and OD combined, regardless of HIV status. p-values in the box plots were derived from Mann-Whitney tests and are 1-sided. AUC: area under the curve. Bracketed values in the ROC curve panel represent 95% confidence intervals.

Levels of beta-2-M by ELISA were higher in patients infected with HIV. This was the case when all groups were considered together (p = 2.27 e-9) and within each of the TB (p = 8.34 e-5), LTBI (p = 3.93 e-7) and OD groups as well (p = 0.018). This is shown in Figure 47, below.

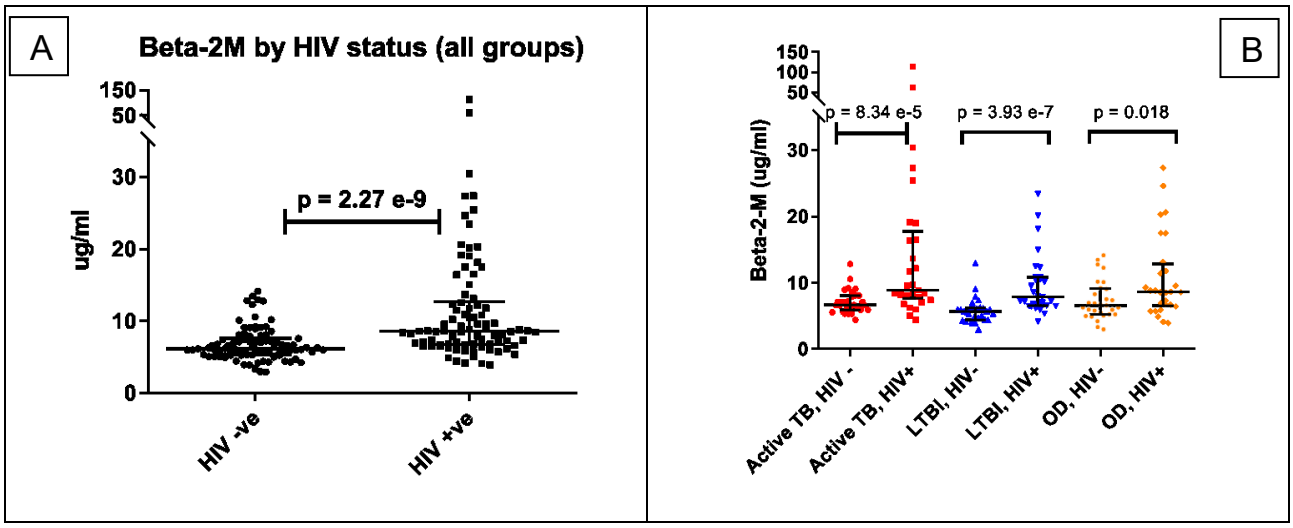


Figure 47 Comparison of levels of beta-2-microglobulin (beta-2-M) in HIV uninfected and infected individuals.

A: Levels in all groups. B: levels in each of TB, LTBI and OD. p-values were derived from Mann-Whitney tests and are 1-sided.

To explore possible reasons for the poor correlation between ELISA and SELDI data for beta-2-M, the distribution of SELDI intensities for beta-2-M was examined more closely. In the same 172 patients, intensities of beta-2-M were also higher in patients co-infected with HIV, but not as significantly as by ELISA (p = 3.96 e-3). Within the individual groups of TB, LTBI and OD, intensities were higher in patients with HIV in the LTBI (p = 3.71 e-5) and OD groups (p = 8.69 e-3), but not in the TB group, where the median was in fact slightly lower (two-sided p = 0.35). Even when all the TB patients whose sera were analysed by SELDI were considered (n = 219), levels of beta-2-M still trended towards being lower in the HIV infected (TB, HIV+) group than the uninfected (TB, HIV-). This was almost statistically significant from a two-sided Mann-Whitney test (p = 0.114). These findings are displayed below in Figure 48.

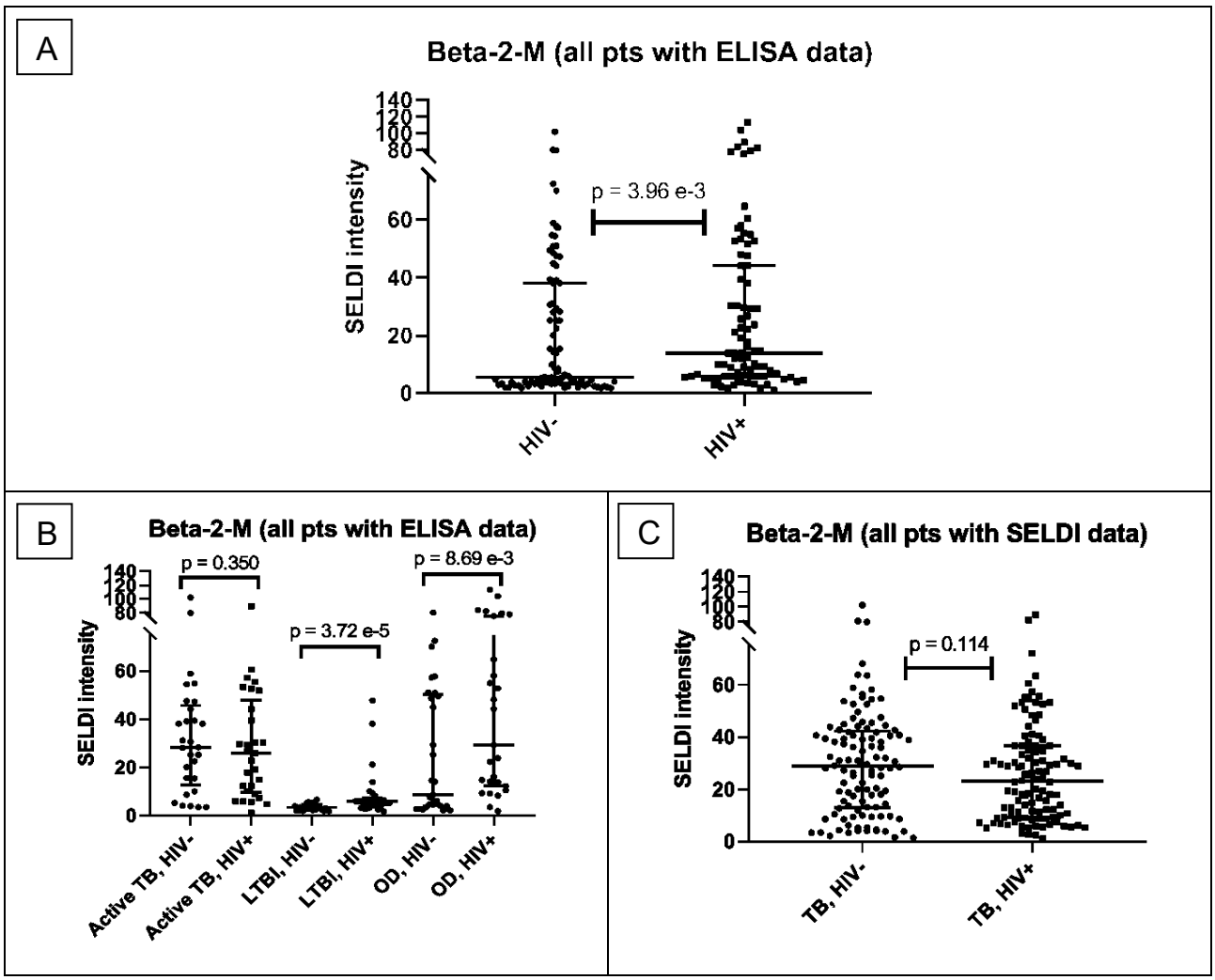


Figure 48 Serum levels of beta-2-M as measured by SELDI intensity and stratified by HIV status.

A: Beta-2-M intensities in the patients whose sera were assayed by ELISA, stratified by HIV status. B: Intensities in the patients whose sera were assayed by ELISA, stratified by clinical group and HIV status. C: Intensities in all TB patients in whom sera were assayed by SELDI, stratified by HIV status. HIV-: HIV uninfected. HIV+: HIV infected. p-values are derived from Mann-Whitney tests and are one-sided, with the exception of panel C in which a two-sided p-value is shown, since the difference in intensities between TB, HIV- and TB, HIV+ was in an opposite direction to what was expected from the ELISA data.

The ELISA kit used for beta-2-M was polyclonal. We therefore hypothesised that the difference of the impact of HIV status on beta-2-M level in the active TB patients when measured by ELISA vs. SELDI may reflect a difference in the forms of the protein that are being detected in each case. It seemed likely that this problem was a cause of the poor correlation between the ELISA and SELDI data. We therefore repeated the correlation analysis between ELISA levels and SELDI intensities after stratifying the patients by HIV status. We also calculated the ROC AUC for distinguishing TB from LTBI/OD for each of the HIV infected and uninfected halves of the cohort. Results are shown in Table 25 below.

	N	Correlation between beta-2-M level (ELISA) and SELDI intensity (rho)	ROC AUC to distinguish TB from LTBI/OD (ELISA data)	ROC AUC to distinguish TB from LTBI/OD (SELDI data)
HIV uninfected	87	0.40 (0.20 – 0.57)	0.64 (0.53 – 0.76)	0.77 (0.67 – 0.87)
HIV infected	85	0.11 (-0.11 – 0.32)	0.61 (0.48 – 0.73)	0.63 (0.50 – 0.75)

Table 25 Correlation of ELISA and SELDI data and diagnostic accuracy of beta-2-microglobulin (beta-2-M) levels by ELISA (ROC AUC) after stratifying the patients by HIV status.

TB, LTBI and OD patients were all included (N = 172). Correlation coefficients are from Spearman's rank correlations for non-parametric data.

In the HIV uninfected patients, the correlation between ELISA and SELDI data was moderately strong and the ROC AUC was higher than when HIV infected patients were included. The point estimate for ROC AUC by ELISA still fell below the 95% confidence interval of the ROC AUC by SELDI, however. In HIV infected patients, there was no longer a statistically significant correlation and the ROC AUC was also no longer significant. The point estimate for the ROC AUC in HIV infected patients (0.61) was higher than it had been with both HIV uninfected and infected patients together (0.60, Figure 46E), however, and was well within the 95% confidence interval of the ROC AUC by SELDI for these patients.

If the poor ROC AUC of beta-2-M by ELISA compared to by SELDI had purely been due to the difference in the impact of HIV co-infection on beta-2-M levels in TB (i.e. TB, HIV- vs. TB, HIV+), then we would've expected the ROC AUC in HIV uninfected patients to have been closer to the corresponding ROC AUC derived from the SELDI data. As described already, however, this was not the case (Table 25). There must therefore have been another factor which accounted for the poor ROC AUC by ELISA which affected the HIV uninfected patients as well. On re-inspection of the data, the spread of levels appeared to be greater in the HIV infected groups compared to their uninfected counterparts (Figure 47B, above). This might have been one factor which preferentially preserved the performance of beta-2-M by ELISA in the HIV infected half of the cohort, since any given variability in the ELISA process would be expected to affect the ROC AUC to a lesser degree if the spread of the levels themselves is greater. We therefore compared the interquartile ranges (IQRs) of beta-2-M between HIV uninfected and infected individuals across the three clinical groups. As shown below in Table 26, the IQRs appeared substantially greater in HIV infected individuals in all three clinical groups.

	IQR in HIV – (ug/ml)	IQR in HIV + (ug/ml)
Active TB	2.2	10.1
LTBI	1.8	4.3
OD	3.9	6.3

Table 26 Comparison of interquartile ranges of beta-2-M levels as measured by ELISA between HIV uninfected and infected individuals.

Each of the three clinical groups is considered in turn. IQR: interquartile range. HIV-: HIV uninfected. HIV+: HIV infected.

A structural variant of beta-2-M deficient in lysine-58 (ΔK58-b2m) has been detected in patients with various different medical disorders (kidney dialysis, autoimmune disease and small cell lung cancer).(4-6) It is possible

that this variant, or another one, arises in HIV/TB, but is still detected by polyclonal ELISA assays. This would account for the higher levels of beta-2-M in TB, HIV+ vs. TB, HIV- when measured by ELISA, but not by SELDI since the structural variation would lead to a different m/z value. To test this hypothesis, we re-inspected the SELDI data to see if any nearby peaks of slightly lower m/z value (approximately 11,600 m/z according to Trenchevska *et al*) had also been depleted by immunoprecipitation of the 11,732 m/z peak previously ascribed to beta-2-M.(6) On the spectra from the chromatographic SELDI surface on which the 11,732 m/z peak had been discovered ("CM10", i.e. weak anionic exchange chromatography, and at pH 4.0), there were no nearby peaks that were co-depleted.

In terms of the performance of beta-2-M by ELISA for distinguishing TB from LTBI/OD by HIV status and site, the ROC AUCs for HIV uninfected and infected patients have been shown above in Table 25. Regarding site, performance was good at Cape Town but was not significant at Karonga. These results are shown below in Table 27.

	Karonga	Cape Town
ROC AUC	0.51 (0.38 – 0.64)	0.70 (0.59 – 0.81)

Table 27 ROC AUCS for beta-2-microglobulin when measured by ELISA for distinguishing TB from LTBI/OD in patients from Karonga and Cape Town.

Bracketed values represent 95% confidence intervals.

iii) Alpha-1-antichymotrypsin

Serum levels of alpha-1-antichymotrypsin (ACT) were obtained by ELISA for a total of 188 patients. As explained above, ACT was selected for both signatures (i.e. TB vs. LTBI/OD and TB vs. OD). Data relating to each are shown separately below. In Figure 49, the correlation plot, box plots and ROC curve relating to translation of the 4,357 m/z biomarker to ELISA to distinguish TB from LTBI/OD are shown. In Figure 6, the equivalent graphs are shown for translation of the 4,468 m/z biomarker to ELISA for distinguishing TB from OD. Table 28 shows the ROC AUCs from ELISA and SELDI for each biomarker.

The correlation between levels of ACT by ELISA and SELDI intensity for the 4,357 m/z biomarker (TB vs. LTBI/OD signature) was strong (rho = 0.66 (0.57 – 0.77)) (Figure 49A). As shown in the Figure 49C, by ELISA, levels of ACT were significantly higher in the TB group than the LTBI/OD groups. This is in keeping with the SELDI data. The corresponding ROC AUC was high at 0.76 (0.69 – 0.83) (Figure 49D), although the point estimate fell below the 95% confidence intervals for the ROC AUC from the SELDI data on the same 188 patients (Table 28).

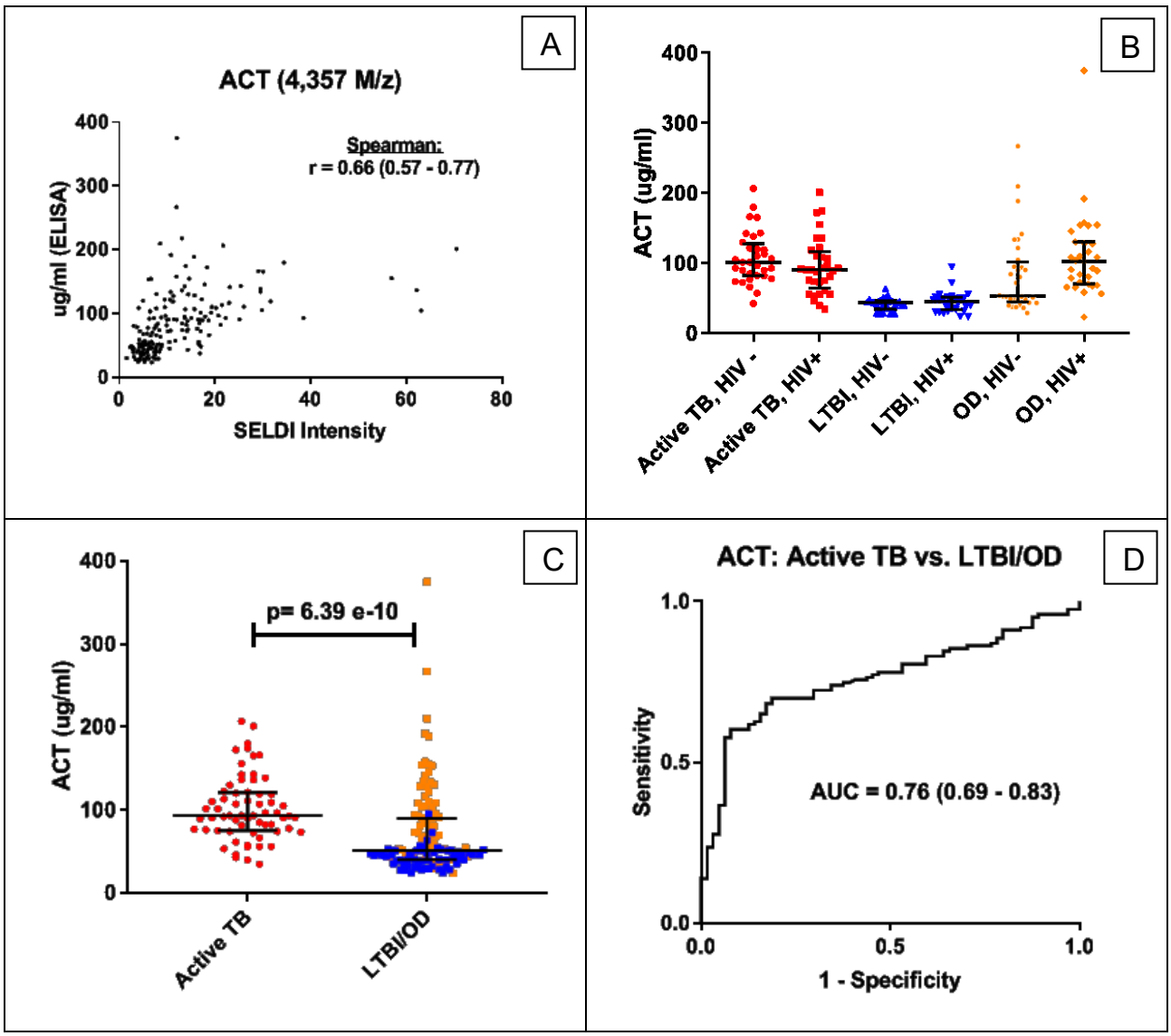


Figure 49 Translation of the 4,357 m/z alpha-1-antichymotrypsin (ACT) fragment to ELISA to distinguish TB from LTBI/OD.

A: correlation plot of serum ACT levels by ELISA with the 4357 m/z SELDI biomarker for TB vs. LTBI/OD (N = 188). B: ACT levels by ELISA in each of the six clinical groups. C: levels in active TB compared with LTBI and OD groups combined, regardless of HIV status. D: levels in LTBI, active TB and OD groups separately, regardless of HIV status. E: ROC curve for use of ACT level by ELISA for distinguishing TB from LTBI and OD combined. p-values in the box plots were derived from Mann-Whitney tests and are 1-sided. AUC: area under the curve. Bracketed values in the ROC curve panel represent 95% confidence intervals.

The correlation between levels of ACT by ELISA and SELDI intensity for the 4,468 m/z biomarker (TB vs. OD signature) was weak (rho = 0.38 (0.21 – 0.52)) (Figure 50A, below). In association with this, levels of ACT by ELISA trended towards being higher in TB than OD (as in the SELDI data) but this did not reach statistical significance (Figure 50B). The corresponding ROC AUC was also low and was not statistically significant (0.58 (0.48 – 0.68)) (Figure 50C). Furthermore, the point estimate for the ROC AUC by ELISA fell below the 95% confidence intervals for the ROC AUC by SELDI on the same 127 patients. In addition, the confidence intervals for the ROC AUC by each of the two methods only just overlapped (Table 28, below).

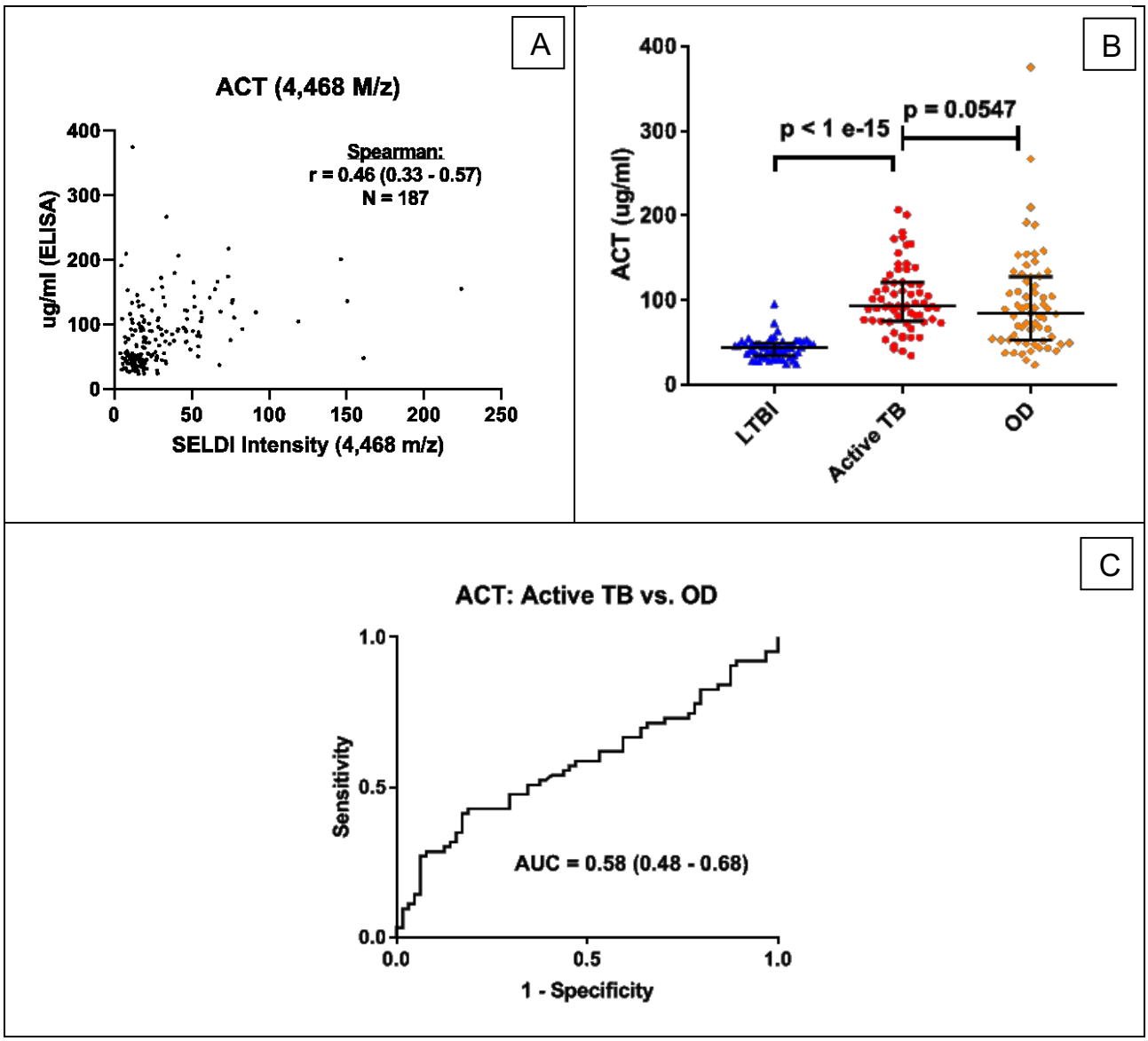


Figure 50 Translation of the 4,468 m/z alpha-1-antichymotrypsin (ACT) fragment to ELISA to distinguish TB from OD.

A: correlation plot of ACT levels by ELISA with the 4,468 m/z SELDI biomarker for TB vs. OD (N = 187). B: ROC curve for use of ACT level by ELISA to distinguish TB from OD. p-values in the box plots were derived from Mann-Whitney tests and are 1-sided. AUC: area under the curve. Bracketed values in the ROC curve panel represent 95% confidence intervals.

As shown below in Table 28, the correlation of the ELISA and SELDI data was significantly weaker for the 4,468 m/z biomarker than the 4,357 m/z (95% confidence intervals not overlapping). In addition, whilst the ROC AUCs by ELISA were lower than those by SELDI in both cases, in the case of the 4,468 m/z marker, this difference appeared to be greater.

	N	Correlation with levels by ELISA (rho)	ROC AUC by ELISA	ROC AUC by SELDI
m/z 4,357	188	0.66 (0.57 – 0.77)	0.76 (0.69 – 0.83)	0.84 (0.78 – 0.90)
m/z 4,468	127	0.46 (0.33 – 0.57)	0.58 (0.48 – 0.68)	0.75 (0.66 – 0.83)

Table 28 Correlation of ELISA and SELDI data for each of the two alpha-1-antichymotrypsin fragments and corresponding ROC AUCs by each of ELISA and SELDI.

Rho values represent Spearman's rank correlation coefficient. M/z: mass to charge ratio. Brackets represent 95% confidence intervals.

To explore possible reasons for this, the distributions of SELDI intensities in each clinical group were examined more closely. In the ELISA data, levels in the OD, HIV+ were higher than in OD, HIV- ($p = 3.59$ e-3) (Figure 49B). This was not the case for either of the fragments in the SELDI data, in fact the intensities trended towards being lower in the OD, HIV+ group in both cases (two-sided $p = 0.35$ and 0.45, respectively) (Figures 51A and 51B).

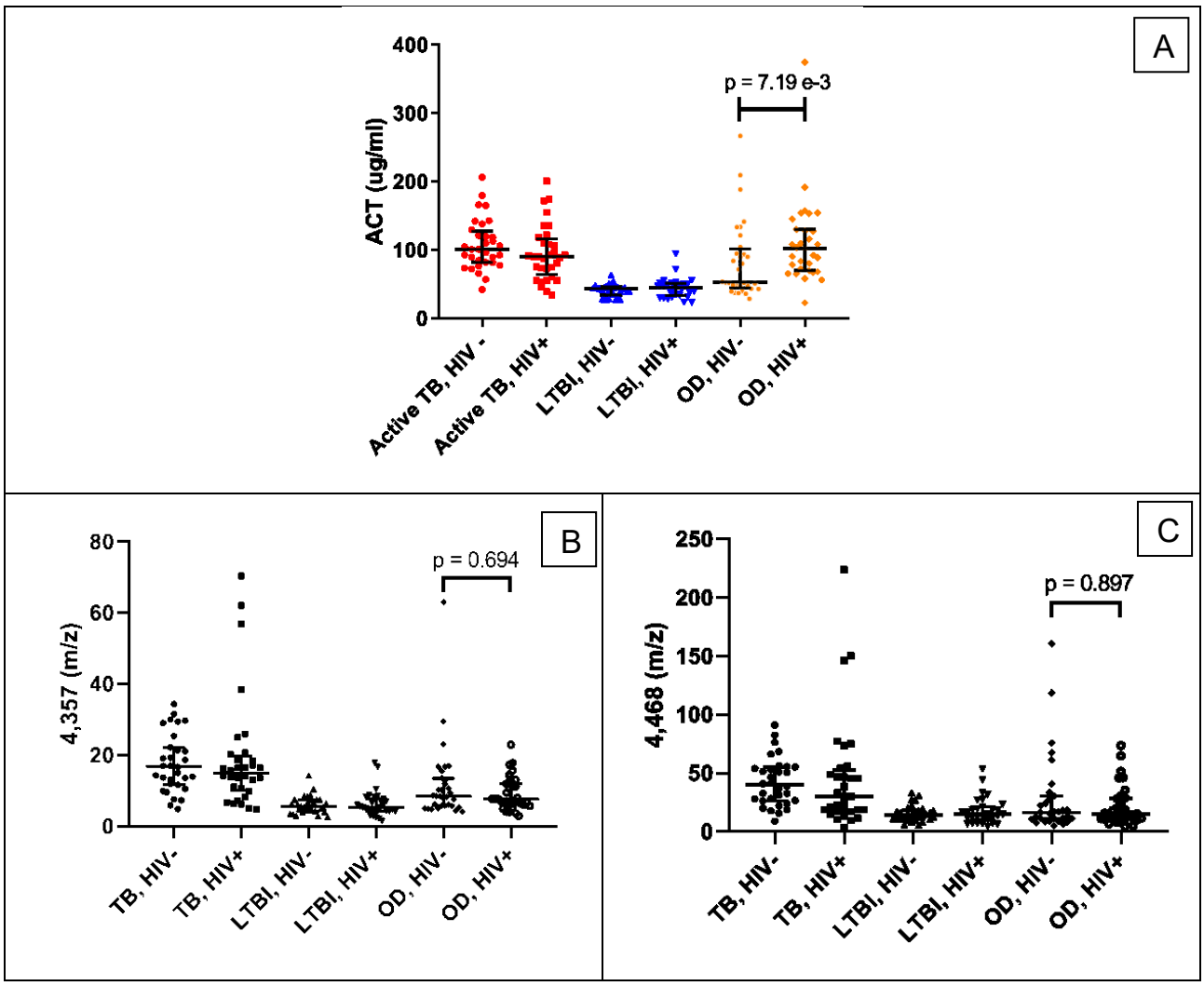


Figure 51 Box plots showing serum levels of alpha-1-antichymotrypsin (ACT) in each of the six clinical groups by ELISA and SELDI.

P-values for the comparison of levels between patients with other diseases (OD) with and without HIV infection are shown. A: levels of ACT by ELISA. B: levels of the 4,357 m/z SELDI biomarker. C: levels of the 4,468 m/z SELDI biomarker. P-values were derived from two-sided Mann-Whitney tests. For B and C, units on the y axis represent SELDI intensity.

As described by the manufacturer, the ELISA used for these assays was designed to detect the full-length protein and used lyophilised full-length protein as the standard (product details in Chapter II, Section 2). It would therefore be expected to have detected the full-length ACT protein, as well as any fragments. To test our hypothesis that the ELISA was detecting both fragments and full-length protein, and that all of these are elevated in TB compared to LTBI (as well as in OD, HIV+), we calculated the ROC AUC for TB vs. LTBI by each of ELISA and SELDI (using the m/z 4,357 biomarker). The AUC was greater using the ELISA data (0.95 (0.91 – 0.99)) than the SELDI data (0.91 (0.86 – 0.96)).

3. Discussion

Even aside from the complex and often lengthy process of protein identification that were beyond the scope of this thesis, translating protein biomarkers from SELDI or other forms of proteomic mass spectrometry (MS) to immunocapture is a complex process. There are multiple sources of variability in both MS and immunocapture, including non-operator-related factors, as well as the stereochemical challenge of detecting and quantifying the exact protein isoform or fragment using an antibody that was selected from the MS spectra. Sources of variation inherent in both of these methods will now be considered in turn.

From a MS protein quantification point of view, serum is a complex mixture. For the EU-TB project, since any resulting POCT would analyse whole blood, no upstream fractionation steps such as depletion of the top 10 or 12 most abundant proteins were performed. Sera were analysed on five different types of chromatographic array (see Introduction IV part ii)), but nonetheless there would have been thousands of proteins (of widely varying abundance) competing for ionisation and subsequent detection on each of those surfaces. Levels of all serum proteins are dynamic, affected by the disease severity and also other factors specific to each patient (e.g. age, gender, metabolic and nutritional state, and other comorbidities). Ionisation of potential biomarkers can therefore be "cross-suppressed" by other proteins whose levels vary from patient to patient, leading to an alteration in the quantification of the target in a way that wouldn't occur in immunocapture. This so called "matrix effect" is impossible to overcome completely, even after effectively fractionating the sera by running each sample on multiple different arrays.

With regards to immunocapture, there are also multiple possible sources of variation. Firstly, workflows of platforms such as ELISA or Luminex involve multiple steps during which error can be introduced. Secondly, some patients' serum aliquots may have been through an extra freeze-thaw cycle due to having been re-run at a different dilution after their optical density (OD) fell above or below the standard curve. Thirdly, there may be non-specific binding of primary or secondary antibodies to other proteins with similar epitopes to the target protein, resulting in falsely high ODs. Fourthly, a matrix effect may also occur, due in this instance for example to excessive numbers of lipid particles physically interfering with the sequential binding of assay components with one another. Finally, even if a series of assays is done in parallel, the nature of manufacturers' supplies may mean that not all the plates in the series come from the same lot. The resulting lot-lot variation can be difficult to fully correct.

As illustrated in the results above, a further challenge arises from the fact that proteins selected from SELDI spectra are selected on the basis of a one single peak (i.e. one m/z species). That peak may represent the full-length protein, or alternatively may be an isoform or even fragment of that protein. Isoforms, or variants, may arise through alternative splicing of messenger RNA, or post-translational modifications (PTMs). There are multiple types of PTM, one being enzymatic cleavage which could result in removal of one or more amino acids from the termini or a specific site and may also generate specific fragments. By their nature, such processes will alter the molecular weight and mean either that there are fragments or isoforms circulating that are not intended to be detected at immunocapture (but are), or, if the fragment is the biomarker, then it is the detection of the full-length protein that must be avoided.

The results for transthyretin overall were good. By ELISA, the correlation with SELDI intensity was moderately strong, and the ROC AUCs for each of the two diagnostic aims (TB vs. LTBI/OD and TB vs. OD) fell within the 95% confidence intervals of the corresponding ROC AUCs by SELDI. The ROC AUC by ELISA for TB vs. OD had a 95% confidence interval that crossed 0.5, however it was very similar to the ROC AUC by SELDI. Furthermore, transthyretin was the third protein in the signature to distinguish TB vs. OD, and so in view of the fact that the variable selection method used (FSPLS) selects biomarkers that by their sequential addition to the model result in an incremental improvement to the performance overall, it is not necessarily expected that each biomarker (after the first one) would have a large ROC AUC when used on its own.

The fact that we had also quantified transthyretin by Luminex made for a very helpful comparison of the two immunocapture platforms. By Luminex, the correlation with SELDI intensity and both ROC AUCs were better than they were with ELISA. Luminex has a wider dynamic range than ELISA, (Ooi, '06 and Song, '05) hence it is a less frequent occurrence to need to re-run sera at different dilutions. This minimises two different types of variation. Firstly, the number of freeze-thaw cycles is minimised, which can affect the structural integrity of many proteins, including transthyretin.(7) Secondly, there may be less lot-lot variation due to further kits needing to be purchased, which has been shown to affect accuracy of both ELISA and Luminex assays.(8, 9) In head-head comparisons, Luminex is reliably at least as accurate as ELISA, and sometimes more accurate.(10-13) Regardless of which specific factors led to the superior performance of the Luminex assay over the ELISA used in this study, the excellent performance of transthyretin as a biomarker when quantified by Luminex is relevant to the second Primary Aim of this Thesis, which was to "to re-test the proteins from the EU-TB signatures by immunocapture". We were not able within the scope of this work to assess performance of transthyretin by lateral flow immunoassay (LFA), which would likely be the final platform used at POC. However, LFA has been shown to be accurate across a dynamic range of 1000-fold, which is greater than the range of the standard curve used in our commercially-produced transthyretin ELISAs.(14) Assessment of performance of transthyretin to detect TB by LFA in our cohort is now warranted.

The results for beta-2-M were good, albeit less straightforward than for transthyretin. The correlation with SELDI intensity was weak, and whilst the ROC AUC was statistically significant 0.6 (0.51 – 0.69), it fell below the 95% confidence intervals for the AUC by SELDI. Various possible explanations for this were explored. Firstly, it was noted that levels of beta-2-M as measured by ELISA were higher in individuals infected with HIV, both across all groups and within each clinical group. This had not been seen in the SELDI data however, since intensities of beta-2-M actually trended towards being lower in TB, HIV+ than TB, HIV-, including when all patients in the EU-TB study were considered. Stratification of the patients by HIV status revealed a stronger correlation with the SELDI data in the HIV uninfected half of the cohort, and an almost absent correlation in the HIV infected half. This was in keeping with the observed difference in pattern of levels between TB, HIV- and TB, HIV+ between assay modalities. The ROC AUCs in each of the HIV uninfected and infected halves, however, were surprising, in that although the ROC AUC in HIV uninfected patients was higher than it was in HIV infected patients, in relation to the SELDI data, beta-2-M by ELISA performed better in the HIV infected patients. It therefore seems likely that the main contributor to the poor performance of beta-2-M by ELISA as compared with SELDI is linked to the compact distribution of levels in patient sera. This may make ELISA an unforgiving method to quantify beta-2-M for the purposes of a diagnostic test, since any form of assay variability will hinder the ROC AUC to a greater extent. The fact that the performance of beta-2-M by ELISA deteriorated

least in the HIV infected half of the cohort is in keeping with this, since the IQRs were appreciably wider than in the HIV uninfected clinical groups. Performance of beta-2-M by LFA remains to be seen and will likely depend upon the accuracy of such a platform. As to the reason for the higher levels in TB, HIV+ vs. TB, HIV- by ELISA but not by SELDI, we were not able to identify a reason based on our existing SELDI spectra, however it seems plausible that another isoform or fragment of beta-2-M is circulating in HIV/TB that was detected by the ELISA platform but did not manifest as the same peak on the SELDI. Further investigation of this is warranted as it may shed light on the disease process in HIV/TB.

The translation of the ACT SELDI biomarkers to immunocapture was the most complex of the three tested in this thesis but illustrates an important challenge. ACT was identified as the top biomarker for both EU-TB signatures, but clearly as fragments and a different fragment in each case. As a first step, a commercially available polyclonal ELISA kit was used, with some degree of expectation that the results might be suboptimal, however the full-length protein is above the optimal molecular weight range of SELDI. As shown above, this ELISA assay did not reproduce the pattern that was seen by SELDI, in that the levels were clearly higher in OD, HIV+ than OD, HIV-, whereas this was not the case by SELDI. This explains the poor performance of ACT as measured by ELISA as a biomarker in either signature, with the performance in the TB vs. LTBI/OD signature being better than in the TB vs. OD signature simply due to the fact that the levels in the LTBI groups were low by either method (the full-length protein isn't elevated in LTBI as it is an acute phase reactant). ROC AUC for distinguishing TB from LTBI (albeit not an actual aim of the project) was greater by ELISA than by SELDI, which suggests that serum levels of both the full-length protein and fragments are elevated in TB, but of only the full-length protein in OD. Clearly another approach will have to be tried, for example an assay based on monoclonal antibodies specific to the non-terminal ends of each fragment which wouldn't be exposed in the full-length protein, or an aptamer designed for the same purpose.

In summary, the diagnostic performance of transthyretin level in serum translated well from SELDI to immunocapture, but this was not the case for beta-2-M or ACT. Regarding the latter two, however, it is possible that with further optimisation of the immunocapture process, exclusive detection of the exact biomarker that was selected from the SELDI spectra can be achieved. This is certainly feasible for ACT, and possibly for beta-2-M as well, depending on whether there is an isoform circulating or another proteomic explanation for the poor correlation that can be overcome.

CHAPTER VI RESULTS: host serum proteins and the biology of TB

1. Introduction

As stated in the Introduction Chapter, changes to the biological state of the host in TB can be exploited either for biomarker discovery, or for understanding more about the pathogenesis of the disease. This chapter focuses on what can be learned from the concentrations of the selected serum proteins in light of the clinical and demographic information available. Broadly speaking, the latter includes disease category (i.e. TB, OD or LTBI) and HIV status, and then within the clinical category of TB, several other variables on which data was collected. These included age, gender, symptoms, symptom duration, history of previous TB, temperature, body mass index (BMI), smear grade and HIV status. For HIV co-infected patients (HIV+), CD4+ count and antiretroviral therapy (ART) status were also recorded. We sought to make use of this information for the dual objectives of acquiring new knowledge about the biology of human TB, and to find novel biomarkers of severity, since BMI, smear grade and (in the case of HIV/TB) CD4+ count are all known to associate with outcome of TB treatment.(1-7)

Rationales for including each of the clinical variables were developed and are shown in Table 29 below, together with hypotheses and rates of missing data. Symptoms and symptom duration were excluded as they were felt to be subjective and hence unreliable. Temperature was excluded as it was a one-off measurement and therefore unlikely to be meaningful in the context of culture positive TB. Previous TB was excluded as it was not possible to distinguish between re-infection and relapse, each of which might entail different changes to the immune response. Antiretroviral status was excluded due to very small numbers on ART (n = 10).

For this thesis, since data were gathered on a much larger number of patients for the 22 proteins from the AE-TBC study than they were for the three proteins from the EU-TB study, we decided to focus on the former for the purposes of finding novel associations with clinical variables.

Results for associations between the 22 AE-TBC biomarkers and “TB disease state” (i.e. TB vs. LTBI) will be shown first. Within “The TB disease state”, we will examine the changes in TB in comparison with LTBI, without the influence of HIV, and then the difference that HIV co-infection makes to the response to TB. Results of associations between the clinical variables and protein levels in TB will then be shown, with those variables that are known to associated with severity shown first: smear grade, BMI and (in the case of HIV/TB) CD4+ count.

	Reason(s) for selecting variable	Hypothesis(/es) re. associations with serum protein levels	No. missing values (%)
Age	Older age confers susceptibility to TB(8), hence identifying associated proteins may inform biology.	Older age will be associated with a greater disturbance to the proteome.	0 (0)
Gender	Gender impacts the immune response to TB however the mechanisms of this are poorly understood.(9)	Levels of some proteins will differ by gender.	0 (0)
BMI	Low BMI is a marker of severity in TB,(10-13) hence biomarkers of this may inform treatment monitoring; Low BMI is a risk factor for TB,(14) hence identifying associated proteins will inform understanding of susceptibility.	Lower BMI will be associated with a greater disturbance to the proteome.	10 (4.5)
Smear	Smear grade is a marker of severity in TB,(3-6) hence biomarkers of this may inform treatment monitoring and understanding of biology.	Higher smear grades will be associated with greater disturbance to the proteome.	4 (1.8)
CD4+	CD4+ is a marker of severity in HIV/TB(15-18), hence biomarkers of this may inform treatment monitoring and understanding of biology.	Lower CD4+ count will be associated with greater disturbance to the proteome.	12 (10.8)

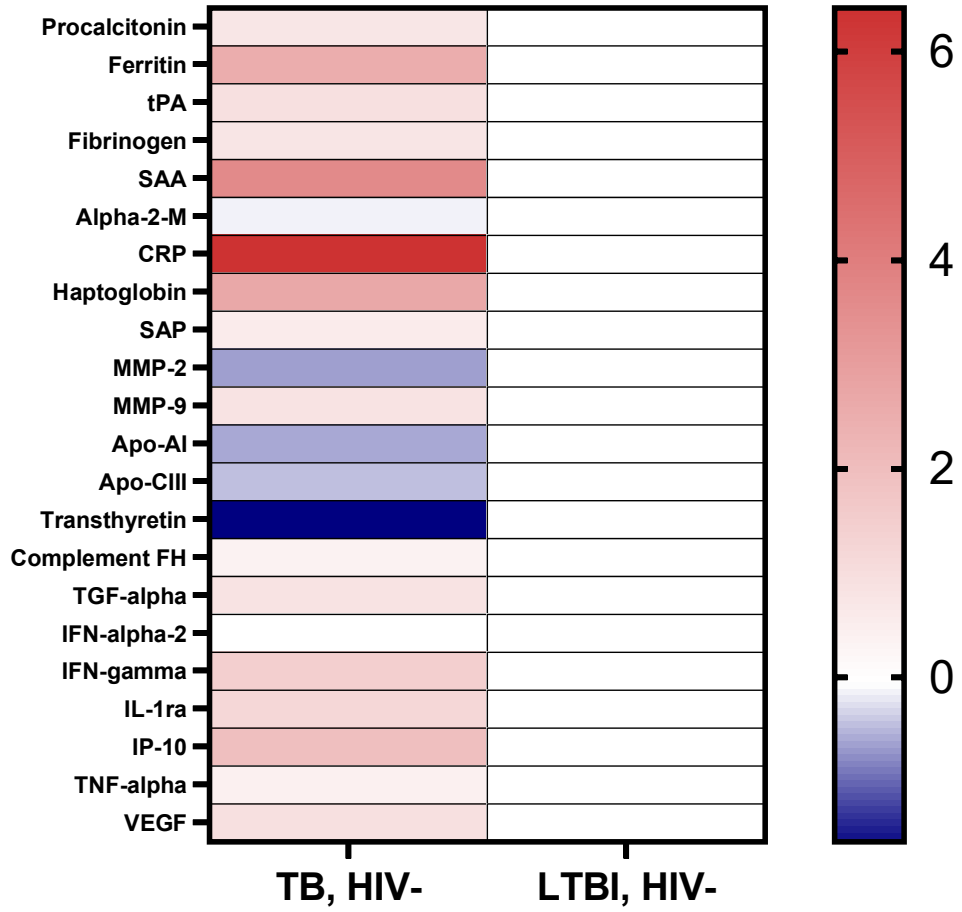
Table 29 A list of the clinical variables that were included for analysis of association with the AE-TBC biomarkers within the TB group.

Left-hand column: reason(s) for selecting clinical variable; Central column: hypothesis(/es) relating to possible associations between clinical variables and serum protein levels ("the proteome"); right-hand column: quantity of missing data (numbers relate to the entire EU-TB cohort of active TB patients (n = 220), except for CD4+ count which relates to TB+HIV+ patients only (n = 111)). BMI: body mass index.

2. The TB disease state and the impact of HIV

i) The TB disease state

To study the change to the serum proteome in TB, differences in protein concentrations were analysed between the TB, HIV- and LTBI, HIV- groups. HIV co-infected patients were excluded in order to remove any contribution of HIV disease progression to the observed changes. Results are shown first of all by virtue of the fold change in median levels in each of the TB and OD groups relative to the LTBI group (Heat Map 2), and then as medians and interquartile ranges in Table 30, with results of significance testing. As shown in the heat map, the protein whose levels were highest in TB, HIV- relative to LTBI, HIV- was CRP; and the protein whose levels were lowest was transthyretin.



Heat Map 2. Differences in median protein levels between TB, HIV- and LTBI, HIV- groups. Differences were plotted as log-tranformed fold-changes (base 2) in comparison to the LTBI, HIV- group. The legend on the right shows the colour scale in relation to log fold-change. Red: increased level; blue decreased.

Of the 22 proteins whose concentrations were measured, all but three were significantly different in TB compared to LTBI, in HIV- individuals (Table 30, below). Those that were not different were alpha-2-M, IFN-alpha-2, and IL-1ra. Of the 19 that were different in TB, the direction of change was as expected in the light of the raw data, namely that all proteins were higher except for apo-AI, apo-CIII and transthyretin, whose concentrations are known to decrease in acute inflammation and TB ("negative biomarkers"), and MMP-2.

	TB, HIV-		LTBI, HIV-			
Protein	**Median**	**IQR**	**Median**	**IQR**	**P-value**	**Corrected P**
Procalcitonin	**1.90**	**1.18 - 2.45**	**1.11**	**0.73 - 1.79**	**0.00**	**0.00**
Ferritin	**61.88**	**29.92 - 112.5**	**10.62**	**5.67 - 19.43**	**0.00**	**0.00**
tPA	**2.49**	**1.7 - 3.08**	**1.29**	**0.91 - 1.92**	**0.00**	**0.00**
Fibrinogen	**2.63**	**1.77 - 3.11**	**1.52**	**0.97 - 2.06**	**0.00**	**0.00**
Serum amyloid A	**6.45**	**4.52 - 7.84**	**0.52**	**0.19 - 1.35**	**0.00**	**0.00**
Alpha-2-M	1.53	1.36 - 1.89	1.62	1.28 - 2.02	0.32	1.00
CRP	**63.84**	**26.27 - 95.66**	**0.75**	**0.42 - 1.97**	**0.00**	**0.00**
Haptoglobin	**7.10**	**0.39 - 13.8**	**1.08**	**0.17 - 2.84**	**0.00**	**0.00**
Serum amyloid P	**46.31**	**36.22 - 54.17**	**30.35**	**25.75 - 37.15**	**0.00**	**0.00**
MMP-2	**82.99**	**65.01 - 105.7**	**125.80**	**105.5 - 160.8**	**0.00**	**0.00**
MMP-9	**479.00**	**290 - 614.9**	**263.00**	**174.6 - 423.2**	**0.00**	**0.00**
apo-AI	**241.50**	**171.8 - 308.7**	**351.80**	**271.3 - 436.5**	**0.00**	**0.00**
apo-CIII	**90.91**	**68.78 - 115.7**	**120.10**	**90.27 - 149.7**	**0.00**	**0.00**
Transthyretin	**70.97**	**51.44 - 106.3**	**212.20**	**169.9 - 295.9**	**0.00**	**0.00**
Complement factor H	**413.30**	**356.1 - 471.7**	**314.80**	**289 - 355.7**	**0.00**	**0.00**
TGF-alpha	**13.66**	**7.9 - 23.69**	**7.49**	**3.46 - 12.09**	**0.00**	**0.00**
IFN-alpha-2	0.00	0 - 20.95	0.00	0 - 15.99	0.32	1.00
IFN-gamma	**13.60**	**5.44 - 28.73**	**4.87**	**2 - 18.56**	**0.00**	**0.01**
IL-1ra	15.82	1.1 - 54.29	6.65	0 - 27.14	0.03	0.47
IP-10	**0.76**	**0.43 - 1.11**	**0.19**	**0.13 - 0.26**	**0.00**	**0.00**
TNF-alpha	**18.13**	**12.47 - 28.58**	**13.32**	**9.16 - 18.69**	**0.00**	**0.00**
VEGF	**307.20**	**130.3 - 810.7**	**158.50**	**78.8 - 264.7**	**0.00**	**0.00**

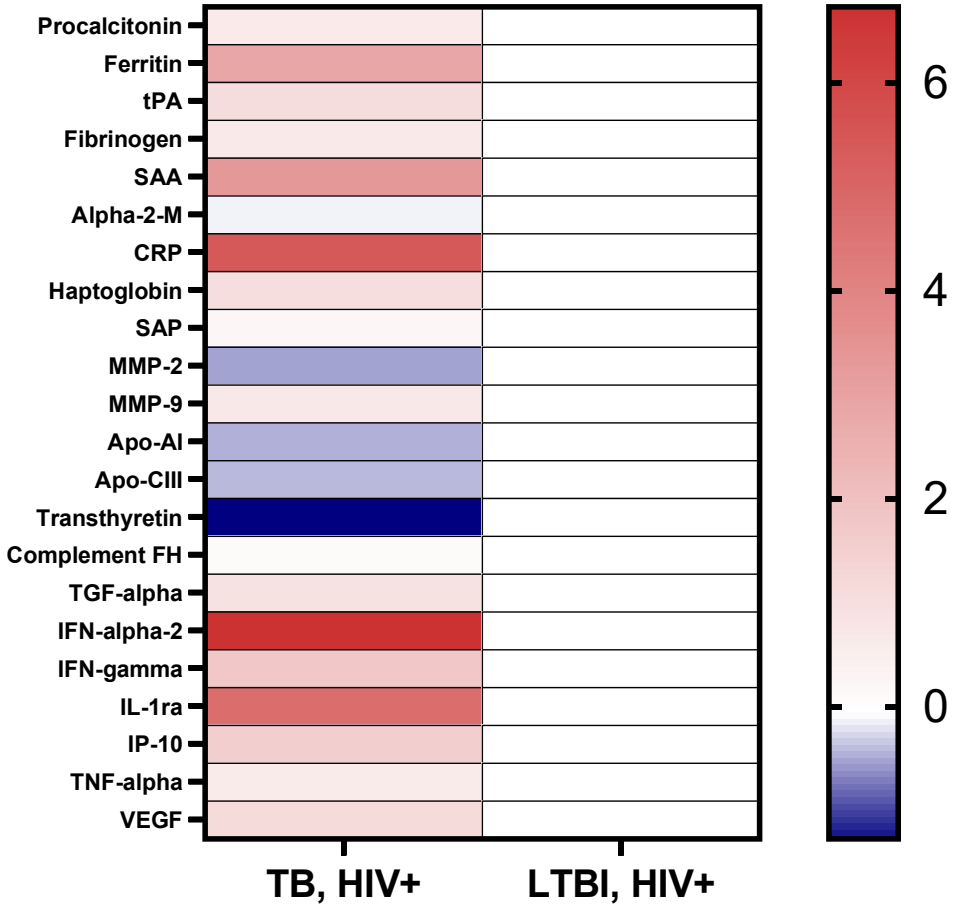
Table 30 A comparison of serum protein concentrations in TB with those in LTBI, focusing on HIV uninfected patients.

Bold font represents proteins whose concentrations were significantly different in TB after p-value correction. Green font represents proteins whose concentrations were higher in TB, red font for those that were lower. Units of protein levels are as follows: mg/ml: Alpha-2-M and Haptoglobin; ug/ml: fibrinogen, serum amyloid A, CRP, serum amyloid P, MMP-2, MMP-9, apo-AI, apo-CIII, transthyretin and complement FH; ng/ml: procalcitonin, ferritin, tPA and IP-10; and pg/ml: TGF-alpha, IFN-alpha-2, IFN-gamma, IL-1ra, TNF-alpha and VEGF. IQR: interquartile range. P-values were determined using one-sided Mann-Whitney tests and corrected using the Bonferroni-Sidak (Sidak) method.

ii) The impact of HIV

With regards to studying changes to the serum proteome in HIV/TB, this was more difficult to do from our data given that stage of HIV-1 infection is itself known to affect it,(19, 20) and TB, HIV+ patients would likely have had more advanced HIV disease than LTBI, HIV+ patients. Nonetheless, to gain an insight into the processes involved in HIV/TB, we firstly compared protein concentrations in TB, HIV+ with LTBI, HIV+, and secondly in TB, HIV- with TB, HIV+. Results from these analyses are shown in Heat Map 3 and Table 31, and Heat Map 4 and Table 32, below.

As shown in Heat Map 3, the protein whose levels were most greatly increased in TB, HIV+ compared with LTBI, HIV+ was IFN-alpha-2; and the protein most greatly decreased was transthyretin.



Heat Map 3. Differences in median protein levels between TB, HIV+ and LTBI, HIV+ groups. Differences were plotted as log-tranformed fold-changes (base 2) in comparison to the LTBI, HIV+ group. The legend on the right shows the colour scale in relation to log fold-change. Red: increased level; blue decreased.

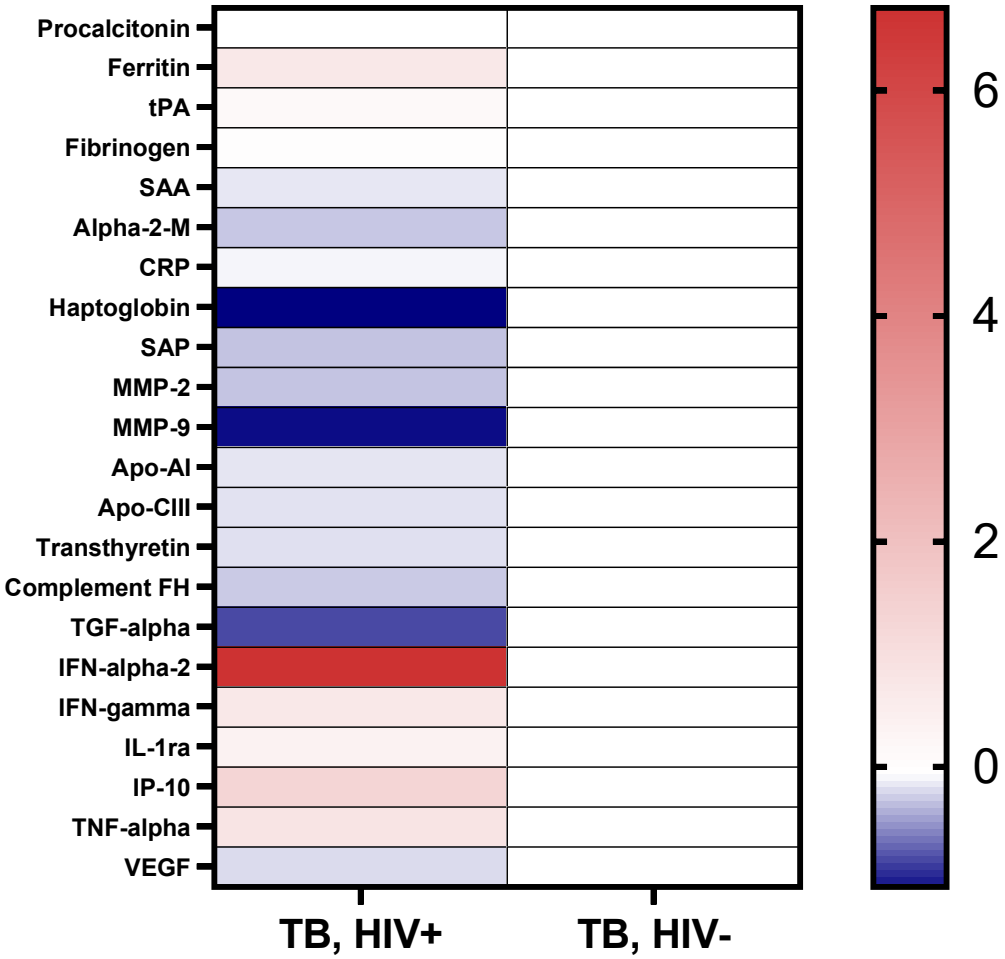
Concentrations of 17 of the 22 proteins were significantly different in TB, HIV+ patients compared with healthy LTBI, HIV+ individuals. Those that were not significant after p-value correction were alpha-2-M, haptoglobin, apo-CIII, complement factor H (FH), and IFN-alpha-2. Of the 17 that were different, the direction of change was as expected in the light of the raw data, namely that all proteins were higher except for MMP-2, apo-AI and transthyretin (Table 31, below).

	TB, HIV+		LTBI, HIV+			
	Median	**IQR**	**Median**	**IQR**	**P-value**	**Corrected P**
Procalcitonin	**1.90**	**(1.50 - 2.76)**	**1.20**	**(0.65 - 1.90)**	**0.00**	**0.00**
Ferritin	**104.20**	**(30.91 - 178.90)**	**14.07**	**(4.98 - 22.18)**	**0.00**	**0.00**
tPA	**2.82**	**(2.04 - 4.27)**	**1.32**	**(0.95 - 2.18)**	**0.00**	**0.00**
Fibrinogen	**2.73**	**(2.13 - 3.93)**	**1.67**	**(0.91 - 2.56)**	**0.00**	**0.00**
Serum amyloid A	**6.01**	**(3.92 - 8.84)**	**0.58**	**(0.27 - 2.93)**	**0.00**	**0.00**
Alpha-2-M	1.30	(1.09 - 1.62)	1.36	(1.13 - 1.66)	0.32	1.00
CRP	**61.97**	**(29.71 - 101.80)**	**1.42**	**(0.75 - 11.31)**	**0.00**	**0.00**
Haptoglobin	3.39	(0.31 - 9.12)	1.63	(0.24 - 4.75)	0.03	0.45
Serum amyloid P	**38.85**	**(30.58 - 49.29)**	**32.25**	**(24.64 - 39.76)**	**0.00**	**0.00**
MMP-2	**69.91**	**(58.53 - 90.51)**	**96.27**	**(72.29 - 121.50)**	**0.00**	**0.00**
MMP-9	**236.20**	**(141.70 - 390.60)**	**140.80**	**(90.71 - 258.20)**	**0.00**	**0.01**
Apo-AI	**223.70**	**(167.10 - 272.40)**	**294.50**	**(222.90 - 370.90)**	**0.00**	**0.00**
Apo-CIII	83.45	(61.64 - 129.60)	106.50	(82.12 - 134.30)	0.01	0.17
Transthyretin	**64.71**	**(46.26 - 97.87)**	**155.70**	**(92.03 - 233.20)**	**0.00**	**0.00**
Complement factor H	354.10	(285.30 - 438.30)	321.50	(283.80 - 371.00)	0.02	0.29
TGF-alpha	**8.04**	**(4.81 - 13.70)**	**4.17**	**(1.48 - 7.50)**	**0.00**	**0.00**
IFN-alpha-2	5.30	(0.00 - 48.70)	0.00	(0.00 - 13.92)	0.01	0.11
IFN-gamma	**22.48**	**(10.00 - 63.61)**	**6.41**	**(2.30 - 24.17)**	**0.00**	**0.00**
IL-1ra	**21.26**	**(3.00 - 49.25)**	**0.00**	**(0.00 - 10.74)**	**0.00**	**0.00**
IP-10	**1.95**	**(1.06 - 3.24)**	**0.64**	**(0.40 - 1.09)**	**0.00**	**0.00**
TNF-alpha	**33.09**	**(20.51 - 57.10)**	**21.21**	**(15.10 - 32.42)**	**0.00**	**0.00**
VEGF	**276.50**	**(137.60 - 533.50)**	**123.80**	**(62.91 - 314.20)**	**0.00**	**0.01**

Table 31 A comparison of serum protein concentrations in TB, HIV+ with those in LTBI, HIV+ patients.

Bold font represents proteins whose concentrations were significantly different in TB after p-value correction. Green font represents proteins whose concentrations were higher in TB, HIV+, red font for those that were lower. IQR: interquartile range. P-values were determined using one-sided Mann-Whitney tests and corrected using the Bonferroni-Sidak (Sidak) method.

With regards to levels in TB, HIV+ compared with TB, HIV-, as shown below in Heat Map 4, the protein that was most greatly increased in TB, HIV+ was IFN-alpha-2; and the protein most greatly decreased was haptoglobin.



Heat Map 4. Differences in median protein levels between TB, HIV+ and TB, HIV- groups. Differences were plotted as log-tranformed fold-changes (base 2) in comparison to the TB, HIV- group. The legend on the right shows the colour scale in relation to log fold-change. Red: increased level; blue decreased.

Concentrations of six proteins were significantly different in TB, HIV+ compared with TB, HIV-. Namely, concentrations of alpha-2-M, MMP-9, complement factor H and TGF-alpha were lower in TB, HIV+. Concentrations of IP-10 and TNF-alpha were higher in TB, HIV+ (Table 32, below).

	TB, HIV-		TB, HIV+			
	Median	**IQR**	**Median**	**IQR**	**P-value**	**Corrected P**
Procalcitonin	1.90	(1.18 - 2.5)	1.90	(1.50 - 2.76)	0.17	0.98
Ferritin	61.88	(29.92 - 112.50)	104.20	(30.91 - 178.90)	0.01	0.23
tPA	2.49	(1.70 - 3.08)	2.82	(2.04 - 4.27)	0.02	0.38
Fibrinogen	2.63	(1.77 - 3.11)	2.73	(2.13 - 3.93)	0.02	0.39
Serum amyloid A	6.45	(4.52 - 7.84)	6.01	(3.92 - 8.84)	0.36	1.00
Alpha-2-M	**1.53**	**(1.36 - 1.89)**	**1.30**	**(1.09 - 1.62)**	**0.00**	**0.00**
CRP	63.84	(26.27 - 95.66)	61.97	(29.71 - 101.80)	0.49	1.00
Haptoglobin	7.10	(0.39 - 13.80)	3.39	(0.31 - 9.12)	0.04	0.55
Serum amyloid P	46.31	(36.22 - 54.17)	38.85	(30.58 - 49.29)	0.00	0.09
MMP-2	82.99	(65.01 - 105.70)	69.91	(58.53 - 90.51)	0.01	0.14
MMP-9	**479.00**	**(290.00 - 614.90)**	**236.20**	**(141.70 - 390.60)**	**0.00**	**0.00**
Apo-AI	241.50	(171.80 - 308.70)	223.70	(167.10 - 272.40)	0.08	0.83
Apo-CIII	90.91	(68.78 - 115.70)	83.45	(61.64 - 129.60)	0.49	1.00
Transthyretin	70.97	(51.44 - 106.30)	64.71	(46.26 - 97.87)	0.11	0.93
Complement factor H	**413.30**	**(356.10 - 471.70)**	**354.10**	**(285.30 - 438.30)**	**0.00**	**0.02**
TGF-alpha	**13.66**	**(7.90 - 23.69)**	**8.04**	**(4.81 - 13.70)**	**0.00**	**0.00**
IFN-alpha-2	0.00	(0.00 - 20.95)	5.30	(0.00 - 48.70)	0.03	0.47
IFN-gamma	13.60	(5.44 - 28.73)	22.48	(10.00 - 63.61)	0.00	0.06
IL-1ra	15.82	(1.10 - 54.29)	21.26	(3.00 - 49.25)	0.31	1.00
IP-10	**0.76**	**(0.43 - 1.11)**	**1.95**	**(1.06 - 3.24)**	**0.00**	**0.00**
TNF-alpha	**18.13**	**(12.47 - 28.58)**	**33.09**	**(20.51 - 57.10)**	**0.00**	**0.00**
VEGF	307.20	(130.30 - 810.70)	276.50	(137.60 - 533.50)	0.21	0.99

Table 32 A comparison of serum protein concentrations in TB, HIV- with those in TB, HIV+ patients.

Bold font represents proteins whose concentrations were significantly different in TB after p-value correction. Green font represents proteins whose concentrations were higher in TB, HIV+, red font for those that were lower. Units of protein levels are as follows: mg/ml: Alpha-2-M and Haptoglobin; ug/ml: fibrinogen, serum amyloid A, CRP, serum amyloid P, MMP-2, MMP-9, apo-AI, apo-CIII, transthyretin and complement FH; ng/ml: procalcitonin, ferritin, tPA and IP-10; and pg/ml: TGF-alpha, IFN-alpha-2, IFN-gamma, IL-1ra, TNF-alpha and VEGF. IQR: interquartile range. P-values were determined using one-sided Mann-Whitney tests and corrected using the Bonferroni-Sidak (Sidak) method.

3. Associations between the 22 AE-TBC biomarkers and clinical variables in TB

i) Smear grade

High bacterial load in sputum has long been recognised as a feature of severity in TB, correlating with poorer microbiological and clinical outcomes.(3-6, 21) POCT biomarkers representative of bacterial load would therefore be extremely useful for patient care in resource-limited settings, for prognostication and treatment monitoring. Culture-based quantitation of load (i.e. colony counting on solid media or time to detection (TTD) in liquid media) was not available for this study, however both sputum samples that were collected from each patient were subjected to smear grading, as well as culture, following the system of the International Union Against Tuberculosis and Lung Disease.(22) Whilst subject to a degree of operator dependency, smear grade has been shown to correlate either moderately strongly or strongly with other quantitative techniques that measure bacterial load as a continuous variable, such as TTD or PCR cycle threshold.(23-25)

In the studies that correlated smear grade with continuous measures of bacterial load, smear grade was analysed as an ordinal variable, i.e. negative, scanty, 1+, 2+ or 3+.(23-25) The study by Hanrahan *et al* ('14) recruited 366 sputum culture positive adult patients, of which 217 were HIV infected and 149 were HIV uninfected. The correlation was stronger in HIV uninfected patients (rho = 0.62 (95% confidence intervals 0.53 – 0.71)), but still moderately strong in HIV infected patients: 0.50 (0.40 – 0.60) in patients with CD4+ counts >/= 200 and 0.45 (0.33 – 0.57) in patients with CD4+ counts < 200.(24) When assessing association between smear grade and outcome, however, investigators have most commonly considered smear grade as a binary variable with >/= 2+ being "high grade", and < 2+ as "low grade".(3, 6, 26, 27) Prevalence of HIV in these latter studies, however, was either low or non-existent, and in the case of HIV-associated TB, it has been more common to consider smear positive (i.e. any grade above negative) as representing higher bacterial loads, and smear negative as representing lower loads.(28-30) A direct relationship between sputum bacterial load and outcome in HIV+TB+ is also more difficult to define, due to the confounding effect of CD4+ on both sputum load and outcome,(31) and the fact that the negative impact of treatment delay in smear negativity may be more pronounced in HIV+TB+.(32) Moreover, the distribution of smear grades was significantly different between TB, HIV and TB, HIV+ in the EU-TB cohort (Figure 52, below).

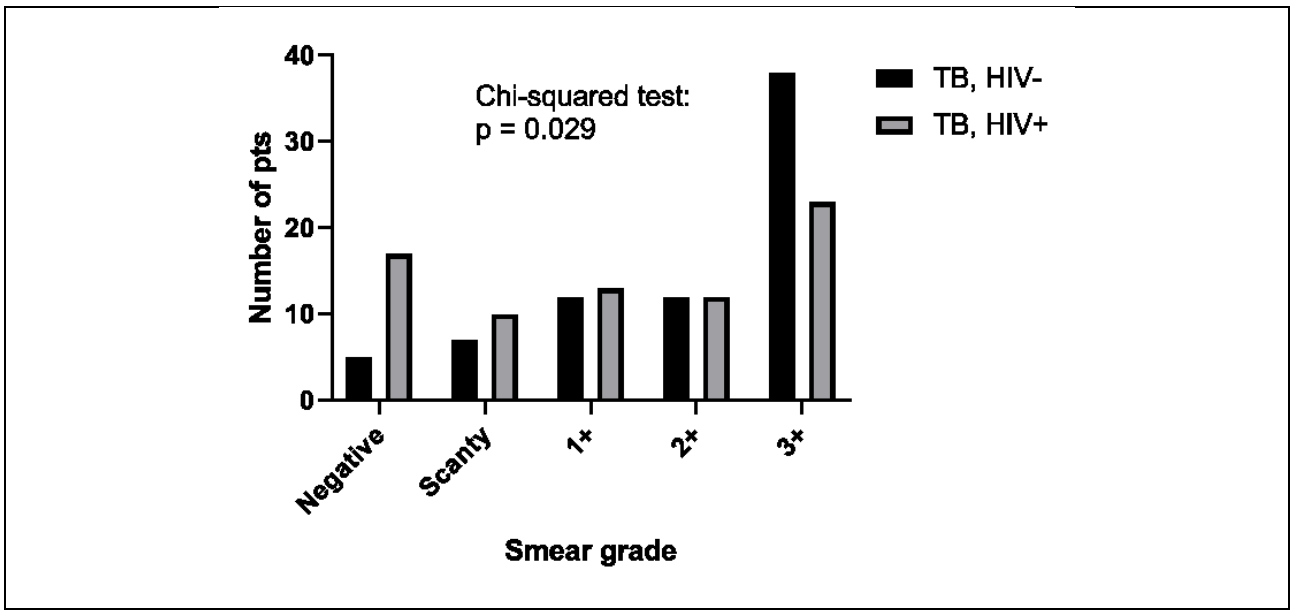


Figure 52 Distribution of smear grades in each of TB, HIV- and TB, HIV+.

The higher of two smear grades for each patient is plotted. The smear grading system of the International Union Against Tuberculosis and Lung Disease was used.(33)

We therefore proceeded to look for associations between smear grade and serum protein level in the following manner:

a) All TB patients: smear grade as an ordinal variable (negative, scanty, 1+, 2+, 3+) to assess correlation with bacterial load regardless of HIV status;
b) TB+HIV-: smear grade as a binary variable (< 2+ vs. >/= 2+) to assess association with bacterial load using a clinically meaningful categorisation;
c) TB+HIV+: smear grade as a binary variable (-ve vs. +ve) to assess association with bacterial load using a clinically meaningful categorisation.

In each case, the higher of the two grades for each patient was used. Since our hypothesis was that higher smear grades would be associated with a greater degree of disturbance to the serum proteome, one-sided statistical tests were performed. For each family of analyses, p-values were corrected using the Bonferroni-Sidak (or "Sidak") method.

Coefficients of correlation of serum protein levels with smear grade as an ordinal variable in all TB patients (i.e. analysis a) above) are shown below in Table 33. There were weak but statistically significant correlations with smear grade for CRP (rho 0.28), MMP-9 (rho 0.31) and TGF-alpha (rho 0.28).

Protein	Rho	95% CIs	Uncorrected P	Corrected P
Procalcitonin	0.09	-0.09 to 0.26	0.15	0.97
Ferritin	0.15	-0.03 to 0.32	0.05	0.64
tPA	0.19	0.01 to 0.35	0.02	0.29
Fibrinogen	0.08	-0.10 to 0.26	0.18	0.99
Serum amyloid A	0.10	-0.08 to 0.27	0.13	0.96
Alpha-2-M	0.04	-0.14 to 0.21	0.33	1.00
CRP	**0.28**	**0.11 to 0.43**	**0.00**	**0.01**
Haptoglobin	0.15	-0.02 to 0.32	0.04	0.59
Serum amyloid P	0.10	-0.08 to 0.27	0.14	0.96
MMP-2	-0.20	-0.36 to -0.04	0.01	0.15
MMP-9	**0.31**	**0.15 to 0.46**	**0.00**	**0.00**
Apo-AI	-0.15	-0.31 to 0.02	0.04	0.56
Apo-CIII	-0.18	-0.34 to -0.01	0.02	0.29
Transthyretin	-0.19	-0.35 to -0.03	0.01	0.19
Complement FH	0.09	-0.08 to 0.25	0.15	0.97
TGF-alpha	**0.28**	**0.12 to 0.43**	**0.00**	**0.01**
IFN-alpha-2	-0.06	-0.23 to 0.11	0.24	1.00
IFN-gamma	0.00	-0.17 to 0.17	0.49	1.00
IL-1ra	0.10	-0.07 to 0.27	0.11	0.92
IP-10	-0.08	-0.25 to 0.09	0.17	0.99
TNF-alpha	-0.02	-0.19 to 0.15	0.39	1.00
VEGF	0.11	-0.07 to 0.27	0.11	0.92

Table 33 Correlation of serum protein levels with smear grade in patients with active TB, regardless of HIV status (n = 150).

The higher of two smear grades was used for each patient and smear grade was treated as an ordinal variable: negative; scanty; 1+; 2+; 3+. Spearman's rank correlation for non-parametric data was performed. p-values are 1-sided and shown before and after Sidak's correction for multiple comparisons. Rows in bold represent proteins that were significantly correlated after correction.

Box plots to display levels of these three proteins by each smear grade category are shown below in Figure 53.

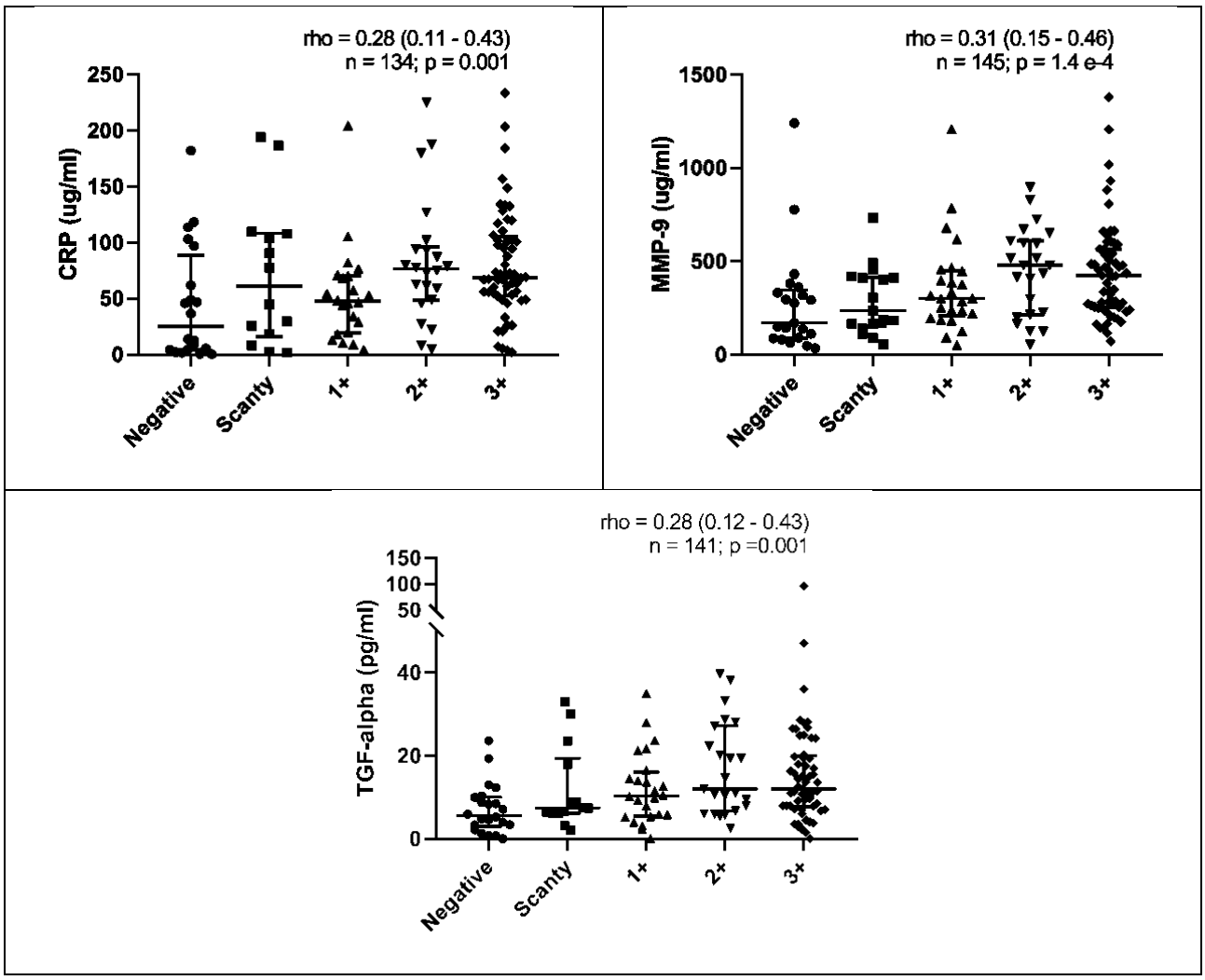


Figure 53 Box plots of protein level by smear grade category for the three proteins that correlated significantly with smear grade as an ordinal variable, regardless of HIV status.

Rho: correlation coefficient from Spearman's ranked analyses.

Regarding TB mono-infected patients (i.e. analysis b)), levels in each of smear <2+ and >/= 2+ patients are shown below in Table 34. Levels are shown by their medians and interquartile ranges. Levels of MMP-2 were significantly lower in TB mono-infected patients with smear grades >/= 2+ vs. <2+ (Table 34, below).

	Smear <2+ (median (IQR))	Smear >/= 2+ (median (IQR))	Uncorrected P	Corrected P
Procalcitonin	1.79 (1.35 - 2.14)	1.97 (0.95 - 2.76)	0.29	1.00
Ferritin	38.1 (22.2 - 82.6)	70.9 (46.0 - 133)	0.00	0.09
tPA	1.81 (1.49 - 2.80)	2.72 (1.97 - 3.58)	0.00	0.09
Fibrinogen	2.39 (1.90 - 3.05)	2.66 (1.64 - 3.20)	0.46	1.00
Serum amyloid A	5.70 (3.64 - 7.41)	6.48 (4.86 - 9.02)	0.17	0.98
Alpha-2-M	1.53 (1.35 - 1.80)	1.53 (1.36 - 1.92)	0.48	1.00
CRP	34.0 (8.81 - 76.9)	72.9 (48.6 - 104)	0.00	0.09
Haptoglobin	3.73 (0.265 - 12.9)	10.3 (0.598 - 16.2)	0.05	0.70
Serum amyloid P	44.7 (34.9 - 54.2)	47.4 (36.8 - 53.9)	0.36	1.00
MMP-2	**102 (83.6 - 121)**	**74.6 (63.6 - 98.9)**	**0.00**	**0.00**
MMP-9	418 (295 - 633)	512 (283 - 615)	0.10	0.90
Apo-AI	265 (181 - 402)	235 (172 - 283)	0.08	0.84
Apo-CIII	101 (76.6 - 138)	83.3 (65.1 - 101)	0.02	0.29
Transthyretin	78.5 (59.3 - 142)	64.9 (48.9 - 94.2)	0.04	0.57
Complement FH	407 (323 - 429)	413 (372 - 481)	0.10	0.90
TGF-alpha	10.1 (6.19 - 19.9)	16.4 (8.98 - 25.8)	0.03	0.51
IFN-alpha-2	0 (0 - 9.43)	0 (0 - 23.4)	0.09	0.89
IFN-gamma	7.41 (3.39 – 19.0)	16.0 (8.00 - 38.4)	0.00	0.11
IL-1ra	12.6 (3.59 - 40.9)	21.2 (1.09 - 54.5)	0.37	1.00
IP-10	0.580 (0.348 - 0.845)	0.890 (0.510 - 1.21)	0.01	0.25
TNF-alpha	16.2 (10.2 - 29.8)	22.3 (13.0 – 29.0)	0.18	0.99
VEGF	279 (150 - 672)	438 (120 - 826)	0.24	1.00

Table 34 Protein levels in TB mono-infected patients (HIV -ve) with smear grades <2+ and >/= 2+.

Medians and interquartile ranges (IQRs) are shown. One-sided p-values before and after Sidak's correction for multiple comparisons are shown. Proteins whose levels were significantly different between smear grade categories are highlighted in bold. Units of protein levels: mg/ml: Alpha-2-M and Haptoglobin; ug/ml: fibrinogen, serum amyloid A, CRP, serum amyloid P, MMP-2, MMP-9, apo-AI, apo-CIII, transthyretin and complement FH; ng/ml: procalcitonin, ferritin, tPA and IP-10; and pg/ml: TGF-alpha, IFN-alpha-2, IFN-gamma, IL-1ra, TNF-alpha and VEGF.

The box plot displaying the data this is shown below in Figure 54.

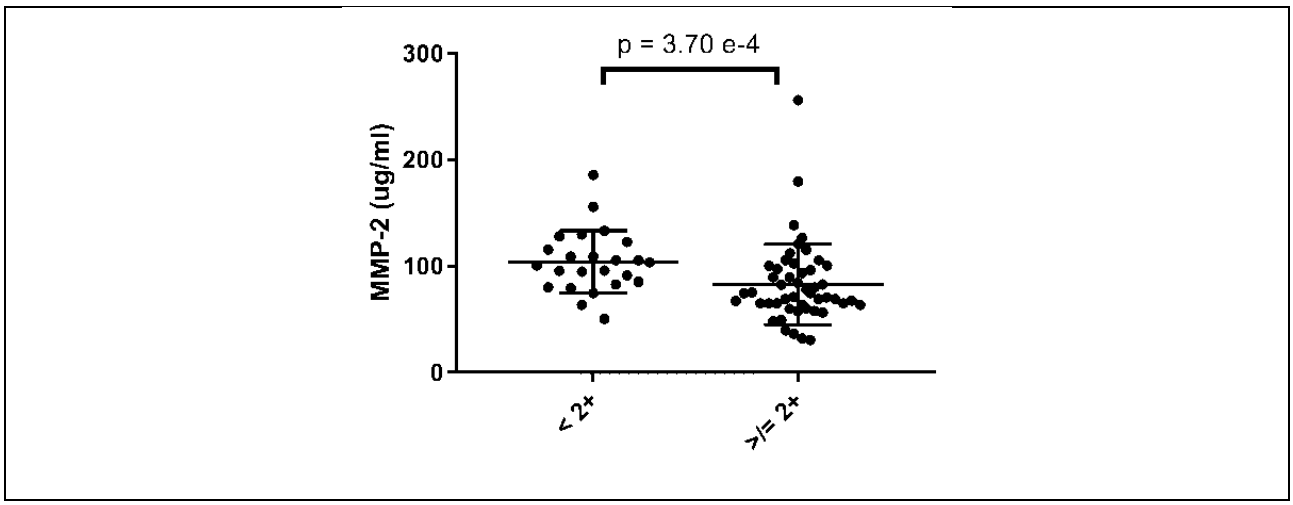


Figure 54 Serum MMP-2 levels in TB mono-infected patients (HIV -ve) with smear grades < 2+ vs. >/= 2+.

p-value is 1-sided, uncorrected and derived from a Mann-Whitney test.

Regarding HIV co-infected patients (i.e. analysis c)), levels in each of smear negative and positive patients are shown below in Table 35. Levels of tPA, fibrinogen, MMP-9, apo-AI and TGF-alpha were significantly different between patients with smear -ve and +ve disease in the context of HIV co-infection.

	Smear -ve (median (IQR))	Smear +ve (median (IQR))	Uncorrected P	Corrected P
Procalcitonin	1.63 (1.20 - 1.79)	2.14 (1.60 - 2.88)	0.01	0.19
Ferritin	70.6 (29.4 - 131)	109 (33.8 - 198)	0.18	0.99
tPA	2.02 (1.51 - 2.79)	3.06 (2.37 - 4.53)	0.00	0.05
Fibrinogen	**2.26 (1.22 - 2.73)**	**3.24 (2.35 - 4.42)**	**0.00**	**0.01**
Serum amyloid A	4.14 (0.470 - 8.55)	6.37 (4.32 - 8.91)	0.10	0.90
Alpha-2-M	1.47 (1.19 - 1.65)	1.25 (1.06 - 1.61)	0.18	0.99
CRP	46.0 (4.64 - 97.2)	67.1 (46.4 - 102)	0.02	0.34
Haptoglobin	3.70 (1.21 - 22)	2.76 (0.303 - 8.94)	0.21	1.00
Serum amyloid P	36.5 (31.9 - 47.3)	40.2 (29.8 - 49.9)	0.34	1.00
MMP-2	82.2 (59.4 - 98.9)	69.2 (58.5 - 87.2)	0.10	0.89
MMP-9	**141 (82.5 - 291)**	**257 (166 - 414)**	**0.00**	**0.04**
Apo-AI	**327 (211 - 421)**	**199 (155 - 246)**	**0.00**	**0.01**
Apo-CIII	124 (86.4 - 166)	76.5 (57.8 - 123)	0.02	0.33
Transthyretin	125 (42.2 - 177)	56.6 (46.3 - 80.7)	0.00	0.14
Complement FH	342 (292 - 490)	360 (284 - 432)	0.45	1.00
TGF-alpha	**4.81 (1.71 - 8.62)**	**9.53 (5.79 - 15)**	**0.00**	**0.02**
IFN-alpha-2	20.3 (0 - 54.6)	0.690 (0 - 46.6)	0.24	1.00
IFN-gamma	14.1 (5.90 - 73.2)	23.5 (11.9 - 64.3)	0.24	1.00
IL-1ra	3.80 (0 - 20.9)	26.7 (3.70 - 66.2)	0.00	0.12
IP-10	1.60 (0.98 - 2.56)	1.98 (1.07 - 3.27)	0.14	0.96
TNF-alpha	21.4 (20.1 - 39.6)	33.9 (21.9 - 62.5)	0.04	0.59
VEGF	244 (145 - 432)	285 (116 - 578)	0.34	1.00

Table 35 Serum protein levels in smear -ve vs. smear +ve patient with active TB and HIV. Median and interquartile ranges

(IQRs) are shown in each case, followed by one-sided p-values from Mann-Whitney tests. One-sided p-values are shown before and after Sidak's correction for multiple comparisons. Units: mg/ml: Alpha-2-M and Haptoglobin; ug/ml: fibrinogen, serum amyloid A, CRP, serum amyloid P, MMP-2, MMP-9, apo-AI, apo-CIII, transthyretin and complement FH; ng/ml: procalcitonin, ferritin, tPA and IP-10; and pg/ml: TGF-alpha, IFN-alpha-2, IFN-gamma, IL-1ra, TNF-alpha and VEGF.

Box plots to display these comparisons are shown below in Figure 55. Levels of fibrinogen, MMP-9 and TGF-alpha were higher in smear +ve patients that were co-infected with HIV, but levels of Apo-AI were lower.

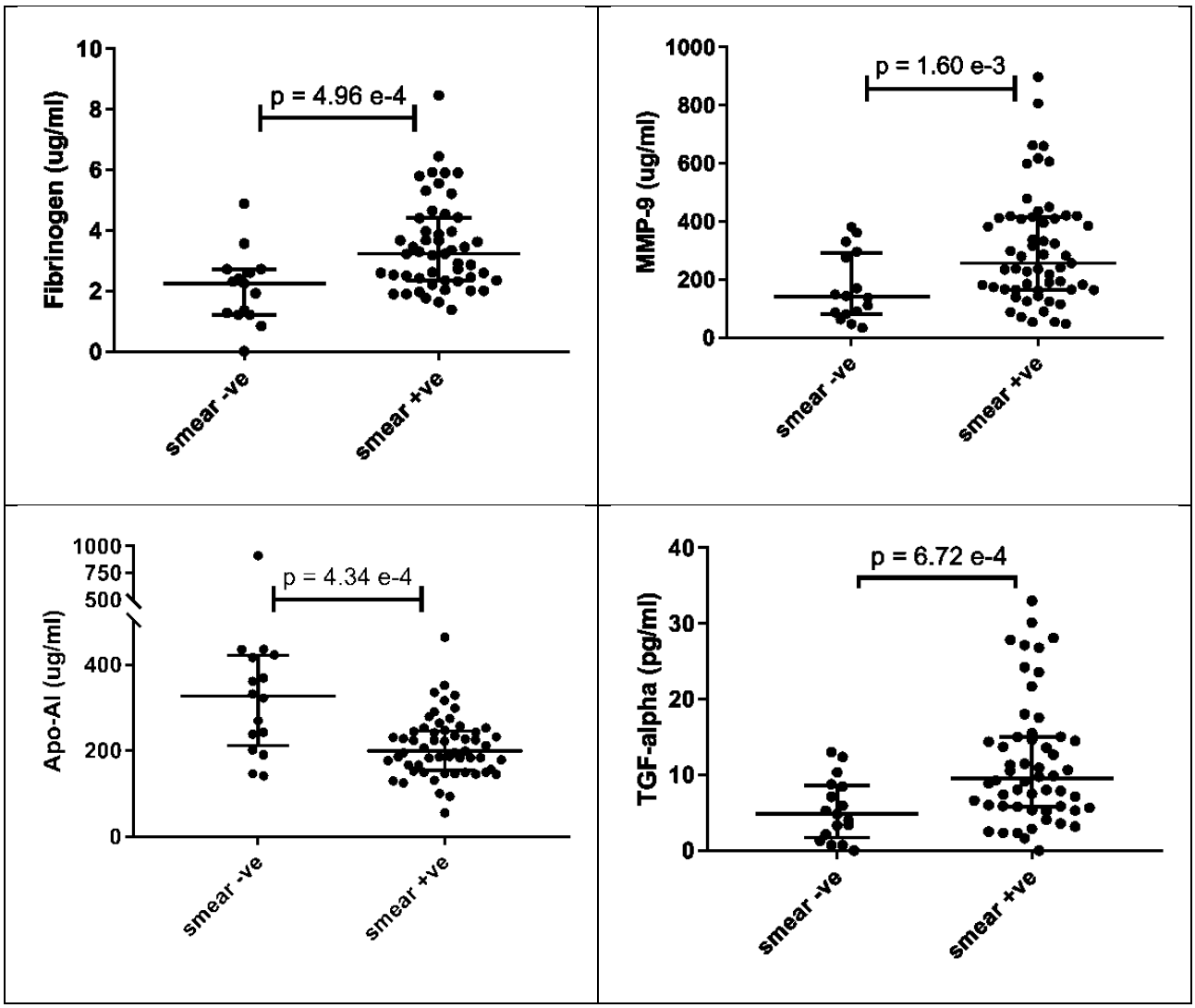


Figure 55 Levels of fibrinogen, MMP-9, apo-AI and TGF-alpha in HIV co-infected TB patients that were smear negative (-ve) vs. positive (+ve).

The differences were statistically significant after allowing for multiple comparisons. The p-values were derived from Mann-Whitney tests and are shown uncorrected and 1-sided.

ii) Body Mass Index

A link between body mass index (BMI) and severity of TB has been long known, and demonstrated recently by the following studies:

a) A study of predictors of treatment failure or mortality in 75 patients with multidrug-resistant (MDR-)TB in Peru found that a “low BMI” at outset was independently associated with risk of death by 4 months. Low BMI was defined as < 18.5 for women or <20 for men.(10)

b) A study of 204 patients with MDR-TB in Latvia found that a BMI < 18.5 at the outset was independently associated with treatment failure or death (adjusted HR 2.3 (1.1 – 4.9)).(11)

c) A study of 316 adult TB patients in Malawi, with and without HIV co-infection, found correlations between radiological grade of TB by CXR and nutritional state, independent of HIV status, gender or age. Nutritional state was estimated using BMI, which was analysed as a continuous variable and < 19 vs. >/= 19.(12)

d) A study of 1,502 TB patients from Bissau, Guinea Buissau and Gondar, Ethiopia found that a field-friendly six-marker model comprising BMI, middle upper arm circumference, cough, dyspnoea, chest pain and anaemia was predictive of both treatment failure and death.(13) BMI < 18 scored one point and < 16 scored two. A comparison of BMI between patients that failed or died and those that did not was not shown but the combined ROC AUCs at each of Bissau and Gondar for treatment failure were significant at 0.71 and 0.66 respectively, and for death were 0.73 and 0.68.(13)

e) A retrospective cohort study of 212 MDR-TB patients in Jakarta, Indonesia showed that patients who were severely underweight at the outset (BMI < 16, n = 93) took significantly longer to culture convert than those with normal BMI (> 18.5, n = 60, adjusted hazard ratio 0.55 (95% confidence intervals 0.37 – 0.84)). In addition, 16% of the patients with initial BMI < 16 died compared to 6% of those with normal BMI (p = 0.015). Notably, however, patients with BMI 16 - 18.5 (i.e. underweight but not severely) did not have worse outcomes compared to those with normal BMI in this study. A comparison of outcomes between those with BMI < vs. >/= 18.5 was not shown.(2)

The other important relationship between BMI and TB is that BMI is inversely associated with risk of TB.(14) It could be envisaged, therefore, that the the associations between serum protein responses and BMI could be either positive or negative: positive if lower BMI results from a more severe response, or negative if results in a less competent response. We therefore decided to seek associations with two-sided tests.

In our study, BMI did not differ significantly between TB, HIV- and TB, HIV+ patients. Levels of 11 of the 22 serum proteins measured, however, were significantly different between TB, HIV- and TB, HIV+, being higher in TB, HIV+ in the case of TNF-alpha, ferritin, fibrinogen, IFN-gamma and IP-10, but lower in the case of alpha-2-M, serum amyloid P, complement FH, MMP-2, MMP-9 and TGF-alpha (Chapter III).

In the studies in the literature listed above, BMI was analysed as a continuous variable by van Lettow *et al* ('04),(12) but as a binary variable by all the other investigators. The WHO definition of underweight is BMI < 18.5.(34) BMI < 16 is referred to as "severe thinness" but is a less well validated indicator of nutritional state and is also used less frequently in studies of TB.(34) We therefore decided to look for associations between BMI and serum protein levels in the following manner:

i. All TB patients: BMI as a continuous variable to assess correlation with BMI regardless of HIV status;

ii. TB, HIV- or TB, HIV+: BMI as a binary variable (< 18.5 vs. >/= 18.5) to assess association with BMI using a well validated marker of prognosis whilst controlling for HIV status.

When all TB patients were considered together (regardless of HIV status), levels of five proteins correlated weakly but significantly with BMI: procalcitonin, tPA, fibrinogen, TGF-alpha and VEGF (Table 36, below). In all cases the correlations were negative, which is in keeping with the fact that they were all also found to be positive biomarkers of TB (i.e. higher in TB than LTBI: Chapter III, Section 2).

	Rho	95% CIs	P	Corrected P
Procalcitonin	**-0.35**	**-0.50 to -0.18**	**0.00**	**0.00**
Ferritin	-0.26	-0.42 to -0.08	0.00	0.08
tPA	**-0.29**	**-0.45 to -0.11**	**0.00**	**0.03**
Fibrinogen	**-0.38**	**-0.53 to -0.22**	**0.00**	**0.00**
Serum amyloid A	-0.22	-0.38 to -0.04	0.01	0.30
Alpha-2-M	-0.18	-0.35 to -0.00	0.02	0.60
CRP	-0.12	-0.29 to 0.07	0.10	0.99
Haptoglobin	-0.09	-0.27 to 0.09	0.15	1.00
Serum amyloid P	0.15	-0.03 to 0.32	0.05	0.90
MMP-2	-0.15	-0.31 to 0.03	0.04	0.86
MMP-9	-0.22	-0.38 to -0.05	0.01	0.21
Apo-AI	0.17	-0.00 to 0.33	0.02	0.67
Apo-CIII	0.21	0.04 to 0.37	0.01	0.26
Transthyretin	0.24	0.07 to 0.39	0.00	0.11
Complement FH	0.11	-0.07 to 0.28	0.10	0.99
TGF-alpha	**-0.34**	**-0.49 to -0.18**	**0.00**	**0.00**
IFN-alpha-2	0.09	-0.09 to 0.26	0.16	1.00
IFN-gamma	0.07	-0.11 to 0.24	0.21	1.00
IL-1ra	-0.16	-0.33 to 0.01	0.03	0.75
IP-10	0.05	-0.13 to 0.22	0.29	1.00
TNF-alpha	-0.05	-0.22 to 0.13	0.28	1.00
VEGF	**-0.27**	**-0.43 to -0.10**	**0.00**	**0.03**

Table 36 Correlation of serum protein levels with body mass index (BMI) in the EU-TB active TB patients (n = 150).

All TB patients are included, regardless of HIV status. Spearman's rank correlation coefficients (rho) are shown for each protein. p-values are 2-sided and shown before and after Sidak's correction for multiple comparisons. Rows in bold represent proteins whose levels were significantly associated with BMI after correction.

The raw data for these correlations is displayed in Figure 56 below.

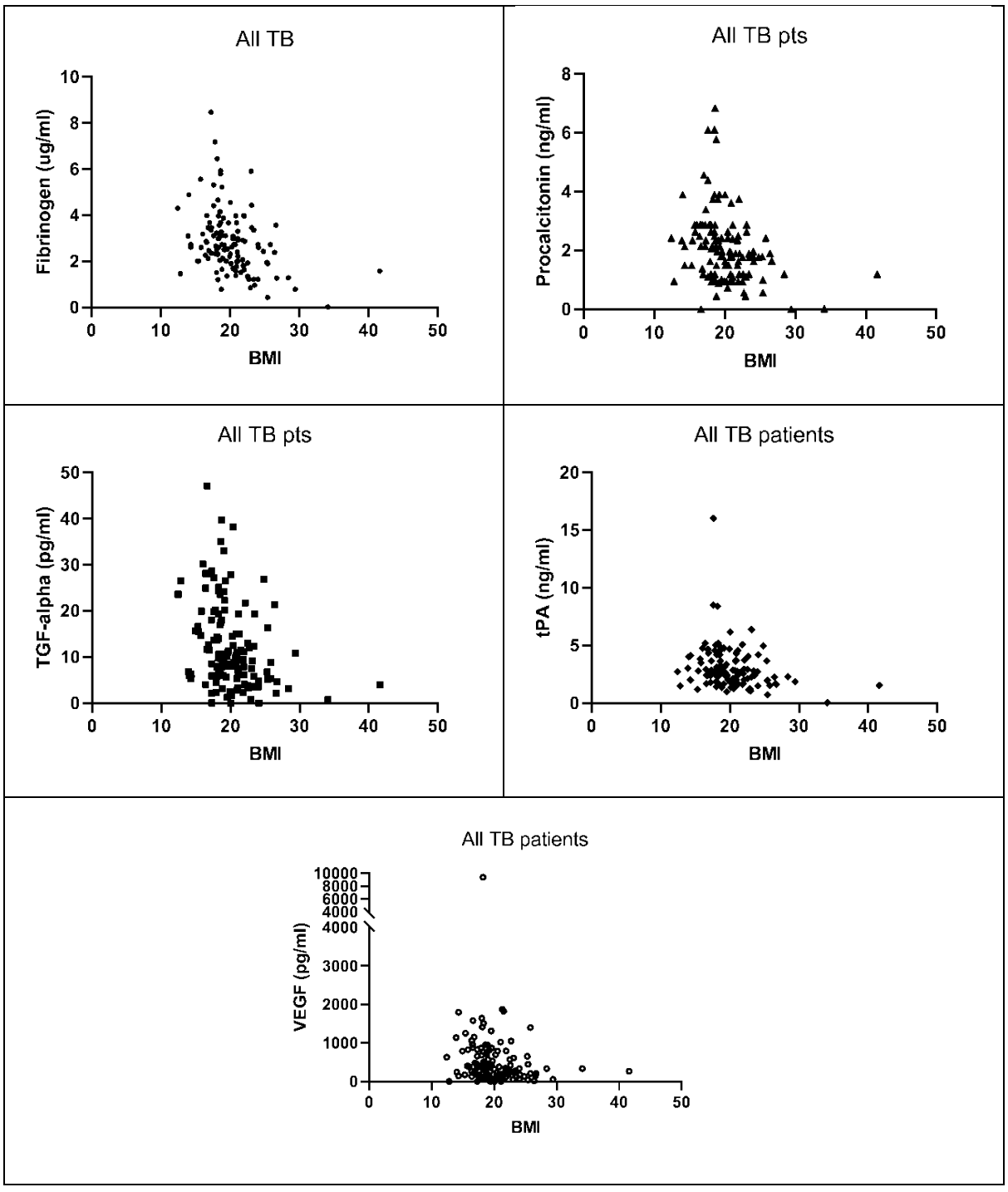


Figure 56 XY scatter plots of serum protein level vs. BMI in the TB patients (HIV- and HIV+ together).

The number of patients represented in each graph was 123 for Fibrinogen, PCT and tPA, 144 for TGF-alpha and 143 for VEGF.

As explained above, the TB patients were then stratified by HIV status and associations sought between protein level and BMI, with BMI divided into <18.5 and >/= 18.5. The distribution of patients across these categories is shown below in Table 37. In both TB, HIV- and TB, HIV+, there were more patients in the BMI >/= 18.5 group.

	BMI < 18.5	BMI >/= 18.5
TB, HIV-	27	40
TB, HIV+	23	51
Total	**50**	**91**

Table 37 Numbers of TB patients with BMI < 18.5 and >/= 18.5 when stratified by HIV status, and in total.

Two-sided tests were again performed. In both TB, HIV- and TB, HIV+ patients, there were no proteins with levels that were significantly different after p-value correction across these two BMI categories (Tables 38 and 39).

	Med (IQR)			
	< 18.5	**>/= 18.5**	**P**	**Corrected P**
Procalcitonin	2.14 (1.155 - 2.76)	1.79 (0.95 - 2.35)	0.15	1.00
Ferritin	76.91 (34.38 - 137)	54.59 (26.52 - 93.93)	0.11	0.99
tPA	2.72 (1.76 - 3.575)	2.49 (1.643 - 2.973)	0.19	1.00
Fibrinogen	2.86 (2.215 - 3.51)	2.18 (1.565 - 2.99)	0.01	0.29
Serum amyloid A	6.03 (4.68 - 8.91)	6.415 (3.553 - 7.798)	0.26	1.00
Alpha-2-M	1.585 (1.363 - 1.94)	1.525 (1.318 - 1.9)	0.28	1.00
CRP	53.59 (26.19 - 94.96)	61.92 (18.29 - 112.2)	0.50	1.00
Haptoglobin	6.86 (2.45 - 12.8)	4.945 (0.3475 - 17.08)	0.50	1.00
Serum amyloid P	36.73 (31.58 - 48.81)	49.7 (41.23 - 55.59)	0.00	0.10
MMP-2	89.7 (70.13 - 110.5)	81.5 (60.03 - 100.7)	0.07	0.97
MMP-9	518.2 (367.8 - 731.9)	450.6 (280.4 - 587)	0.03	0.74
Apo-AI	235.2 (165 - 311)	248.3 (210.7 - 317.6)	0.17	1.00
Apo-CIII	75.04 (61.11 - 93.01)	99.26 (76.41 - 120.2)	0.00	0.15
Transthyretin	59.8 (41.91 - 81.75)	80.41 (54.94 - 125.5)	0.02	0.59
Complement FH	384.4 (332.1 - 432.4)	418.1 (394.9 - 472.1)	0.03	0.73
TGF-alpha	19.33 (7.76 - 25.01)	10.4 (7.09 - 19.36)	0.03	0.80
IFN-alpha-2	0 (0 - 5.52)	0 (0 - 33.81)	0.08	0.97
IFN-gamma	10.98 (2.74 - 20.94)	15.95 (6.85 - 44.19)	0.04	0.81
IL-1ra	24.5 (3.405 - 58.74)	11.58 (0 - 34.68)	0.15	1.00
IP-10	0.67 (0.405 - 0.955)	0.79 (0.45 - 1.24)	0.16	1.00
TNF-alpha	19.18 (14.57 - 28.66)	17.18 (11.21 - 29.39)	0.31	1.00
VEGF	661.5 (180.1 - 919.9)	242 (114.3 - 513)	0.01	0.21

Table 38 Comparison of protein levels between TB, HIV- patients with BMI < 18.5 vs. >/= 18.5.

Medians and interquartile ranges of each group are shown. P-values are 2-sided, result from Mann-Whitney tests and are shown before and after Sidak's correction for multiple comparisons. Units: mg/ml: Alpha-2-M and Haptoglobin; ug/ml: fibrinogen, serum amyloid A, CRP, serum amyloid P, MMP-2, MMP-9, apo-AI, apo-CIII, transthyretin and complement FH; ng/ml: procalcitonin, ferritin, tPA and IP-10; and pg/ml: TGF-alpha, IFN-alpha-2, IFN-gamma, IL-1ra, TNF-alpha and VEGF.

	Med (IQR)			
	< 18.5	**>/= 18.5**	**P**	**Corrected P**
Procalcitonin	2.565 (2.088 - 2.88)	1.78 (1.2 - 2.43)	0.00	0.09
Ferritin	148.6 (43.1 - 292.6)	94.62 (27.53 - 141.4)	0.04	0.86
tPA	3.47 (2.558 - 4.745)	2.75 (1.85 - 4.143)	0.02	0.53
Fibrinogen	3.21 (2.378 - 4.833)	2.68 (1.908 - 3.678)	0.04	0.84
Serum amyloid A	7.2 (6.508 - 9.643)	5.145 (3.483 - 7.805)	0.01	0.43
Alpha-2-M	1.43 (1.25 - 1.788)	1.215 (1.03 - 1.545)	0.02	0.59
CRP	64.23 (49.21 - 104.7)	61.64 (21.47 - 98.61)	0.14	1.00
Haptoglobin	4.44 (0.3275 - 8.328)	3.365 (0.2975 - 11.25)	0.41	1.00
Serum amyloid P	45.36 (30.56 - 49.83)	37.73 (30.55 - 48.62)	0.29	1.00
MMP-2	77 (66.61 - 95.29)	68.31 (57.85 - 90.51)	0.24	1.00
MMP-9	303.9 (168.8 - 409.2)	211.1 (122.7 - 382.6)	0.12	1.00
Apo-AI	225.5 (162.9 - 253.1)	217.2 (164.5 - 293.1)	0.42	1.00
Apo-CIII	76.71 (61.88 - 126.8)	88.13 (56.43 - 132.3)	0.42	1.00
Transthyretin	55.32 (39.45 - 97.82)	70.07 (48.48 - 108.7)	0.25	1.00
Complement FH	375.7 (277.8 - 436.1)	353.2 (285.7 - 448.4)	0.45	1.00
TGF-alpha	11.49 (5.92 - 15.07)	7.52 (3.365 - 11.85)	0.02	0.56
IFN-alpha-2	5.3 (0 - 34.64)	1.38 (0 - 59.87)	0.36	1.00
IFN-gamma	25.58 (12.33 - 93.17)	19.41 (8.39 - 50.91)	0.07	0.96
IL-1ra	44 (0 - 145.3)	11.58 (3.055 - 39.5)	0.15	1.00
IP-10	2.17 (1.13 - 3.78)	1.8 (1.06 - 3.115)	0.12	1.00
TNF-alpha	35.73 (21.42 - 59.87)	29.14 (20.38 - 54.06)	0.28	1.00
VEGF	386.4 (220.9 - 917.9)	239.8 (114.3 - 488)	0.03	0.69

Table 39 Comparison of protein levels between TB, HIV+ patients with BMI < 18.5 vs. >/= 18.5.

Medians and interquartile ranges of each group are shown. P-values are 2-sided, result from Mann-Whitney tests and are shown before and after Sidak's correction for multiple comparisons. Units: mg/ml: Alpha-2-M and Haptoglobin; ug/ml: fibrinogen, serum amyloid A, CRP, serum amyloid P, MMP-2, MMP-9, apo-AI, apo-CIII, transthyretin and complement FH; ng/ml: procalcitonin, ferritin, tPA and IP-10; and pg/ml: TGF-alpha, IFN-alpha-2, IFN-gamma, IL-1ra, TNF-alpha and VEGF.

iii) CD4+ count

In HIV-associated TB (HIV/TB), peripheral blood CD4+ cell count (CD4+ count) has been shown in multiple studies to be associated with outcome of TB treatment. These studies include:

a) a prospective cohort study of 209 adults with HIV/TB that were admitted due to deterioration during TB treatment in South Africa, in which the risk of admission was inversely associated with CD4+ count (categorised as <200, 200 - 350 and >/= 350 cells/uL), and 15 of 17 deaths occurred in patients with CD4+ < 200;(15)

b) a prospective cohort study of 723 adult outpatients with smear negative HIV/TB in Zimbabwe, in which the main risk factor for death was immunosuppression, with a CD4+ count < 50 conferring a 13-fold increase in risk (95% CIs: 5.06 – 33.08);(16)

c) another prospective cohort study of 227 in- and outpatients with HIV/TB in Malaysia, in which a CD4+ < 200 was independently associated with death (adjusted HR 3.9, 95% CI 1.20 - 12.63);(17)

d) a retrospective cohort study of 337 hospitalised adults with HIV/TB in Cameroon, in which the most strongly assocated independent risk factor for death was CD4+ < 50 (adjusted OR =16.43 (95% CIs: 1.05 - 258.04);(7)

e) a large retrospective cohort study that included 10,379 patients with HIV/TB in South Africa, in which CD4+ count was again inversely associated with risk of death. At every CD4+ count category > 100 (i.e. 100 - 200, 200 - 350 and >/= 350), the adjusted hazard ratio dropped further in comparison to the CD4+ < 50 group.(18)

To our knowledge, associations between serum protein levels and CD4+ count in HIV/TB have not been comprehensively studied. As with BMI, we expected that any associations could be bidirectional: if a lower CD4+ count indicated more severe disease then we might expect a greater degree of disturbance to the serum proteome (i.e. negative association), however for those protein responses that are driven by CD4+ cell activity, there might be a positive association. We therefore performed two-sided analyses.

Analyses of CD4+ count as a binary variable in the studies listed above (a– e) was done using either a count of 50 or 200 cells/uL as the cut-off. In our study, 66 TB, HIV+ patients had CD4+ count data available. The distribution of CD4+ counts among these patients is shown below in Table 40, using CD4+ count categories as per Pepper *et al.*, ’15.(18)

CD4+ count	0 - 49	50-99	100-199	200-349	350-499	>/= 500
N	15	10	13	17	6	5

Table 40 Distribution of TB, HIV+ patients by CD4+ count category.

The total number of patients with CD4+ count data available was 66.

At a CD4+ cut-off of 50, the numbers of patients in each group was 15 and 51. At a CD4+ cut-off of 200, the numbers were 38 and 28. We therefore aimed to look for associations between protein levels and CD4+ in our TB, HIV+ group in the following ways:

i. with both CD4+ count and protein level as continuous variables to assess correlation;
ii. with CD4+ count stratified as < 200 vs. >/= 200 cells/uL to assess association of protein level with CD4+ count at a biologically meaningful stratification.

For ii., the cut-off of 200 was chosen to maintain approximately equal patient numbers in both groups and hence maximise statistical power.

The correlations of protein level with CD4+ count as a continuous variable are shown below in Table 41. Levels of ferritin, tPA, apo-AI, IP-10 and TNF-alpha all correlated with CD4+ count. Correlations were negative for all proteins, except apo-AI for which it was positive. All correlations were moderately strong, and statistically significant after correction for multiple comparisons.

	Rho	95% CIs	P	Corrected P
Procalcitonin	-0.13	-0.38 to 0.14	0.32	1.00
Ferritin	**-0.53**	**-0.7 to -0.31**	**0.00**	**0.00**
tPA	**-0.49**	**-0.67 to -0.26**	**0.00**	**0.00**
Fibrinogen	-0.06	-0.32 to 0.21	0.67	1.00
Serum amyloid A	0.03	-0.24 to 0.29	0.85	1.00
Alpha-2-M	0.20	-0.06 to 0.44	0.12	0.94
CRP	-0.01	-0.27 to 0.25	0.93	1.00
Haptoglobin	-0.01	-0.27 to 0.25	0.91	1.00
Serum amyloid P	0.10	-0.17 to 0.35	0.47	1.00
MMP-2	0.05	-0.21 to 0.29	0.72	1.00
MMP-9	0.00	-0.25 to 0.25	0.97	1.00
Apo-AI	**0.40**	**0.16 to 0.59**	**0.00**	**0.03**
Apo-CIII	0.12	-0.13 to 0.36	0.33	1.00
Transthyretin	0.23	-0.02 to 0.46	0.06	0.77
Complement FH	0.16	-0.09 to 0.4	0.20	0.99
TGF-alpha	-0.01	-0.26 to 0.25	0.97	1.00
IFN-alpha-2	-0.05	-0.30 to 0.20	0.67	1.00
IFN-gamma	-0.17	-0.41 to 0.08	0.17	0.99
IL-1ra	-0.30	-0.52 to -0.05	0.02	0.30
IP-10	**-0.43**	**-0.62 to -0.2**	**0.00**	**0.01**
TNF-alpha	**-0.45**	**-0.64 to -0.23**	**0.00**	**0.00**
VEGF	-0.03	-0.28 to 0.23	0.82	1.00

Table 41 Correlation of serum protein levels with CD4+ count as a continuous variable in the TB, HIV+ patients (n = 66).

Rho: Spearman's rank correlation coefficients. P-values are 2-sided and shown before and after Sidak's correction for multiple comparisons.

The raw data on these correlations is displayed below in Figure 57.

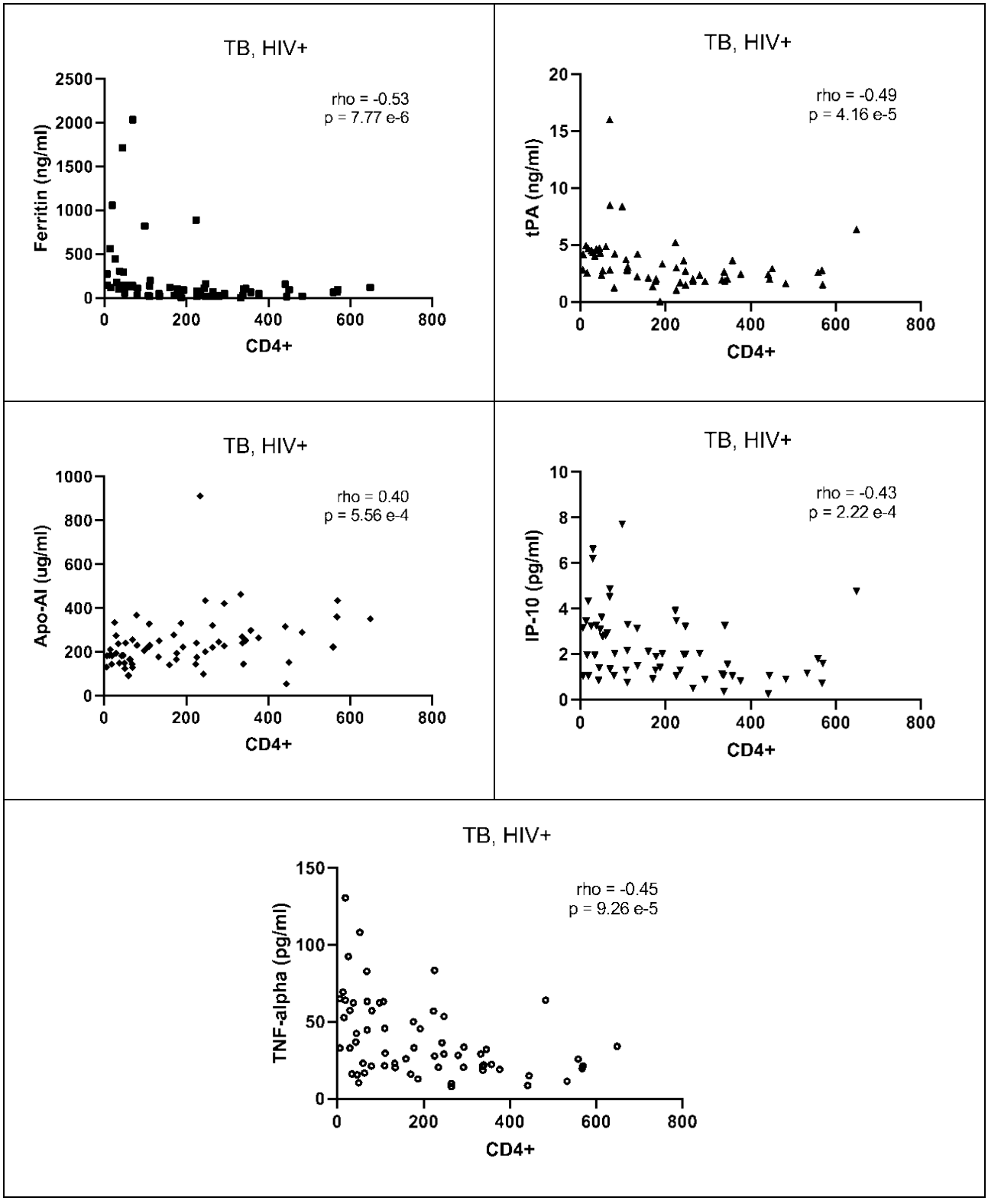


Figure 57 Levels of proteins in TB, HIV+ patients by CD4+ count as a continuous variable.

The associations between protein levels and CD4+ count as a binary variable, stratified as < vs. >/= 200 cells/uL are shown below in Table 42, below. Levels of tPA were significantly higher in patients with CD4+ < vs. >/= 200 cells/uL. These were no other statistically significant differences after correction for multiple comparisons.

	Med (IQR)			
	< 200	**>/= 200**	**P**	**Corrected P**
Procalcitonin	2.07 (1.63 - 2.64)	1.63 (1.2 - 2.43)	0.21	1.00
Ferritin	131.03 (57.72 - 274.94)	66.42 (24.36 - 100.74)	0.00	0.07
tPA	**3.74 (2.58 - 4.54)**	**2.35 (1.85 - 2.79)**	**0.00**	**0.04**
Fibrinogen	2.6 (2.21 - 3.67)	2.73 (2.26 - 3.88)	0.99	1.00
Serum amyloid A	5.61 (3.96 - 7.81)	6.27 (4.18 - 9.06)	0.50	1.00
Alpha-2-M	1.21 (1.02 - 1.49)	1.47 (1.17 - 1.65)	0.07	0.77
CRP	66.68 (46.1 - 95.31)	56.11 (22.39 - 97.84)	0.48	1.00
Haptoglobin	2.51 (0.35 - 8.06)	3.4 (0.26 - 7.75)	0.80	1.00
SAP	38.91 (30.62 - 48.34)	41.89 (31.91 - 50.72)	0.61	1.00
MMP-2	71.39 (59.4 - 80.76)	69.15 (57.85 - 96.12)	0.85	1.00
MMP-9	236.55 (146.25 - 408.27)	236.17 (143.77 - 371.91)	0.79	1.00
Apo-AI	195.46 (167.14 - 237.08)	253.22 (212.17 - 337.22)	0.01	0.16
Apo-CIII	75.75 (58.07 - 128.69)	105.56 (77.52 - 127.41)	0.14	0.97
Transthyretin	55.18 (41.61 - 88.37)	77.56 (50.92 - 125.12)	0.06	0.75
Complement FH	356.38 (290.53 - 441.46)	360.27 (329.6 - 415.42)	0.53	1.00
TGF-alpha	7.72 (4.88 - 13.17)	8.04 (5.01 - 11.67)	0.98	1.00
IFN-alpha-2	8.69 (0 - 66.04)	9.97 (0 - 45.39)	0.94	1.00
IFN-gamma	27.57 (12.96 - 70.22)	21.76 (8.77 - 30.04)	0.19	0.99
IL-1RA	26.74 (5 - 52)	5.02 (0.04 - 25.43)	0.01	0.11
IP-10	2.14 (1.39 - 3.28)	1.14 (0.91 - 2.01)	0.00	0.05
TNF-alpha	43.71 (22.64 - 62.66)	22.45 (19.49 - 32.93)	0.01	0.17
VEGF	347.26 (172.07 - 632.32)	239.83 (108.54 - 397.79)	0.22	1.00

Table 42 Protein levels in TB, HIV+ patients with CD4+ counts < 200 vs. >/= 200 cells/uL (n = 66).

IQR: interquartile range. P-values are 2-sided and derived from Mann-Whitney tests. p-values were corrected using Sidak's method. Results tPA are highlighted in bold as the difference between the two groups was statistically significant.

The raw data on the association bewteen tPA level and CD4+ as a binary variable is shown below in Figure 58.

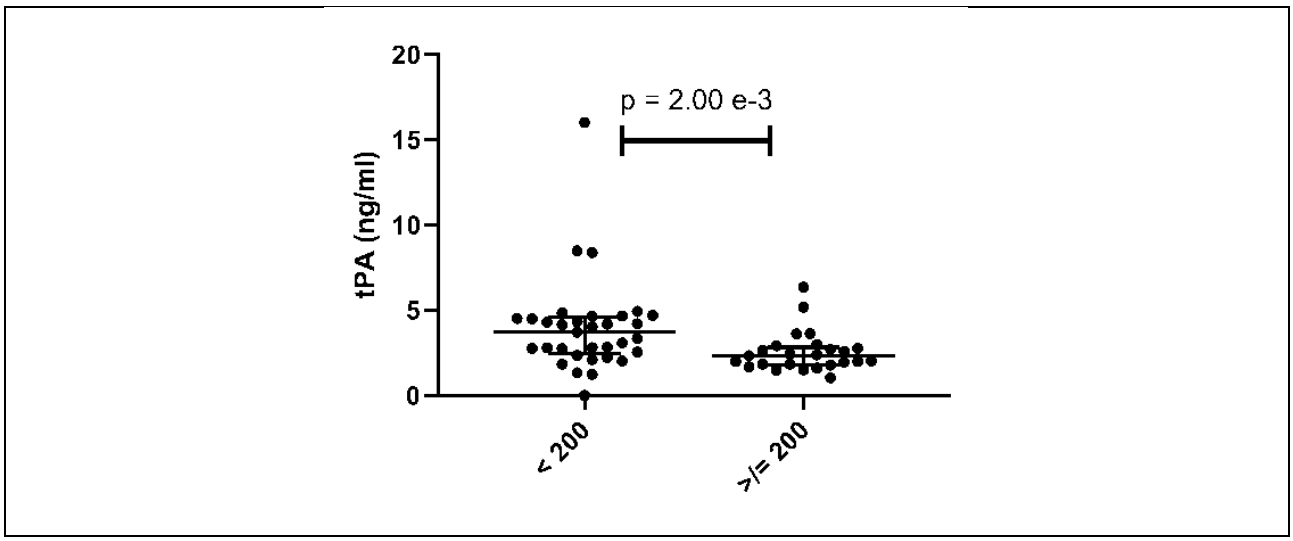


Figure 58 Levels of tPA in TB, HIV+ patients with CD4+ counts either < or >/= 200 cells/uL.

The p-value is 2-sided and derived from a Mann-Whitney test. Error bars represent medians with interquartile ranges.

Next, proteins that associated with each of the three clinical markers of severity (smear grade, BMI and, in the case of TB, HIV+, CD4+ count) were compared to look for overlaps that might increase the probability of their being an association between protein level and outcome itself. This analysis was done on all TB patients as well as on each of TB, HIV- and TB, HIV+ patients. As shown below in Table 43, in all TB patients, TGF-alpha was significantly associated with both smear grade and BMI. In TB, HIV+ patients, apo-AI was associated with both smear grade and CD4+ count.

	Smear grade	BMI	CD4+
All TB pts	CRP* MMP-9* TGF-alpha	Procalcitonin tPA Fibrinogen* TGF-alpha VEGF	*n/a*
TB, HIV-	MMP-2	-	*n/a*
TB, HIV+	Fibrinogen* MMP-9 Apo-AI TGF-alpha	-	Ferritin tPA Apo-AI IP-10* TNF-alpha*

Table 43 Comparison of biomarkers significantly associated with each of the clinical markers of severity (smear grade, BMI and CD4+ count).

Associations were considered significant if $p < 0.05$ after correction for multiple comparisons. Proteins that were associated with more than one clinical marker of severity are underlined. Proteins that were part of the 9-protein PReMS signature are marked with an asterisk. Protein levels were positively associated with smear grade and negatively associated with BMI and CD4+ count, except for MMP-2 and apo-AI for which the directions were opposite.

Correlations were then sought between the relevant pairs of clinical markers of severity themselves, to exclude any associations that might have confounded these findings. In all TB patients, there was a trend towards a negative correlation between smear grade and BMI (Spearman's rho -0.09 (95% CIs: -0.25 – 0.08), p = 0.15). In the TB, HIV+ patients, there was a trend towards a higher proportion of smear +ve patients in those with CD4+ < 200 cells/uL (p = 0.06, one-sided Fisher's exact test). The numbers of smear +ve and -ve TB, HIV+ patients by CD4+ count category are shown below in Table 44.

	CD4+ < 200	**CD4+ >/= 200**
Smear -ve (%)	5 (36)	9 (64)
Smear +ve (%)	33 (63)	19 (37)

Table 44 Numbers of TB, HIV+ patients who were smear -ve vs. +ve by each of CD4+ count < vs. >/= 200 cells/uL.

Bracketed values represent percentages of row totals.

To better define the relationship between TGF-alpha level and each of smear grade and BMI in all TB patients, and between apo-AI level and each of smear grade and CD4+ count in TB, HIV+ patients, we went on to perform multivariate analyses. We aimed to evaluate the independence of the associations between the clinical variables and protein level in each case. As shown below in Table 45, in all TB patients, both smear grade and BMI were associated with TGF-alpha level. As shown in Table 46, however, in TB, HIV+ patients, CD4+ count was associated with apo-AI level, but smear grade was not.

Variable	**|t|**	**p-value**
Intercept	6.753	0.00
BMI	**3.778**	**0.00**
Smear grade	**2.780**	**0.01**

Table 45 Multivariate analysis of associations between smear grade and BMI with TGF-alpha level in all TB patients.

TGF-alpha levels were decile-normalised before analysis. Least squares multiple linear regression was then performed in Prism v. 8.3. Residuals were normally distributed by Anderson-Darling and Kolmogorov-Smirnov tests. Degrees of freedom: 130; R^2: 0.16. Bold font indicates exposure variables that were significantly associated.

Variable	|t|	p-value
Intercept	7.339	<0.0001
CD4+ count	**3.430**	**0.0011**
Smear grade	1.745	0.0859

Table 46 Multivariate analysis of associations between smear grade and CD4+ count with apo-AI level in TB, HIV+ patients.

tPA levels were decile-normalised before analysis. Least squares multiple linear regression was then performed in Prism v. 8.3. Residuals were normally distributed by Anderson-Darling, D'Agostino-Pearson, and Shapiro-Wilk tests. Degrees of freedom: 62; R^2: 0.19. Bold font indicates exposure variables that were significantly associated.

iv) Age

The definition of “old age” varies depending on which part of the world or population is being referred to and on the metric that is being used. The Office for National Statistics in the UK define old age as being over 65 years since birth, whereas the World Health Organization refer to “older people” as those who have merely passed the median life expected for that population.(35, 36) From a biological perspective, ageing results from the gradual accumulation of damaging insults at the molecular and cellular level, resulting in a gradual decline in physiological reserve.

The distribution of ages across the active TB group in our study, including by each of HIV status and site was examined in order to inform how best to look for associations between serum protein level and age in TB. The results are shown below in Figure 59. There was no significant difference in age by either HIV status or site. There was one patient from the Cape Town TB, HIV+ group (female) that was excluded from these analyses as she was recorded as having an age of 9 years. In the clinical database, however, it was noted that this was an estimated rather than actual age. The height and weight were in keeping with adulthood (160cm and 50 kg), hence this patient had been kept in the study for all other analyses other than age.

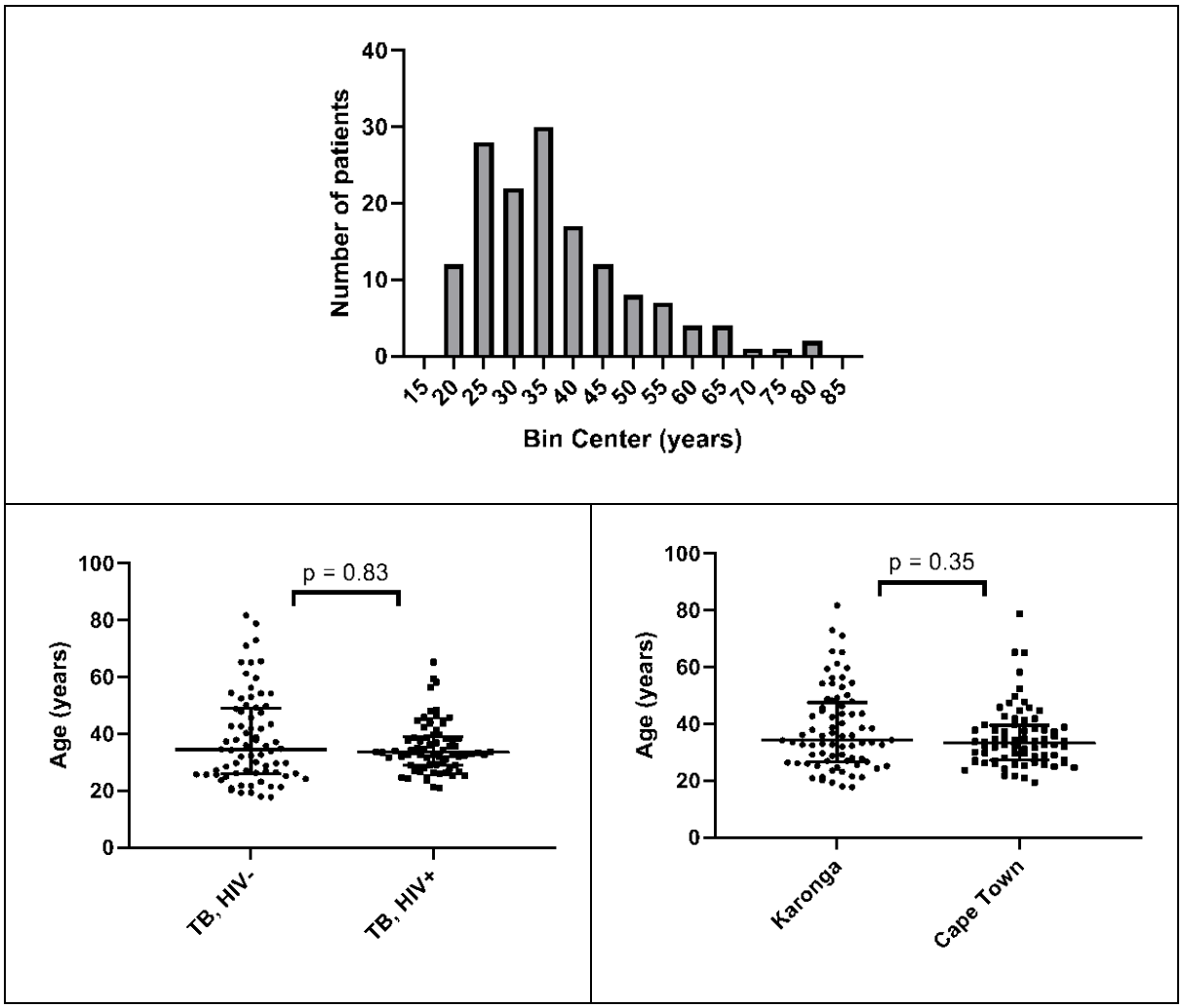


Figure 59 Distribution of ages in patients with active TB.

A: All TB patients. B and C: Comparison of ages by HIV status and site respectively. P-values are two-sided and were derived from Mann-Whitney tests.

Because the application of a cut-off to signify older age was felt to be arbitrary, and because the range of ages across our TB cohort was wide (18 – 82 years), we decided to analyse age as a continuous variable only. All TB patients were also considered together since neither HIV status nor site were felt to represent significant confounders. Older age is a risk factor for acquiring TB,(8) but TB disease can also be more severe in older age groups.(37) We envisaged that serum protein responses might either be deficient or exaggerated in older age, and thus performed two-sided tests.

The correlations of each protein with age is shown below in Table 47. Levels of haptoglobin were weakly but significantly negatively correlated with age.

	Rho	95% CIs	P	Corrected P
Procalcitonin	0.00	-0.18 to 0.18	0.87	1.00
Ferritin	-0.14	-0.31 to 0.03	0.14	0.96
tPA	0.05	-0.13 to 0.22	0.43	1.00
Fibrinogen	-0.01	-0.18 to 0.17	0.90	1.00
Serum amyloid A	-0.13	-0.30 to 0.05	0.23	1.00
Alpha-2-M	-0.07	-0.24 to 0.11	0.43	1.00
CRP	-0.22	-0.38 to -0.05	0.02	0.40
Haptoglobin	**-0.26**	**-0.42 to -0.09**	**0.00**	**0.03**
Serum amyloid P	-0.17	-0.34 to 0.00	0.06	0.76
MMP-2	0.00	-0.16 to 0.17	0.83	1.00
MMP-9	-0.26	-0.41 to -0.10	0.00	0.07
Apo-AI	0.27	0.10 to 0.42	0.00	0.06
Apo-CIII	0.00	-0.17 to 0.17	0.81	1.00
Transthyretin	0.00	-0.17 to 0.17	0.81	1.00
Complement FH	-0.16	-0.32 to 0.01	0.07	0.79
TGF-alpha	-0.27	-0.42 to -0.11	0.00	0.07
IFN-alpha-2	-0.09	-0.26 to 0.08	0.34	1.00
IFN-gamma	-0.16	-0.32 to 0.02	0.05	0.70
IL-1ra	-0.20	-0.36 to -0.03	0.03	0.43
IP-10	-0.08	-0.25 to 0.09	0.39	1.00
TNF-alpha	-0.22	-0.38 to -0.06	0.01	0.19
VEGF	-0.06	-0.23 to 0.12	0.64	1.00

Table 47 Correlations of serum protein levels with age in patients with active TB.

All patients were included in the analyses, regardless of HIV status or site. Rho refers to Spearman's rank correlation coefficient. P-values are 2-sided and shown before and after Sidak's correction for multiple comparisons.

The scatter plot for Haptoglobin is shown below in Figure 60.

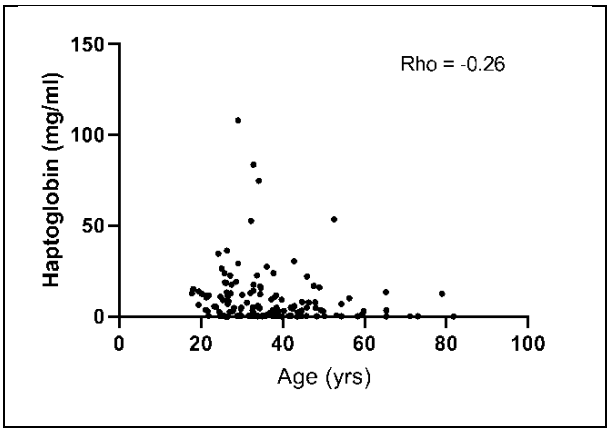


Figure 60 Correlation of serum levels of Haptoglobin with age in patients with active TB.

All TB patients were included, regardless of HIV status. Rho represents Spearman's rank correlation coefficient.

v) Gender

Protein levels were compared between male and female TB patients regardless of HIV status or site. Two-sided Mann-Whitney tests were used to look for a difference in levels in either direction since the literature was conflicting with regards to what to expect (as reviewed by Seddon *et al*., '18).(9) As shown below in Table 48, the only protein whose levels were significantly different by gender after correction for multiple comparisons was ferritin, which was higher in male TB patients.

	Female	Male		
Protein	**Median (IQR)**	**Median (IQR)**	**P-value**	**Corrected P**
Procalcitonin	1.79 (1.35 - 2.415)	2.02 (1.2 - 2.88)	0.24	1.00
Ferritin	**49.92 (25.3 - 95.42)**	**107.8 (52.05 - 176.4)**	**0.00**	**0.03**
tPA	2.4 (1.665 - 3.185)	2.87 (2.13 - 4.125)	0.01	0.21
Fibrinogen	2.54 (1.995 - 3.325)	2.8 (2.02 - 3.675)	0.26	1.00
Serum amyloid A	6.66 (4.15 - 8.61)	6.05 (4.313 - 8.893)	1.00	1.00
Alpha-2-M	1.36 (1.16 - 1.625)	1.52 (1.27 - 1.875)	0.02	0.37
CRP	56.55 (24.27 - 91.97)	67.14 (34.95 - 102.3)	0.27	1.00
Haptoglobin	4.81 (0.385 - 12.85)	4.86 (0.32 - 12.8)	0.68	1.00
Serum amyloid P	43.36 (33.67 - 51.38)	41.97 (34.43 - 52.56)	0.85	1.00
MMP-2	69.94 (57.85 - 95.13)	81.45 (65.01 - 102.7)	0.14	0.97
MMP-9	308.5 (167.1 - 545.5)	362.4 (248.9 - 511.9)	0.26	1.00
Apo-AI	225.7 (178.5 - 318.5)	232.2 (167.1 - 278.7)	0.66	1.00
Apo-CIII	100.1 (73.66 - 139.4)	80.44 (60.35 - 115.7)	0.00	0.09
Transthyretin	70.07 (44.72 - 94.02)	69.18 (50.66 - 105.9)	0.43	1.00
Complement factor H	383.2 (323.5 - 431.1)	403 (323.4 - 465.6)	0.37	1.00
TGF-alpha	9.88 (5.35 - 17.97)	11.32 (6.1 - 20.19)	0.31	1.00
IFN-alpha-2	0 (0 - 35.65)	0 (0 - 30.22)	0.87	1.00
IFN-gamma	15.95 (7.73 - 41.02)	16.36 (7.5 - 46.65)	0.88	1.00
IL-1ra	17.06 (3 - 42.74)	16.36 (1.54 - 54.7)	0.87	1.00
IP-10	1.12 (0.76 - 2.05)	1.05 (0.61 - 2.08)	0.59	1.00
TNF-alpha	25.85 (16.21 - 38.95)	23.32 (15.03 - 40.41)	0.62	1.00
VEGF	252.7 (106.4 - 442.4)	373.9 (156.1 - 792.2)	0.04	0.56

Table 48 Comparison of levels between female and male TB patients, regardless of HIV status or site.

P-values are two-sided, derived from Mann-Whitney tests and shown before and after Sidak's correction.

Ferritin levels on each individual TB patient are displayed by gender in Figure 61 below.

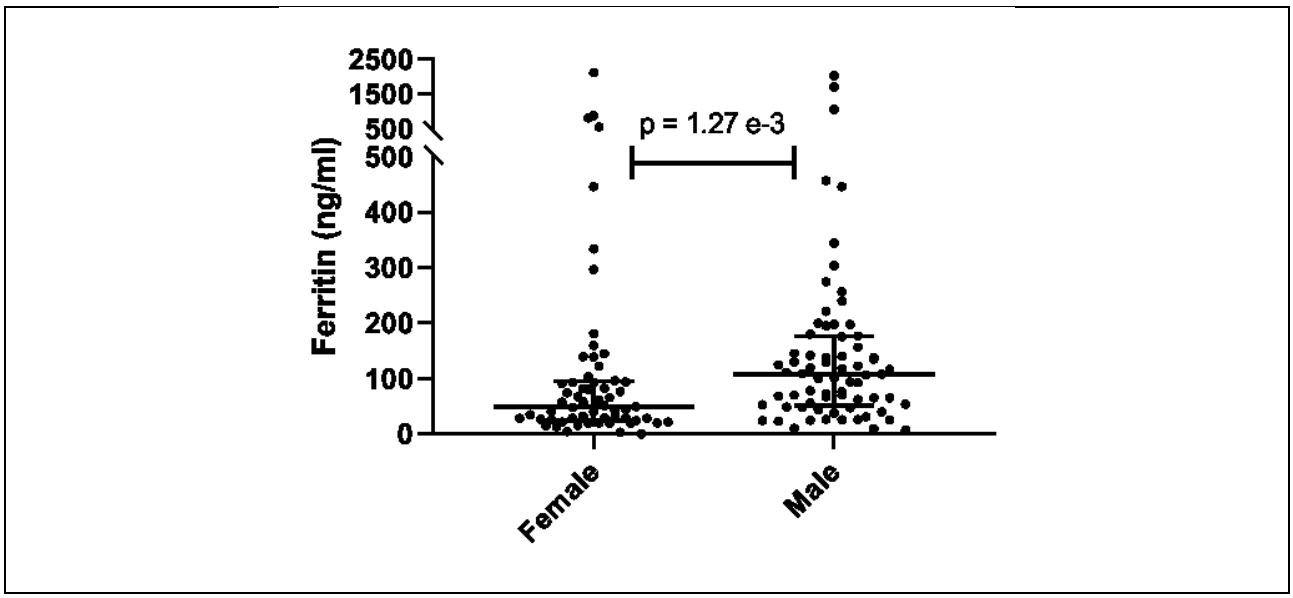


Figure 61 Comparison of levels in female vs. male TB patients, regardless of HIV status or site.

P-value is derived from two-sided Mann-Whitney test and is shown before correction.

In order to explore possible reasons for this association, we performed a multivariate linear regression analysis, since we had already seen a trend towards associations between ferritin levels and HIV status (Table 30), smear grade (Table 32), BMI (Table 35), and age (Table 46). In addition, ferritin levels were significantly associated with CD4+ count in TB, HIV+ patients (Table 40). We had not previously explored association of any variables with study site, however we included this variable for completeness as well. Ferritin levels were not normally distributed, so we decile-normalised them before performing the analysis. As shown below in Table 49, age, gender, HIV status and BMI were all independently associated with ferritin level. The most statistically significant exposure variables were gender (p = 0.008) and HIV status (p = 0.004).

Variable	|t|	P-value
Intercept	4.091	0.00
Age	**2.210**	**0.03**
Gender	**2.681**	**0.01**
Site	0.5065	0.61
HIV status	**2.951**	**0.00**
BMI	**2.058**	**0.04**
Smear grade	1.414	0.16

Table 49 Results of a multivariate linear regression analysis to assess independent association of multiple different variables with serum ferritin level.

All TB patients were included in this analysis, which was performed in Prism v8.3.0 after decile-normalising ferritin levels. Degrees of freedom: 116; R^2: 0.22. Residuals were normally distributed according to Andersen-Darling, Shapiro-Wilk and Kolmogorov-Smirnov tests.

Because of the association with CD4+ count, we repeated this analysis after stratifying the TB group by HIV status. Results for TB, HIV- patients are shown first (Table 50) and for TB, HIV+ are shown second (Table 51). In TB, HIV- patients, there was no independent association between any of the exposure variables and ferritin level, including gender, with a trend towards smear grade being independently associated (p = 0.05).

Variable	|t|	P-value
Intercept	3.101	0.00
Age	1.706	0.09
Gender	0.4999	0.62
Site	0.8112	0.42
BMI	1.295	0.20
Smear grade	1.979	0.05

Table 50 Results of a multivariate linear regression analysis to assess independent association of multiple different variables with serum ferritin level in HIV uninfected TB patients.

Degrees of freedom: 53; R^2: 0.25. Residuals were normally distributed according to Andersen-Darling, D'Agostino-Pearson, Shapiro-Wilk and Kolmogorov-Smirnov tests.

For TB, HIV+ patients, the extra exposure variables of CD4+ count and antiretroviral therapy (ART) status were added. As shown in Table 51, BMI, CD4+ count and ART status were all independently associated with ferritin level. Gender was not independently associated with ferritin level in this clinical subgroup (p = 0.16).

Variable	|t|	P-value
Intercept	5.097	0.00
Age	0.7026	0.49
Gender	1.431	0.16
Site	0.5841	0.56
BMI	**2.157**	**0.04**
Smear grade	0.2975	0.77
CD4+ count	**2.238**	**0.03**
ART status	**2.083**	**0.04**

Table 51 Results of a multivariate linear regression analysis to assess independent association of multiple different variables with serum ferritin level in HIV co-infected TB patients.

Degrees of freedom: 49; R^2: 0.36. Residuals were normally distributed according to Andersen-Darling, D'Agostino-Pearson, Shapiro-Wilk and Kolmogorov-Smirnov tests. ART: antiretroviral therapy.

To understand the relationship between ferritin level, CD4+ count and gender further, we plotted the levels of ferritin by gender in each of the TB, HIV- and TB, HIV+ halves of the cohort. As shown below in Figure 62, the difference in ferritin level by gender was most significant in the TB, HIV+ group.

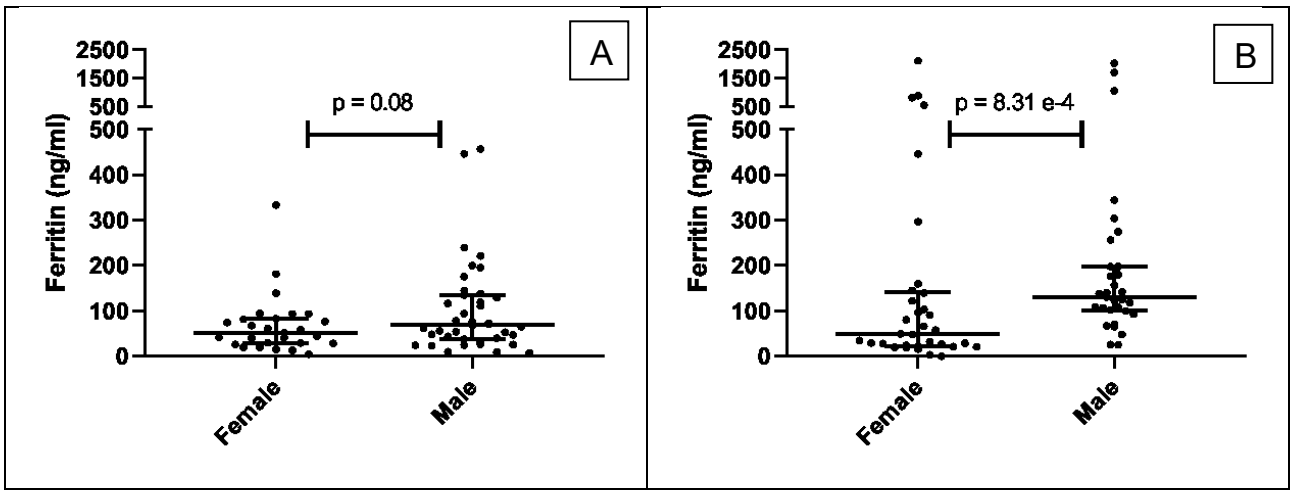


Figure 62 Comparison of levels in female vs. male TB patients after stratification by HIV status.

A: TB, HIV- patients. B: TB, HIV+ patients. P-values are one-sided and derived from Mann-Whitney tests.

We then repeated the regression analysis for TB, HIV+ patients but after removing CD4+ count and ART status. As shown in Table 52, without these extra exposure variables, an independent association between gender and ferritin was again seen.

Variable	|t|	P-value
Intercept	4.740	0.00
Age	0.3207	0.75
Gender	**2.526**	**0.01**
Site	0.3421	0.74
BMI	**2.015**	**0.05**
Smear grade	0.2979	0.77

Table 52 Results of a multivariate linear regression analysis to assess independent association of multiple different variables with serum ferritin level in HIV co-infected TB patients, with CD4+ count and ART status excluded.

Degrees of freedom: 58; R^2: 0.24. Residuals were normally distributed according to Andersen-Darling, D'Agostino-Pearson, Shapiro-Wilk and Kolmogorov-Smirnov tests.

We therefore concluded that the apparent association between gender and ferritin level was confounded by CD4+ count and/or ART status in the TB, HIV+ group. We therefore examined the distributions of CD4+ count and ART status across TB, HIV+ patients by gender. As shown below in Figure 63, there was a trend towards CD4+ counts being lower in male patients (p = 0.16). As shown in Table 50, there was also a trend towards a lower proportion of male TB, HIV+ patients being on ARVs than women (Table 53).

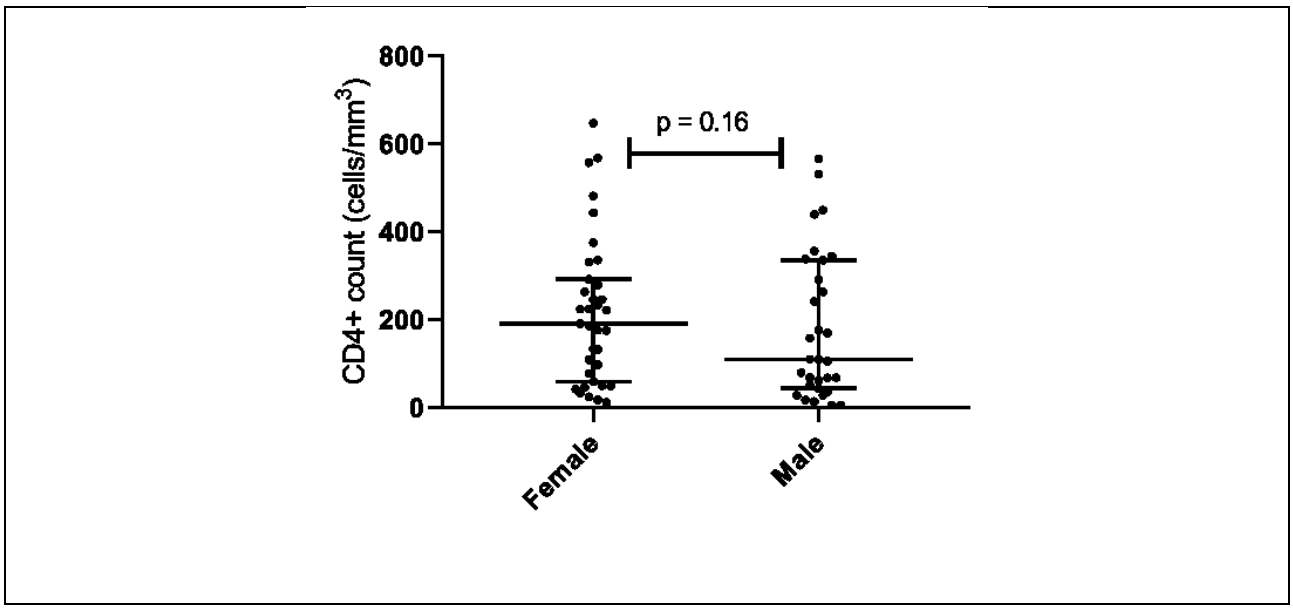


Figure 63 Comparison of CD4+ count by gender in TB, HIV+ patients.

The p-value is one-sided and derived from a Mann-Whitney test.

On ART?	N	Y
Male	33	3
Female	32	7

Table 53 Numbers of TB, HIV+ patients on antiretroviral therapy (ART) by gender.

Fisher's exact test for the proportion of female patients on ART being higher than that of male patients: p = 0.19.

4. **Discussion**

The results from each of the above analyses will now be summarised and discussed in turn.

i) The TB disease state

Whilst this was not an untargeted proteomic study, two main observations can be made from the study of these 22 proteins. To facilitate this discussion, a summary of the biology of each protein together with direction of change in TB and possible effect on the disease process is shown below in Table 26. Firstly, the state of active TB is associated with numerous different biological phenomena. All of these are well-documented, however not always at the level of every individual protein. These include:

a) Acute Phase Response (APR)

This is defined by increased or decreased production from the liver in response to IL-1-beta, IL-6 and TNF-alpha (as reviewed by Gabay and Kushner, '99).(38) The proteins in this study that are affected by this process were procalcitonin, ferritin, fibrinogen, serum amyloid A, CRP, haptoglobin, serum amyloid P, apo-AI, apo-CIII and transthyretin. The presence of an APR in TB is well described and has been long known, however serum amyloid P has not previously been shown to be elevated in TB in comparison with LTBI or healthy controls;(39)

b) Th1 response

One of the cornerstones of the understanding of the response to TB is that antigen-specific CD4+ T cells are directed towards a Th1 phenotype.(40) This is reflected in the elevated levels of IFN-gamma and IP-10 in our data;

c) Matrix Degrading Phenotype

As will be discussed further below, the relevance of decreased serum MMP-2 concentrations in TB is unclear, but increased MMP-9 production is well documented and associated with tissue matrix degradation;(41)

d) Hypoxia

Nuclear medicine imaging has recently demonstrated that whilst TB re-activation preferentially occurs in well-oxygenated areas of the lung, profound local hypoxia often develops, which in turn drives cavitation.(42) The elevated levels of VEGF and TGF-alpha in our study are in keeping with this, and represent a new finding regarding TGF-alpha;

e) Altered lipid metabolism

Apo-AI is well known to be a "negative acute phase reactant", i.e. levels decrease in an APR.(43) Apo-CIII is also in this category, albeit less well documented.(44) To our knowledge, this is the first time that levels of either of these proteins has been shown to be lower in TB than LTBI or healthy controls, although there is supporting evidence in both cases ("smear grade section" below, and AE-TBC study). Apo-AI is the principle protein component of high-density lipoproteins (HDL), which are responsible for cholesterol recovery from tissue. Apo-CIII, by contrast, is a component of all lipoproteins, involved in synthesis and maintainence of integrity. It is also known that levels of apoB and low-density lipoprotein (LDL) are lower in TB,(45) hence what seems to occur is a suppression of systemic lipid transport in both directions (both to and from the liver). This pattern is not unique to TB however two possible

implications are firstly that a lipid-rich diet may not be effective for calorie replacement in the early phases of TB treatment if the proteins required for lipid transport are less readily available, but secondly that there may be scope for therapies targeting lipid retrieval from tissues.

The second observation is that various opportunities for host-directed therapy are revealed. These include modulation of iron metabolism, MMP inhibition and modulation of lipid metabolism. Of these, our data adds to the understanding of altered lipid metabolism and opportunities for intervening in this will therefore be discussed. In addition, our novel findings of alterations in serum concentrations of serum amyloid P and complement factor H will also be briefly discussed with regards to therapeutic opportunities.

The goal of Mtb in the process of reactivation is cavitation, however there are various features in the stages before this that point to a local dysregulation of lipid metabolism. As reviewed by Russell *et al*., ’07, amongst the innate immune cells in the centre of the granuloma are lipid-laden macrophages.(46) These “foam cells” are commonly seen at sites of chronic inflammation, however as TB progresses, this entire granuloma centre breaks down in a process known as caseation. A recent transcriptomic, proteomic and lipidomic analysis of human TB granulomas showed that transcripts and proteins involved in lipid sequestration and metabolism were highly expressed in these caseous granulomas, and that the caseum itself was comprised of lipid species closely resembling the constitution of LDL.(47) Furthermore, infection of human PBMC-derived and mouse macrophages with Mtb reproduced the development of these foam cells, as did (in mice) the adminstration of the Mtb cell wall glycolipid trehalose dimycolate. In short, Mtb induces local lipid sequestration and storage in the macrophage that may then subsequently provide the caseous, liquefactive environment that is needed for expectoration. In the light of our data on apo-AI and apo-CIII, possible therapeutic avenues may therefore include infusion of recombinant HDL particles with the aim of enhancing lipid recovery from tissues and limiting caseum development, or co-administration of drugs such as statins which lower LDL but raise HDL. The former approach may be prohibitively expensive and technologically challenging, though safe, having been trialled in the area of atherosclerosis.(48) It may therefore be worth considering in the era of extensively drug-resistant TB. Lipid-lowering agents, however, may be considerably more feasible, being widely available and, in some cases, indicated for co-existing metabolic syndrome. There is mounting evidence for statins, in fact, including that they improve macrophage function against Mtb in mice and *ex vivo*, probably by facilitating lysosome-phagosome membrane fusion,(49) shorten the duration of TB treatment in mice,(50) and appear to reduce the risk of reactivation in people with LTBI.(51) A well-designed randomised controlled trial of adjunctive statin therapy in TB is awaited.

Serum amyloid P is a pentraxin, which are soluble and innate pathogen recognition receptors produced as part of the APR.(52) Serum amyloid P does not appear to bind mycobacteria, however, and in PBMC-derived human macrophages *in vitro* has been shown to potentiate Mtb growth.(53) This is possibly due to the ability of serum amyloid P to induce alternatively activated (or M2) macrophages, which are less able to control Mtb.(54) Co-administration of compounds that block the interaction of serum amyloid P were shown *in vitro* to mitigate the increased growth, hence opening up an avenue for adjunctive therapy.(53)

Complement factor H is a novel biomarker of TB in secondary care settings, and not previously shown to be elevated in TB in comparison with LTBI/healthy controls. It is an inhibitor of the alternative pathway, serving to

protect the host from excessive cell surface-associated complement activation.(55) Multiple pathogens, including *Mycobacterium bovis BCG* (BCG), have developed the ability to bind complement FH themselves, however, which in the case of BCG has been shown to inhibit phagocytosis by human macrophages.(56) The precise ligand on the surface of BCG to which complement FH binds remains to be determined, as does a study of whether the same phenomenon occurs with Mtb, however there is potential here for interfering with disease progression, whether soon after the initial infection event, or later.

Change	Source(s)	Cytokines implied	Biological function	Effect in TB	Known?
Procalcitonin	Thyroid, adipocytes	IL-6, TNF-alpha(57)	Unknown	*n/a*	Yes(58)
Ferritin	Liver	TNF-alpha, IL-1-beta(59)	Iron storage/transport, IL-10 mediated suppression of delayed type hypersensitivity(59)	Sequestration of Fe in lung tissue; dampening of cell-mediated immunity	Yes(60)
tPA	Vascular endothelium	type I and II IFNs(61)	Fibrinolysis	Uncertain	**No**
Fibrinogen	Liver, lung	IL-6(62)	Clot formation	Uncertain	Yes(63)
Serum amyloid A	Liver	IL-6, TNF-alpha, IL-1-beta, IFN-gamma(64)	Cholesterol recovery and acute phase reactant(64)	Uncertain	Yes(65)
CRP	Liver	IL-1-beta, IL-6, IL-17(66, 67)	Opsonisation and complement fixation	Uncertain	Yes(68)
Haptoglobin	Liver	IL-6, TNF-alpha(69)	Binds free haemoglobin	Possibly limits replication of Mtb by preventing accumulation of Fe in macrophages(70)	Yes(71)
Serum amyloid P	Liver	IL-1, IL-6(72)	Pathogen recognition, inhibition of PMN accumulation, fibrosis	Administration may limit lung pathology but promote growth of Mtb(53)	**No**
MMP-2	Monocytes, smooth muscle	TNF-alpha(73)	Tissue damage/repair	Unclear	Yes(74)
MMP-9	Monocytes, PMNs	IL-1, TGF-beta(75)	Tissue damage/repair	Matrix degradation	Yes(76)
apo-AI	Liver	TNF-alpha, IL-1-beta	Cholesterol recovery	Possibly promotes development of caseum	Yes

Change	Source(s)	Cytokines implied	Biological function	Effect in TB	Known?
apo-CIII	Liver	TNF-alpha, IL-1-beta(44)	Cholesterol delivery	Possibly limits development of caseum	**No**
Transthyretin	Liver	IL-6(77)	T4 and retinol transport	Uncertain	Yes(78)
Complement factor H	Liver, lung(79)	IFN-gamma(80)	Complement pathway inhibitor	Protects host tissues, and possibly mycobacteria	**No**
TGF-alpha	Widespread	*n/a*	Tissue growth and repair	Unknown	**No**
IFN-gamma	Th1, CD8+ and NK cells	*n/a*	Th1 response	Macrophage activation	Yes(39)
IP-10	Monocytes, endothelial cells, fibroblasts	*n/a*	Chemokine	Chemotaxis and activation of monocytes and T cells	Yes(81)
TNF-alpha	Monocytes, PMNs, dendritic cells, T cells	*n/a*	Acute inflammation	Formation and maintenance of granulomata	Yes(82)
VEGF	Widespread (hypoxia)	*n/a*	Angiogenesis	May improve perfusion of disease site	Yes(83)

Table 54 The source(s), causative cytokines, biological functions and possible effects in TB of proteins whose serum concentrations are altered in TB.

Colour is used to indicate broad categories of biological function; red: acute phase response; orange: matrix degradation; green: lipid metabolism; blue: hypoxia; purple: Th1 response. Black indicates ungrouped proteins.

ii) The impact of HIV

The pathogenesis of HIV-associated TB (TB/HIV) is distinct from that of TB in HIV uninfected individuals in that it most commonly involves a progression of primary infection with Mtb, rather than a re-activation of previously asymptomatic latent TB. As a result, the host response is expected to differ. Indeed, HIV/TB occurs in individuals with impaired cell-mediated immunity and is characterised by immune activation, high rates of systemic dissemination, and, in the most unwell patients, a sepsis-like syndrome (as reviewed by Bell and Noursadeghi, '17).(84) Many of these features are well documented, however a comparison of the changes in the serum proteome with that of TB in HIV uninfected patients has not been thoroughly investigated. The two studies listed in the Introduction Chapter that recruited patients with HIV co-infection (Achkar *et al.*, '15 and Chegou *et al.*, '16) both showed a comparison of the individual protein ROC AUCs by HIV status,(85, 86) but this was to distinguish TB from OD, not LTBI/healthy controls. Achkar *et al* did show via heat map the changes in protein abundance relative to LTBI/healthy individuals by HIV status, but no numerical values or statistical tests were applied to this, making it difficult to interpret.(86) A study more recently by Lesosky *et al* measured concentrations of 13 cytokines in plasma of HIV co-infected individuals with and without active TB.(87) No HIV uninfected control group was included, but of the 22 markers that were quantified in our study, concentrations of IFN-gamma, IP-10 and TGF-alpha were significantly higher in the TB group (as in our dataset). In contrast to our dataset, however, concentrations of TNF-alpha and VEGF were not higher in the TB group. As a results of the study design, however, 60% of the TB patients were asymptomatic at recruitment, rendering it a much less severe cohort.

In our study, two comparisons were performed to shed light on the pathogenesis of HIV/TB: protein concentrations in TB, HIV+ vs. LTBI, HIV+, and in TB, HIV- vs. TB, HIV+. Three main observations are made from these results. Firstly, the acute phase response (APR) appears to be relatively preserved in HIV/TB. Levels of haptoglobin and apo-CIII were not significantly different after p-value correction in TB, HIV+ vs. LTBI, HIV+, whereas they had been in the equivalent comparison in HIV uninfected individuals. However, in both cases, concentrations of these proteins were also different between the respective LTBI groups (i.e. LTBI, HIV+ vs. LTBI, HIV-). On comparing concentrations in TB, HIV- and TB, HIV+, these proteins were not significantly different. Their responses may have been attenuated in TB, HIV+, but overall, the APR in HIV/TB appears to be intact. Levels of alpha-2-M were significantly lower in TB, HIV+ vs. TB, HIV-, however this protein appeared to be an outlier compared to the other acute phase proteins and reasons for this are unclear.

Secondly, the production of proteins that result from inflammatory activity at the disease site is impaired in HIV/TB. Namely, levels of complement factor H were not higher in TB, HIV+ vs. LTBI, HIV+, and were also lower in TB, HIV+ vs. TB, HIV-. In addition, MMP-9 and TGF-alpha were also lower in TB, HIV+ vs. TB, HIV-. This is in keeping with the known impairment of granuloma formation and subsequent necrosis and cavity formation in HIV/TB.

Thirdly, what we have labelled as "Th1 responses" appear to be preserved in HIV/TB. This is counter-intuitive, since advancing HIV-1 infection causes a marked CD4+ cytopaenia, including a reduction in Mtb-specific Th1 cells.(88) Nonetheless, serum concentrations of IP-10 and TNF-alpha were significantly higher in TB, HIV+ vs. TB, HIV-. Levels of IFN-gamma were also higher in TB, HIV+, although not significantly after p-value correction. These cytokines are in fact produced by multiple cell types, including NK and CD8+ T cells in the case of IFN-

gamma. It is therefore likely that the phenomenon of immune activation is behind these findings, as a result of rising HIV-1 viral as well as Mtb bacterial loads. In addition, levels of IL-1ra were not higher in TB, HIV+ vs. TB, HIV- after p-value correction, however they were significantly higher in TB, HIV+ vs. LTBI, HIV+, whereas this wasn't the case in the equivalent HIV uninfected comparison (TB, HIV- vs. LTBI, HIV-). IL-1ra is an anti-inflammatory cytokine produced by innate immune cells simultaneously to IL-1 beta and serves to block the activity of IL-1 alpha and -beta at the IL-1 receptor. IL-1 beta itself has a very short half-life in plasma, hence IL-1ra is often measured as a surrogate.(89) IL-1ra has been shown previously to be elevated in plasma in TB vs. healthy controls;(90) elevated in BAL fluid in TB vs. healthy controls;(91) and higher in plasma in patients with HIV/TB that died vs. survived.(92) It has not, to our knowledge, been shown previously to be higher in relation to LTBI in HIV co-infected individuals vs. uninfected, hence this finding in our dataset adds to current knowledge. It has been shown to be a marker of successful treatment response in TB, HIV+ however, which is in keeping with our finding.(93)

iii) Smear grade

The largest study in the literature of associations between serum proteins and sputum bacterial load is a study by Sigal *et al* ('17) in which sera from 319 TB patients (35 with HIV (11%)) were tested for levels of 70 pre-selected markers that had previously been found to be either diagnostic biomarkers of TB or markers of treatment response.(94) Recruitment was performed across 24 sites, of which 16 were in North America, 4 in Africa, 2 in South America, and 1 in each of Europe and South East Asia. Sputum load was measured either by TTD in liquid culture or smear grade >/= 2+ vs. < 2+. Baseline radiographical information from CXR was also used: the presence of cavities, the presence of cavities > 4cm diameter, and >50% lung involvement. Repeat sputum cultures were also obtained on these patients at weeks 8, 12 and 16 of treatment because they were recruited into a clinical trial. Serum protein measurements were also repeated at these timepoints. After correction for multiple comparisons, proteins significantly associated with baseline sputum bacterial load were IFN-gamma (with TTD) and SAA (with smear grade >/= 2+). In addition, out of the 22 proteins measured in our study, CRP, SAA and MMP-9 were associated with the presence of cavities at baseline, CRP and SAA with the presence of cavities > 4cm, and CRP, SAA, MMP-9, IFN-gamma, IP-10, PCT and VEGF with >50% lung involvement.

In our data, when all TB patients were considered, there were associations seen between smear grade and levels of CRP, MMP-9 and TGF-alpha. The correlations were positive in all cases and were weak but statistically significant. These results are unsurprising in the sense that levels of all these proteins were higher in TB patients than the healthy LTBI controls, and so a positive correlation with bacterial load is in keeping with the recently emerged concept of "molecular distance to health".(95) This is simply a recognition of the fact that the more clinically unwell the patient, the more disturbed any of the various host "biomes" can be expected to be. A brief discussion of the associations with each protein in turn is now given.

CRP is the serum protein most commonly used in clinical practice for the diagnosis and treatment monitoring of various infectious and inflammatory conditions.(96) It was the first acute phase protein to be described and is produced by the liver largely in response to IL-6.(66, 97). In TB, it has been studied most in relation to either treatment response,(94, 98-100) or screening for active TB amongst asymptomatic patients, particularly those attending antiretroviral clinics.(68, 101) In both situations it performs well, however studies of its correlation with bacterial load have been fewer. Brown *et al* found serum CRP levels to be higher in sputum culture positive than negative TB patients, and smear positive than negative,(102) but Mendy *et al* found no association with smear grade.(98) In the Sigal study, CRP correlated with radiographic extent of disease but not with sputum bacterial load, although the authors only published those proteins that were significantly associated after correction of p-values for 70 proteins.(94) Our data therefore adds to the literature in strengthening the evidence for CRP as a marker of bacterial load and severity in TB.

Matrix metalloproteinases (MMPs) have long been known to be implicated in the pathogenesis of TB.(103) There are 24 different mammalian MMPs, of which MMPs-1 and -9 are the ones secreted in highest concentrations in response to inflammatory stimuli.(104) Expression of MMPs-1, -7 and -10 have been shown to be directly inducible by Mtb in human macrophages *in vitro*,(105) however multiple lines of evidence implicate MMP-9 as also being produced and involved in the pathogenesis of TB. It is secreted from neutrophils in Mtb-infected cell cultures,(106) present in the same areas of Mtb-infected mouse lungs as acid-fast

bacilli,(107) and detected in lungs of TB patients.(103) In serum, MMP-9 levels have been shown to correlate with radiographic extent of TB disease, and normalise following effective treatment.(76, 94) To our knowledge, this is the first time that an association between serum MMP-9 levels and sputum bacterial load has been shown in human TB patients. MMP-9 may not be directly induced by Mtb, but our data add to the evidence that MMP-9 production from the disease site may provide a reliable readout of Mtb load and/or disease activity.

TGF-alpha is a member of the epidermal growth factor family and stimulates cell proliferation and development, including tissue repair. TGF-alpha levels have been found to be higher in stimulated Quantiferon (QFT) supernatants from QFT positive individuals with TB than those with LTBI,(108) and, in a pan-African cohort, compared with other respiratory diseases.(109) By contrast, a plasma-based study published alongside found no difference in plasma TGF-alpha levels between individuals with TB and LTBI.(110). This may reflect the fact that this cytokine is produced locally at the disease site and so may not have reliable access to the circulation. They did find, however, that lower levels of TGF-alpha at week 1 of treatment were associated with sputum culture negativity at week 8,(110) which fits with our observation that serum TGF-alpha levels pre-treatment associate with sputum bacterial load. The mechanism of TGF-alpha production in TB and its relevance has not yet been elucidated, however it may be in response to tissue damage. To our knowledge, this is the first time that the relationship between sputum bacterial load and serum TGF-alpha level has been studied in TB.(94)

Considering these three proteins together, it can also be observed that smear grade appears to associate with the host response both at the disease site (i.e. MMP-9 in the lung, +/- TGF-alpha) and systemically (i.e. CRP from the liver). It may therefore be that biomarkers from each of these sites would complement each other in marking severity/bacterial load. In addition, both CRP and MMP-9 were part of our 9-protein signature, hence it may be that serum protein diagnostic signatures of TB have a degree of inbuilt prognostication. In terms of direction of causation, it is likely that TGF-alpha and CRP are driven by bacterial load, however the relationship bewteen load and MMP-9 may be bidirectional, since the MMP-driven cavities favour Mtb replication.(111)

In the HIV uninfected half of the TB patient cohort, the only protein whose levels were associated with smear grade (when considered as <2+ or >/= 2+) was MMP-2, which was significantly lower in smear >/= 2+. MMP-2 is a constitutively expressed enzyme, as opposed to the inducible MMPs such as MMP-1, 3, 8 and 9.(112) In the re-test of the AE-TBC biomarkers in the EU-TB cohort (Chapter III), serum MMP-2 levels were found to be lower in TB than LTBI, but not lower than the other diseases (OD) group. The mechanism by which serum MMP-2 levels decrease in TB or severe illness is not clear, although it may be an effect of TNF-alpha.(73) Previous studies have found MMP-2 levels to be lower in serum in TB than in healthy or sick controls,(74, 85) however this is the first time to our knowledge that an association between serum MMP-2 level and smear grade has been sought.

In HIV co-infected TB patients, significant associations were seen between smear grade and levels of fibrinogen, apo-AI, MMP-9 and TGF-alpha. With regards to apo-AI, however, this was later found to have been confounded by CD4+ count (Section 3 iii), Table 44, above). The association with fibrinogen is new to the literature, and this protein was included in our 9-protein diagnostic signature. Biologically, fibrinogen is part of the acute phase hepatic response, hence the patterns of association between smear grade and serum

proteome appear to be similar in HIV/TB: some associated proteins arise from the lung (MMP-9 and TGF-alpha), and some from the acute phase response.

SAA and IFN-gamma were not signficantly associated with smear grade in our study even though they were in the Sigal study.(94) Before correction for multiple comparisons, however, IFN-gamma did correlate positively in TB, HIV- patients and SAA trended towards this. In TB, HIV+ patients, neither protein was significantly associated before correction, but both trended towards. A summary of the proteins that were associated with baseline sputum load, baseline radiographic correlates of load and response to treatment in our study and the Sigal study is shown below in Table 55. The list from the Sigal study is restricted to those that were also measured in our study.

	EU-TB study	**Sigal study**
Baseline sputum bacterial load (all TB patients)	**CRP** **MMP-9** TGF-alpha	SAA IFN-gamma
- **In TB, HIV-**	(MMP-2)	*n/a*
- **In TB, HIV+**	**MMP-9** (TGF-alpha) (Fibrinogen)	*n/a*
Baseline radiographic extent of disease	*n/a*	**CRP** **MMP-9** IP-10 PCT VEGF
Treatment response	*n/a*	**CRP** SAA

Table 55 A comparison of serum protein biomarkers that were found to associate with measures of baseline sputum bacterial load or associated variables in the EU-TB and a published study.(94)

Bold is used to show proteins that were associated in both studies. Brackets are used to identify proteins that were not measured in the Sigal study. (94) In all cases, higher smear grades and radiographic extent of disease were associated with a greater degree of disturbance to the proteome, i.e. higher levels of all proteins except MMP-2.

iv) Body Mass index

The recognition that TB causes weight loss goes back as far as the literature on the disease itself. In 1694, the English physician Richard Morton wrote about what he perceived to be the various causes and remedies for the syndrome of "consumption", also known by it's greek terminology of "phthisis pulmonalis", the word phthisis meaning "decay".(113) This prominent symptom of weight loss occurs commonly in TB, being consistently reported as having over 80% prevalence,(114) including in HIV co-infection.(115) The pathogenesis of weight loss in TB appears to involve breakdown of both fat and muscle regardless of HIV status and to be cytokine-driven, with higher plasma levels of IL-6 seen in patients with BMI < 18.5.(116, 117)

A link between BMI and and TB severity would therefore seem intuitive and multiple studies have corroborated this, as described in the Results section. Links between severity and/or BMI with serum protein or cytokine levels in TB have also been documented previously, although the interpretations are not straightforward. In a study in Indonesia, Karyadi *et al* found that TB patients with BMI < 18.5 had higher levels of TNF-alpha and IL-6.(117) Range *et al* studied over 500 patients with TB in Tanzania, including patients co-infected with HIV, and found that serum alpha-1-antichymotrypsin levels were inversely correlated with BMI.(118) In a study in Japan, Tsukaguchi *et al* found that in TB patients with mild or moderate "malnutrition", levels of IL-1 and TNF-alpha following stimulation of monocytes with lipopolysaccharide were negatively correlated with BMI. In patients with severe malnutrition, however, these responses were impaired in comparison with those in the mild or moderate malnutrition groups.(119) In keeping with this latter finding, a study of 199 adult TB patients in Malawi found that in patients that died or deteriorated during the first eight weeks of treatment, BMI was lower and *ex vivo* TNF-alpha responses to LPS and H37Rv were impaired.(120)

In our study, when all TB patients were considered together, levels of five proteins correlated weakly but significantly with BMI: procalcitonin, tPA, fibrinogen, TGF-alpha and VEGF. The direction of association was negative in every case, which is in keeping with the fact that levels of all five were higher in TB than LTBI (Chapter III, Section 2). After stratification by HIV status, however, none of the serum protein levels were significantly different between the two BMI categories. This may have been due to the decrease in patient numbers on stratification, although the type of statistical test also differed. In addition, it is possible that a slightly higher BMI cut-off would have captured more of the patients with more deranged protein levels. On inspection of the XY scatter plots of the six proteins that correlated significantly with BMI as a continuous variable (Figure 57, above), there is an impression that a BMI cut-off of 20 might lead to a clearer distinction, since protein levels appear to decrease again below a BMI of approximately 18. It is possible that whilst there are clearly negative correlations between these protein levels and BMI, that this relationship does not hold true below a certain BMI. This would be in keeping with Tsukaguchi *et al*, '91 and Waitt *et al*, '11 and implies that intensive nutritional therapy may be indicated for the most underweight patients. It would also represent a note of caution to using host-derived markers alone for prognostication, and even for diagnosis in underweight patients.

From a biological perspective, the negative correlations between procalcitonin, tPA, fibrinogen, TGF-alpha and VEGF reflect similar processes as have been discussed under smear grade. Procalcitonin and fibrinogen are acute phase reactants and TGF-alpha is likely driven by lung damage. tPA and VEGF were not associated with smear grade in our study, although VEGF was associated with baseline radiographic extent of disease in

the Sigal study.(94) Regarding tPA, it is known to be secreted from the vascular endothelium in response to type I and II interferons, which are both produced in TB (Table 54, above).(121) In HIV/TB, tPA has recently been shown to be a marker of severity, along with various other markers of endothelial activation.(122) The difference in tPA levels by BMI category appeared to be slightly more statistically significant in the TB, HIV+ group, which supports the notion of endothelial activation as part of the pathogenesis of severe HIV/TB. The association between BMI and VEGF likely reflects activity at the disease site, since as was described for MMP-9 above, VEGF has also been found to be expressed in samples of resected lung from TB patients, and to decline in level in the serum in response to treatment.(123)

In summary, we found associations between BMI in TB and serum protein markers of lung inflammation, vascular endothelial activation and the acute phase response. These findings imply that low BMI in TB is largely a result of severe, progressive inflammation, although we did observe patterns in the distribution of protein levels suggesting that below a certain BMI, responses may wane.

v) CD4+ count

Peripheral blood CD4+ T cell count (CD4+ count) is known to be decreased in active TB, even in the absence of HIV infection, primarily due to recruitment of Mtb-specific CD4+ to the lungs.(124-127) In some studies, this process appeared to be particularly pronounced in patients with more severe TB.(125-127) In HIV mono-infection, due to the natural history, CD4+ count was recognised from the early days of the epidemic as being the key biomarker of disease progression and a powerful predictor of risk of AIDS and death.(128, 129) In HIV-associated TB (HIV/TB), as might therefore be expected, CD4+ count has also been shown to associate with outcome, as outlined above.

In our study, levels of five proteins were found to correlate significantly and moderately strongly with CD4+ count as a continuous variable: ferritin, tPA, apo-AI, IP-10 and TNF-alpha. In addition, levels of tPA were associated with CD4+ count as a binary variable. Associations were in the direction expected from the data on levels in TB vs. LTBI (Chapter III, Section 2), implying that the main direction of association between CD4+ count in HIV/TB and the serum proteome is negative, i.e. a lower CD4+ count associates with a greater degree of disturbance. In terms of the implications about the biology of HIV/TB, as was the case with BMI, there were manifestations of the acute phase response (ferritin, apo-AI and TNF-alpha), and of interferon-driven vascular endothelial activation (tPA), but also of Th1-type cytokine activity in the form of IP-10. As discussed above in Section 4 ii), this is interesting, since the pathogenesis of HIV/TB is supposed to largely arise from a Th1 immune defect. Whilst the explanation for the source of IFN-gamma in advanced HIV/TB is production by other cell types, it cannot be maintained that lack of IFN-gamma production in general accounts for the pathogenesis of HIV/TB.

It is interesting to note that, of the five proteins that associated with CD4+ count, none were associated with smear grade. Apo-AI appeared to be initially, but this was later found to have been confounded by CD4+ count (Section 3 iii), Table 44, above). It is well established that, in HIV/TB, lower CD4+ counts predispose to dissemination of infection and mycobacteraemia.(122, 130) It would therefore appear that, in HIV/TB, there are some serum proteins that are sputum bacterial load-associated, and some that are dissemination-associated. This notion is supported by the biology of ferritin in particular, which in addition to being produced by the liver, is also secreted by reticuloendothelial macrophages.(131) It is also in keeping with other reports of high serum ferritin levels being found in patients with systemic opportunistic infection in HIV,(132) and also disseminated HIV/TB and HIV-associated disseminated *M. avium intracellulare* infection.(133, 134) Ferritin may therefore represent a helpful marker of disseminated bacterial load in HIV/TB, although further studies are needed to confirm this.

To our knowledge, associations between levels of serum proteins and CD4+ count in HIV/TB have not been comprehensively studied. Skogmar *et al.* (’15) quantified levels of CRP and neopterin in 195 patients with HIV/TB and found that both were negatively correlated with CD4+ count, and neopterin more strongly (Spearman’s rho -0.53).(135) Lawn *et al.* (’00) measured beta-2-microglobulin in sera from 50 HIV/TB patients and found a weak but significant negative correlation with CD4+ count (rho -0.29).(136) No other proteins were quantified in these studies however, hence our study represents a significant addition to the literature in this regard.

vi) Age

The incidence and mortality rate of TB in the elderly is higher, with the highest number of deaths in HIV uninfected patients occuring in the age group of 55-64 years, and the case fatality ratio rising continually after that.(8, 137) As reviewed by Kollman *et al.* (’12), the effects of ageing on the immune system in general are well studied and include a heightened baseline level of circulating pro-inflammatory cytokines but reduced responsiveness to stimulation of pathogen recognition receptors (especially TLR-4), as well as impaired macrophage and neutrophil function.(138) The effects regarding TB specifically are less well documented, however as reviewed by Rajagopalan *et al.* (’00), there is a preponderance (in HIV uninfected patients) towards re-activation rather than primary disease, but an elevated risk of dissemination.

To our knowledge, associations between serum protein levels and host age in adults with TB have not previously been shown. Brown *et al.* (’16) looked for associations between CRP level and various host features and found that CRP levels were lower in those aged < 16 years (i.e. children), but there was no significant difference between their other age categories (16 – 35, 35 – 50 and >/= 50 years).(102) There was a trend towards levels being lower in those over 50, however, which is in keeping with our data (rho -0.22 for CRP, p = 0.02 before correction).

For the one protein which correlated with age in our study (haptoglobin) the direction of correlation was in keeping with the hypothesis that responses to Mtb wane with advancing age. Haptoglobin is an acute phase reactant and binds any free haemoglobin that may be circulating following haemolysis.(139) It is one of the proteins visually detectable by classical serum protein electrophoresis techniques, hence early studies of the serum proteome in TB found it to be elevated in comparison with healthy controls.(71) It also appears to play an important role in the host defence against Mtb, as evidenced by the fact that the Hp 2-2 polymorphism is associated with an increased mortality rate from TB.(70) The Hp 2-2 phenotype is a weaker anti-oxidant and has lower affinity for haemoglobin than the other phenotypes, such that individuals with Hp 2-2 have higher iron concentrations in macrophages and monocytes.(140) It may be therefore that some of the increased mortality from TB in elderly patients may be attributable to impaired haptoglobin production, although there are likely to be many factors.

Overall, there was a trend towards multiple protein reponses being negatively associated with age in our study: CRP, MMP-9, apo-AI, TGF-alpha, IL-1ra and TNF-alpha were all significantly associated before p-value correction. Correlations were weak in every case, but many of these proteins were biomarkers of both diagnosis and severity in our cohort. Caution may therefore be required when interpreting levels of host serum proteins in older age groups with TB.

vii) Gender

In our study, the only protein whose levels were significantly associated with gender was ferritin, which was higher in male patients. After further analysis, however, this appeared to be confounded by a lower number of male TB, HIV+ patients being on ART than female patients, accompanied by a lower CD4+ count in men. These latter associations were not statistically significant, however the distribution of ferritin level by CD4+ count had appeared exponential, which much higher ferritin levels below a CD4+ count of around 100 cells/mm^3 (Section 3 iii), Figure 57, above). It is well known that male patients with HIV/TB in Africa present later than females, however, and also have worse outcomes, including death.(141, 142) Given the biology of ferritin discussed already, it would seem likely that ferritin levels were significantly higher in men due to higher rates of dissemination among the HIV co-infected half of the cohort.

5. Conclusion

Through quantification of levels of the 22 proteins selected for the AE-TBC study in a large sample of the EU-TB cohort,(85) we have shown that levels of multiple serum proteins are altered in TB, ranging from proteins involved in the acute phase response (APR), through those involved in T cell responses, matrix degradation, response to hypoxia and lipid metabolism. According to our dataset, the serum proteome therefore reflects multiple key aspects of the response to TB. In addition, we have been able to capture aspects of the impact of HIV co-infection on the serum proteome in TB, which is that the acute phase response is largely preserved, but that protein responses arising from the disease site are impaired (MMP-9 and complement factor H). In addition, the phenomenon of immune activation in HIV/TB also appears to manifest in the serum proteome, in the form of higher levels of IP-10 and TNF-alpha, and a more noticeable rise in IL-1ra levels compared to LTBI. It is in fact noteworthy that despite the strong association between CD4+ count and TB risk and severity in HIV, that circulating levels of IFN-gamma were not lower in this group.

Whilst not specifically designed for this purpose, data on multiple different demographic and clinical variables was collected as part of the EU-TB study. This allowed us to perform a robust search for associations between protein levels and clinical variables in TB. A number of interesting observations were made in the form of associations between smear grade, BMI, age, and in the case of HIV/TB, CD4+ count. Of particular interest were the facts that a) CRP and MMP-9 appear to reliably associate with bacterial load in the lung, both in our and other studies;(94) b) low BMI in TB appears primarily to result from more severe immune response, although below a certain threshold of BMI, host responses may become impaired; c) in HIV/TB, CD4+ found associates with a different collection of proteins that does sputum smear grade, implying that specific serum proteins could be used to both stage disease and monitor response; d) some serum protein responses may wane in advancing age.

It is not possible to crystallize the precise implications and applications of all these findings, however our conclusion overall is that many different aspects of the biology of TB are reflected in the host serum proteome. Many of the possible biomarkers of severity that we have elucidated are present within diagnostic signatures of TB, suggesting a degree of inbuilt prognostication.

CHAPTER VII GENERAL DISCUSSION

The conclusions of the activities presented in this thesis will now be given under their respective headings. Findings will be considered in the context of the existing literature on host-derived biomarkers and biosignatures of TB, specifically those derived from the blood transcriptome and serum metabolome as the two main alternative resources potentially amenable to analysis by a POCT using blood.

1. Host serum protein diagnostic biomarkers of TB

New diagnostic tests for TB that are rapid, accurate, affordable, user-friendly and robust in endemic settings are needed in order to facilitate early treatment to limit morbidity, mortality and transmission.(1) If used as a screening test, such an intervention could also have the added benefits of reducing the number of unnecessary TB cultures and/or PCR-based tests that are performed, the number of unnecessary courses of TB treatment that are given, and the number of alternative and sometimes life-threatening alternative diagnoses that are missed. In this thesis, we have explored the utility of host serum proteins as biomarkers for this purpose in cohorts reflective of patients presenting with TB-like illness in sub-Saharan Africa, both as individual markers and biosignatures. The biobank we utilised was from adult patients that had presented largely to secondary care facilities. By virtue of the study design, 50% of the patients were HIV co-infected. The downstream application that was envisaged was that selected proteins would be rapidly quantifiable at point of care (POC) by multiplexed lateral flow immunoassay (LFA) in a manner that is affordable, accurate, user-friendly and robust in resource-poor settings. The rationales for selecting host serum proteins were that they are abundant and diverse, arise from a wide range of biological sources, are dynamic in level in response to pathological processes and are amenable to immunocapture.

We evaluated two different approaches to development of host serum protein-based diagnostic signatures. The first was to test the utility of proteins that had been developed in other, primary care-based cohorts (Chapter I).(2) From this, we discovered a 9-protein signature that had an accuracy in our test set that showed potential for use as a screening test in African hospital settings. Among the proteins was complement factor H, which was a new biomarker to the TB field in the forerunner to this study.(2) With an individual area under the receiver operating curve of 70%, complement factor H was the top-performing biomarker out of the 22 tested in this particular study, and performed equally well across the two study sites and HIV status. The second main activity of this thesis and approach that we took to serum protein biomarker development was to test the translatability of proteins that had been discovered by a mass spectrometry-based survey of serum proteomes in our cohort by immunocapture in the form of ELISA (Chapter III). The results from this were enlightening in that, of the three proteins tested, one translated fully (transthyretin), but the other two illustrated specific challenges. Firstly, beta-2-microglobulin (beta-2-M) levels by ELISA correlated weakly with their respective SELDI intensities, and their ability to distinguish TB from latent TB (LTBI) and other diseases was poorer than expected. A reason for this was felt to be that, in HIV/TB disease in particular, the ELISA platform may be detecting another form of beta-2-microglobulin in addition to that which was detected by SELDI. Investigation into this is ongoing, and it may be that by optimising the accuracy of the immunocapture platform, that the performance of beta-2-M can be restored. Secondly, detection of alpha-1-antichymotrypsin (ACT) by ELISA was affected by the fact that, in TB, it appears that both a fragment and the intact protein are elevated in serum. In our other diseases group, however, the intact protein was also elevated, and our commercial

polyclonal ELISA kit detected both of these targets. As with beta-2-M, further work is underway to develop an immunocapture probe that is specific to the ACT fragment of interest, yet can be incorporated into a LFA platform, such as a monoclonal antibody.

The molecular weight of complement factor H is in excess of the accurate range of SELDI. Conclusions that we drew from these efforts to develop serum protein-based signatures were a) host serum protein biomarkers perform differently at different points in the healthcare system, hence studies should be based in the populations for which the diagnostic test is intended; b) untargeted studies are the optimal approach, although complexities in translation to immunocapture are expected; c) targeted approaches to biomarker discovery may complement untargeted ones, and a broad survey of other studies in the literature is warranted.

In comparison to other "omics"-based signatures for the diagnosis of TB, with a ROC AUC in the training set of 0.90, our 9-protein test compares relatively well. With regards to serum proteomics, as mentioned in the Introduction Chapter and also the Discussion section of Chapter 3, no prior study has evaluated the accuracy of a diagnostic signature for TB using a) a clinically relevant recruitment strategy; b) patients co-infected with HIV; c) immunocapture; d) secondary care as the major focus. Since commencement of this thesis, an aptamer-based study of patients with suspected TB recruited from multiple countries, including co-infection with HIV, has been published, having found a six-protein signature with a ROC AUC of 0.92.(3) All patients were recruited from primary care settings, however, thus representing a distinct type of population from that of the EU-TB study, and more akin to that of the AE-TBC study. No other comparable studies have since been published at the time of writing.

The blood transcriptomic arm of the EU-TB study, based on the same cohorts as referred to in this thesis, discovered a signature to distinguish TB from OD which achieved a ROC AUC of 0.95 (95% CI: 0.89 – 0.99), however this was achieved utilising 44 transcripts. Efforts to develop POC testing platforms for translation of transcriptomic signatures have done so with very few numbers of transcripts initially, e.g. four.(4) Since this, a study by Maertzdorf *et al* was published which found that, using a targeted approach in an HIV uninfected population from India, TB patients could be accurately distinguished from healthy controls with a four-gene signature (ROC AUC 0.98).(5) When applied to the discrimination of TB from other diseases using published datasets, however, including HIV, this dropped to 0.72.(5) More successfully, seemingly, was a study by Roe *et al* in which a different four-transcript signature was derived that distinguished patients with active TB from those presenting to hospital with non-TB febrile illnesses with a ROC AUC of 0.94.(6) This is a promising finding, however the OD (or "fever") group in this case was not recruited from a point of differential diagnosis, and were also exclusively HIV uninfected.(6) The accuracy of this signature would therefore be expected to fall when re-tested in patients that were genuinely clinically suspected of having TB, and when applied to HIV infected patients. Similarly, a 3-transcript signature was published by Laux da Costa *et al* that distinguished TB from OD with 95% sensitivity and 93% specificity, however these OD cases were not clinically suspected of having TB either.(7) In addition, the numbers in this study were small (n = 27 TB and 23 OD) which produced wide confidence intervals.(7) Indeed, when this signature (along with 15 others) was re-tested on a collection of 24 publicly available transcriptomic datasets, the sensitivity and specificity dropped to 83% and 65% respectively.(8) The best performing signature in this latter study was a 3-transcript signature referred to as "Sweeney3", which produced 89% sensitivity and 74% specificity. The control group in this instance, however,

comprised more healthy control and LTBI patients that it did ODs, which would likely have inflated the diagnostic accuracy (n = 573 healthy controls, 699 LTBIs, 803 ODs).(8) As discussed further below, Sweeney3 was also evaluated in three further cohorts, and found to have an overall sensitivity and specificity of 90% and 70%, placing it within WHO targets for a screening test.(9) All of these cohorts were HIV uninfected, however, and were also either primary care or community based, and thus a different situation from that of our study.(9) A fresh approach taken recently to transcriptomic signature discovery was published by Singhania *et al,* in which the TB blood transcriptome was divided into "modules", each of which representing a different cluster of co-ordinately expressed genes, and the best individual transcripts then selected from within each module. The aim of this approach was to avoid false classification of viral infections such as influenza as TB, since both predominantly feature interferon-driven responses.(10) The study accessed publicly available transcriptomic data from studies including the EU-TB transcriptomic study. A minimal 20-gene signature was then derived from comparisons of TB transcriptomes with those of LTBI. Whilst this performed very well in distinguishing TB from LTBI, however, it did not so well in distinguishing TB from OD (ROC AUC 0.74 in the EU-TB dataset). This may be because influenza did not make up an appreciable proportion of the ODs that were recruited to the EU-TB study. In summary, there is currently no blood transcriptomic signature that has been shown to distinguish active TB from OD in patients presenting to hospital, regardless of HIV status, with an accuracy greater than that of our 9-protein signature, other than the 44-transcript EU-TB signature. The Sweeney 3-gene signature shows promise for a community-based transcriptomic triage test for TB.

Development of a serum metabolomic signature for TB is lagging behind considerably. In 2012, a 20-metabolite signature was discovered that distinguished TB from LTBI and healthy controls with a cross-validation error rate of 3.3% (n = 136).(11) To date, only two studies have found signatures to discriminate TB from other diseases: one published by Feng *et al*, '15, in which a four-metabolite signature was discovered with a ROC AUC of 0.99 in a cohort of 371 adults in China.(12) As with many of the transcriptomic studies, however, the control group comprised both healthy controls and other diseases (n = 105 and 146, respectively), and it is clear from the levels of metabolites in each group that the differences in levels were less marked between TB and OD than they were between TB and healthy controls. Published at the same time by Lau *et al*, another study found that a two-metabolite combinations had specificities of > 90%, but sensitivities as low as 70% to distinguish TB from community acquired pneumonia, using the optimum cutoffs for each metabolite.(13) The high specificities may be interesting from a biological point of view, but the low sensitivities preclude usage as a diagnostic.

Interestingly, there has been very little overlap in terms of biomarker identities between blood transcriptomic and serum proteomic studies. The closest to a common biomarker across these two categories with regard to our study was haptoglobin, which was part of the four-transcript signature published by Roe *et al.*(6) Haptoglobin was a individual biomarker of TB vs OD in the EU-TB cohort in our study (ROC AUC 0.69), however was not included in the final 9-protein signature. This lack of commonality is not entirely surprising, since the blood transcriptome is a compartment specific to circulating leukocytes, whereas the serum proteome contains products of multiple organ systems, including leukocytes. In addition, the serum proteome is largely restricted to secreted, rather than intracellular proteins. For these reasons, it may in fact be pertinent to amalgamate transcriptomic and proteomic data in the hopes of finding an even better signature. The resulting development of POCTs that can detect both types of biomarker would likely be very challenging, however.

2. Response of host serum proteins to treatment.

In addition to assessing diagnostic performance, we also utilised plasma from a paediatric cohort of TB patients that were recruited from one of the same populations as our adult cohort (Cape Town, South Africa). This was originally planned in order to evaluate whether the high levels of complement factor H in our adult TB cohort were due to genetic polymorphism, or whether they represented a host response. In so doing, however, we captured data on levels of five serum proteins before and after treatment, as well as in healthy controls. Contrary to our hypothesis, levels of factor H returned to normal after treatment in the majority of patients. In addition, levels of transthyretin and apo-AI (which had also been markers of TB in adults) also returned to or towards normal. This study was limited by lack of definite outcome data, but it was interesting that there was very little overlap in terms of which proteins failed to normalise in which patients. This implies that, as with diagnostic testing, a combination or signature of proteins may be required for treatment monitoring purposes as well. As mentioned in the Discussion of Chapter IV, there are currently no published serum protein signatures of TB treatment response that have been derived in association with final clinical outcomes. In keeping with our finding for transthyretin, however, a proteomic study comparing levels of proteins at start and end of intensive phase of treatment found that levels dropped significantly.(14)

Studies of blood transcriptomics have been more successful to date, with the earliest studies finding that signatures resolved with treatment.(15,16) In 2017, Thompson *et al* published a novel 5-transcript signature that could predict treatment failure with a ROC AUC of >/= 0.8 at any of weeks 0, 1, 4 and 24 of treatment.(17) This was in an HIV uninfected cohort. They did not have follow-up data to look for correlates of relapse-free cure, but nevertheless this was a promising start. Later, as mentioned above, Warsinske *et al* evaluated the performance of their 3-transcript diagnostic signature (Sweeney3) at predicting treatment response, again in HIV uninfected patients.(9) They found that at end of treatment, Sweeney3 could distinguish those whose concurrent sputum cultures were negative or positive (i.e. treatment failure; ROC AUC 0.93); and at the start of treatment, it could predict those who would have residual inflammation on PET/CT at end of treatment, of which it is known that some would then go on to relapse.(9) Such patients had more pronounced changes in gene transcription at the outset. The authors did not, curiously, evaluate the performance of Sweeney3 to predict outcome when utilised during treatment. Most recently, a study by Darboe *et al* of HIV co-infected TB patients found that a previously discovered diagnostic signature decayed progressively during treatment, but that there was no difference in rate of decay between those who later relapsed and those who did not.(18) They did not seek to discover a novel signature purely for this purpose, however.

To date, there have been no serum metabolomic signatures of TB treatment response published. In urine, a study by Mahapatra *et al* found a six-metabolite signature that could distinguish samples at month 0 from month 2 with an error rate of 11.8%. The patients were all successfully cured, however, so the clinical relevance of this finding is as yet unclear.(19)

In summary, the proteomic field lags behind the transcriptomic field in delivering signatures of treatment response at present. This may change, however, as what is seen across all these three biomarker categories is that the disturbance to homeostasis that results from the active TB state is reversed by successful treatment, more completely in those that are cured. The findings in our small serum protein study add to the proteomic

treatment response biomarker literature, especially for paediatric TB. As with diagnostic signature studies, validation of any of these types of signatures in other cohorts, including studies of clinical benefit in the real world, will be critical in taking such signatures forward.

3. Host serum proteins and the biology of TB.

The adult EU-TB cohort was a highly characterised cohort of patients that included those with TB that was confirmed by culture, and those with LTBI that were well. In addition, a) 50% of every clinical group were HIV-infected, and b) data was collected on multiple different demographic and clinical variables. Having quantified serum protein levels in large samples of these cohorts, we therefore seized the opportunity to study the data with the aim of discovering new information about the biology of TB.

As presented in Chapter VI, we discovered that the serum proteome in TB contains information reflective of a) the pathogenesis of TB; b) the impact of co-infection with HIV; c) the bacterial load in sputum; d) the patient's body mass index; e) the CD4+ count in HIV/TB; and f) the age of the patient. Of particular interest was that MMP-9 and CRP were validated as reliable biomarkers of bacterial load in sputum, and that, in HIV/TB, different serum proteins associated with each of sputum load and CD4+ count. Ferritin was identified as a candidate biomarker of disseminated bacterial load, and the dogma that progression of HIV/TB relies on loss of Th1 response was challenged by higher levels of IFN-gamma and IP-10 in such patients. These findings will now be discussed in the light of equivalent findings from the transcriptomic and metabolomic literature.

With regards to the pathogenesis of TB, on the serum proteomic side, an interesting study of TB risk recently published by Penn-Nicholson *et al* compared proteomes of adolescents with LTBI who went on to develop active TB with those that didn't over a 2-year study period.(20) Out of 3,040 proteins measured by aptamer, 135 were found to be differentially abundant in those that progressed to get TB. At the top of the list were MMP-1 and CRP, similar to our study, but also Galactose-1-phosphate uridyl transferase 1 (GALT-1), which is involved in galactose metabolism and was more abundant in progressors' serum, and Creatine Kinase type M/type B (CK-MB), which was less abundant. These were not measured in our study and represented new findings to the literature. With regard to transcriptomics, as reviewed by Cliff *et al,* the blood transcriptome does correlate with activity at the disease site, and it is clear that myeloid cell gene transcription is upregulated.(21) This is in keeping with our finding of a marked acute phase response, since activation of macrophage and neutrophils results in production of IL-6. The Th1 response is also represented in both, since we detected higher circulating levels of IFN-gamma and IP-10, and IFN-gamma responsive genes are found to upregulated in the blood. A major finding from blood transcriptomics that wasn't replicated in our study, however, is type I interferon signalling. An early study by Berry *et al* found interferon signalling to be the most over-represented pathway in TB in comparison to healthy controls.(14) This included both types I and II interferon, however, as in our dataset, they did not find IFN-alpha proteins to be elevated in serum. This may relate to limits of detection, however the finding that type I interferon signalling is pronounced in TB was hitherto unappreciated in the field. Increased interferon signalling was also the predominant feature of the transcriptomes in adolescent patients with LTBI that progressed to active TB in the transcriptomic arm of the recent longitudinal study of TB risk mentioned above.(22) Similarly, in a meta-analysis of blood transcriptomic data comparing active TB with healthy controls, the inflammasome was also enriched for, whereas we did not

see significantly higher levels of IL-1ra in serum.(23) So there are some aspects of TB pathogenesis that are captured in the blood transcriptome but not the serum proteome. Conversely, there were some findings in our serum proteomic data which have not been found in transcriptomic data: alpha-1 antichymotrypsin is secreted by liver and lung tissue, and would therefore not be expected to be reflected in leukocyte transcription, however this was a highly significant biomarker of TB in the EU-TB cohort with an individual ROC AUC of 0.84 to distinguish TB from LTBI and OD combined.

With regard to metabolomics, the study by Weiner *et al* mentioned above found markers of "inflammation, immunosuppression and stress" to characterise the TB state.(10) As with the comparison with transcriptomics, there is some overlap with our proteomic findings, but this is incomplete. Specifically, levels of amino acids were found to be lower in TB sera, which fits with our finding that transthyretin levels are lower in serum and that they fail to completely normalise after treatment (in children). These findings together imply a relative protein deficient state in TB. In addition, they found multiple correlations between amino acid levels and cytokine levels, implying that this deficiency hinders the immune response to TB. This would fit with our observation that levels of several serum acute phase proteins, cytokines and growth factors appeared lower in patients with BMIs below 18 (Figure 56, Chapter VI). What was not reflected in our proteomic data was that levels of lipophosphatidylcholines (LPCs) were lower in TB than healthy controls. This has also been seen in patients with cancer or sepsis, but a mechanism was proposed by the authors for TB in that Mtb can inhibit phospholipase-A2 enzyme, a major source of LPC, in the macrophage in order to induce its apoptosis.(24) Such a process would not be expected to be reflected in the serum proteome.

With regards to HIV-associated tuberculosis, the three observations we made were that the acute phase appears to be preserved; activity at the disease site appears to be impaired; and that the Th1 response appears to be preserved, but that the likely results from immune activation and production of cytokines by other cell types. Interestingly, these observations match what has been found to date in the blood transcriptome of HIV/TB relatively well. Firstly, in support of the point about the acute phase response, a pair of studies by Gebremicael *et al* comparing expression levels of 105 immune-related genes in blood between TB patients and healthy controls both with and without HIV found an over-representation of pattern recognition receptors in the HIV infected cohort (i.e. TB, HIV+ vs. LTBI, HIV+), including TLR-2 and -4, which wasn't seen in the HIV uninfected cohort.(25) Increased expression of these receptors would trigger the innate and acute phase responses in HIV/TB. Secondly, with regard to "impaired activity at the disease site", in the transcriptomic arm of the EU-TB study, whilst there was no statistically significant difference in the performance of the 27-transcript signature to distinguish TB from LTBI between the HIV infected and uninfected halves of the cohort, there was a trend towards poorer discrimination in the HIV infected half.(26) This pattern was also seen when another signature (by Berry *et al*) was applied to the cohort. A comparison of individual transcripts between the TB, HIV- and TB, HIV+ groups was not done in this study, however this reduced discriminative ability of the signature implies a less severely deranged transcriptome in HIV/TB. With regard to MMPs specifically, whilst serum MMP protein levels were lower in our data in TB, HIV+ vs TB, HIV-, a comparison of MMP transcript abundance in blood between these groups has never been done. Tadokera *et al* found, however, that MMP transcript abundance in PBMCs post-stimulation were higher in HIV/TB patients with IRIS vs. those without IRIS, implying that MMP transcription in the absence of IRIS and before ART initiation was reduced.(27) It seems likely, therefore, that MMP transcripts in unstimulated whole blood would be less numerous in TB, HIV+

than TB, HIV-. Thirdly, with regard to “preserved Th1 responses”, a recent *post hoc* study of the EU-TB blood transcriptome dataset in combination with three other datasets (performed to find a signature of HIV/TB) revealed dysregulation of CD8+ and NK cell genes as a hallmark.(28) These phenomena would not be detectable in the serum proteome directly, however, as mentioned in Chapter VI Section 4 part ii), these are the main cell types other than CD4+ cells that are able to produce IFN-gamma.

Another important study of the blood transcriptome in HIV/TB was a smaller, longitudinal study of HIV infected adults with LTBI. Heathy patients with LTBI had been previously categorised into subclinical TB or truly “latent” TB on the basis of positron emission computed tomography (PET-CT) and either subsequent development of active TB or radiological response to isoniazid preventive therapy.(29) Comparison of the blood transcriptomes in each of these two groups and with an active TB group revealed that, relative to the LTBI group, the transcripts that were overabundant in the subclinical and active disease groups were those relating to the classical complement pathway. These are transcripts whose resultant proteins should be detected in the serum, if looked for, however no correlations between transcript and protein abundances were seen. Levels of circulating immune complexes in the serum in this study did correlate with complement transcript abundance, however, so the lack of correlation with complement protein levels themselves may have again reflected an abundance and limit of detection issue.(29) Alternatively, this may have related to rapid consumption of complement proteins in the circulation, which could represent another way in which information in the blood transcriptome is not reflected in the serum proteome.

There have been no serum metabolomic studies of HIV/TB to date, except for a small study comparing profiles in patients that did and did not develop TB-IRIS following commencement of ART.(30) This found that arachidonic and glycerophospholipid metabolism were altered in the group that went on get IRIS.

With regard to the blood transcriptome and smear grade/sputum bacterial load, a recent study in Haiti identified a 20-gene signature that associated with high bacterial load as measured by Xpert (Dupnik *et al*, ’18). None of these related to serum proteins that were measured in our study, with the top three being annexin A3, a calcium channel subunit, and complement C3b/C4b receptor 1.(31) No other transcriptomic studies have looked for association with sputum bacterial load, to our knowledge, however it was noted from early studies that signatures correlated with radiographic extent of disease,(14) and this has continued to be seen even in a recent three-transcript signature.(8)

With regard to the blood transcriptome and BMI, to our knowledge no studies have specifically evaluated this, however given associations with radiographic severity, it would be expected that lower BMI would associate with greater disturbances to the transcriptome.

With regard to the blood transcriptome and CD4+ count in HIV/TB, this has not been extensively studied, however a case-control study of patients with low CD4+ counts < 100 cells/uL with and without active TB found that published signatures did discriminate well, and that a *de novo* signature showed significant overlap with published signatures in terms of transcript identities also.(32) It may be that the blood transcriptome is not significantly affected by CD4+ count. This study did not include patients with other diseases, however.

Associations between the blood transcriptome and patient age in TB do not appear to have been sought. This will be an important demographic feature for inclusion in future validation studies of transcriptomic (and other) signatures.

4. Closing remarks

In summary, our results in this thesis support the hypothesis that levels of host serum proteins can form accurate diagnostic biosignatures of TB for use in hospital settings by immunocapture. In comparison with other signatures in the literature derived from host biomarkers (proteomic, transcriptomic or metabolomic), our 9-protein signature is among the best for distinguishing TB from other diseases in hospital settings in Africa, regardless of HIV status.

In addition, we confirmed the second part of our hypothesis by finding that levels of serum proteins reflect the biology of TB. Other compartments such as the blood transcriptome and serum metabolome show agreement with some of these findings, particularly with regards to HIV-associated TB (HIV/TB). However, the degree of overlap between what these different facets of the host response can teach us is far from complete.

The host serum proteome is therefore confirmed as a rich resource for translational research on TB. Future avenues leading on from the work of this thesis include refinement of a multiplexed immunocapture-based process for TB diagnosis in similar populations, together with development of a field-friendly platform, and clinical validation studies at points of care. In addition, further clinical studies are required to explore the utility of serum proteins as markers of bacterial load, severity and treatment response, with prospective follow-up to determine relapse-free cure. Finally, further characterisation of the HIV-induced immune defect in HIV/TB is needed to inform intervention in this disease. Serum proteomic studies of progression to active TB, and also of treatment failure, are warranted in HIV infected populations.

REFERENCES

Chapter I

1. World Health Organization. *TB: a global emergency, WHO report on the TB epidemic.* WHO/TB/94.177. 1994.
2. Toungoussova OS, Bjune G, Caugant DA. Epidemic of tuberculosis in the former Soviet Union: social and biological reasons. *Tuberculosis (Edinb)*. 2006;86(1):1-10.
3. Stop TB Partnership. *Our History.* Available from: http://www.stoptb.org/about/history.asp [Accessed 26th September 2019].
4. World Health Organization. *Global tuberculosis report 2018.* WHO/CDS/TB/2018.20. 2018.
5. Dye C, Glaziou P, Floyd K, Raviglione M. Prospects for tuberculosis elimination. *Annu Rev Public Health*. 2013;34:271-86.
6. World Health Organization. *WHO End TB Strategy.* Available from: https://www.who.int/tb/post2015_strategy/en/ [Accessed 26th September 2019].
7. The World Bank. *Gross Domestic Product*. Available from: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD [Accessed 26th September 2019].
8. The World Bank. *Atlas of Sustainable Development Goals*. Available from: http://datatopics.worldbank.org/sdgatlas/ [Accessed 26th September 2019].
9. Peeling RW, Holmes KK, Mabey D, Ronald A. Rapid tests for sexually transmitted infections (STIs): the way forward. *Sex Transm Infect*. 2006;82 Suppl 5:v1-6.
10. Rageade F, Picot N, Blanc-Michaud A, Chatellier S, Mirande C, Fortin E, et al. Performance of solid and liquid culture media for the detection of *Mycobacterium tuberculosis* in clinical materials: meta-analysis of recent studies. *Eur J Clin Microbiol Infect Dis*. 2014;33(6):867-70.
11. van Zyl-Smit RN, Binder A, Meldau R, Mishra H, Semple PL, Theron G, et al. Comparison of quantitative techniques including Xpert MTB/RIF to evaluate mycobacterial burden. *PLoS One*. 2011;6(12):e28815.
12. Hanrahan CF, Theron G, Bassett J, Dheda K, Scott L, Stevens W, et al. Xpert MTB/RIF as a measure of sputum bacillary burden. Variation by HIV status and immunosuppression. *Am J Respir Crit Care Med*. 2014;189(11):1426-34.
13. Olaru ID, Albert H, Zallet J, Werner UE, Ahmed N, Rieder HL, et al. Impact of quality improvement in tuberculosis laboratories in low- and lower-middle-income countries: a systematic review. *Int J Tuberc Lung Dis*. 2018;22(3):309-20.
14. Steingart KR, Henry M, Ng V, Hopewell PC, Ramsay A, Cunningham J, et al. Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *Lancet Infect Dis*. 2006;6(9):570-81.
15. Squire SB, Belaye AK, Kashoti A, Salaniponi FM, Mundy CJ, Theobald S, et al. 'Lost' smear-positive pulmonary tuberculosis cases: where are they and why did we lose them? *Int J Tuberc Lung Dis*. 2005;9(1):25-31.
16. Hawken MP, Muhindi DW, Chakaya JM, Bhatt SM, Ng'ang'a LW, Porter JD. Under-diagnosis of smear-positive pulmonary tuberculosis in Nairobi, Kenya. *Int J Tuberc Lung Dis*. 2001;5(4):360-3.
17. Addo KK, Owusu-Darko K, Dan-Dzide M, Yeboah-Manu D, Ablordey A, Caulley P, et al. Situation analysis of TB microscopy centres in Ghana. *Int J Tuberc Lung Dis*. 2006;10(8):870-5.

18. Boehme CC, Nicol MP, Nabeta P, Michael JS, Gotuzzo E, Tahirli R, et al. Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. *Lancet*. 2011;377(9776):1495-505.

19. World Health Organization. *WHO Tuberculosis Programme: framework for effective tuberculosis control*. WHO/TB/94.179. 1994.

20. World Health Organization. *Chest Radiography in Tuberculosis Detection: Summary of current WHO recommendations and guidance on programmatic approaches.* WHO/HTM/TB/2016.20. 2016

21. van Cleeff MR, Kivihya-Ndugga LE, Meme H, Odhiambo JA, Klatser PR. The role and performance of chest X-ray for the diagnosis of tuberculosis: a cost-effectiveness analysis in Nairobi, Kenya. *BMC Infect Dis*. 2005;5:111.

22. World Health Organization. *TB/HIV: A clinical manual*. ISBN: 9241546344. 2004.

23. World Health Organization. *WHO monitoring of Xpert MTB/RIF roll-out.* Available from: https://www.who.int/tb/areas-of-work/laboratory/mtb-rif-rollout/en/ [Accessed 26th September 2019].

24. Cepheid. *Xpert MTB/RIF.* Available from: https://www.cepheid.com/us/cepheid-solutions/clinical-ivd-tests/critical-infectious-diseases/xpert-mtb-rif [Accessed 26th September 2019].

25. Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. Xpert(R) MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst Rev*. 2014(1):CD009593.

26. World Health Organization. *Automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB/RIF assay for the diagnosis of pulmonary and extrapulmonary TB in adults and children: policy update*. WHO/HTM/TB/2013.16. 2013.

27. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A*. 1997;94(18):9869-74.

28. World Health Organization. *Rapid Implementation of the Xpert MTB/RIF diagnostic test: Technical and Operational 'How-to' Practical considerations.* WHO/HTM/TB/2011.2. 2011.

29. Dorman SE, Schumacher SG, Alland D, Nabeta P, Armstrong DT, King B, et al. Xpert MTB/RIF Ultra for detection of *Mycobacterium tuberculosis* and rifampicin resistance: a prospective multicentre diagnostic accuracy study. *Lancet Infect Dis*. 2018;18(1):76-84.

30. Stop TB Partnership. *TB REACH Xpert Budget Estimation Tool*. Available from: http://www.stoptb.org/global/awards/tbreach/bet.asp [Accessed 26th September 2019].

31. World Health Organization. *Prerequisites to country implementation of Xpert MTB/RIF and key action points at country level: checklist.* WHO/HTM/TB/2011.12. 2011.

32. The World Bank. *Current health expenditure per capita.* Available from: https://data.worldbank.org/indicator/SH.XPD.CHEX.PP.CD [Accessed 27th September 2019].

33. World Health Organization. *TB finance: global financial monitoring.* 2019 Available from: https://www.who.int/tb/areas-of-work/monitoring-evaluation/financing/global-financing/en/ [Accessed 26th September 2019].

34. Theron G, Zijenah L, Chanda D, Clowes P, Rachow A, Lesosky M, et al. Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. *Lancet*. 2014;383(9915):424-35.

35. Churchyard GJ, Stevens WS, Mametja LD, McCarthy KM, Chihota V, Nicol MP, et al. Xpert MTB/RIF versus sputum microscopy as the initial diagnostic test for tuberculosis: a cluster-randomised trial embedded in South African roll-out of Xpert MTB/RIF. *Lancet Glob Health*. 2015;3(8):e450-e7.
36. Vassall A, Siapka M, Foster N, Cunnama L, Ramma L, Fielding K, et al. Cost-effectiveness of Xpert MTB/RIF for tuberculosis diagnosis in South Africa: a real-world cost analysis and economic evaluation. *Lancet Glob Health*. 2017;5(7):e710-e9.
37. Abdurrahman ST, Emenyonu N, Obasanya OJ, Lawson L, Dacombe R, Muhammad M, et al. The hidden costs of installing Xpert machines in a tuberculosis high-burden country: experiences from Nigeria. *Pan Afr Med J*. 2014;18:277.
38. Clouse K, Page-Shipp L, Dansey H, Moatlhodi B, Scott L, Bassett J, et al. Implementation of Xpert MTB/RIF for routine point-of-care diagnosis of tuberculosis at the primary care level. *S Afr Med J*. 2012;102(10):805-7.
39. Walzl G, McNerney R, du Plessis N, Bates M, McHugh TD, Chegou NN, et al. Tuberculosis: advances and challenges in development of new diagnostics and biomarkers. *Lancet Infect Dis*. 2018;18(7):e199-e210.
40. Notomi T, Okayama H, Masubuchi H, Yonekawa T, Watanabe K, Amino N, et al. Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res*. 2000;28(12):E63.
41. Bojang AL, Mendy FS, Tientcheu LD, Otu J, Antonio M, Kampmann B, et al. Comparison of TB-LAMP, GeneXpert MTB/RIF and culture for diagnosis of pulmonary tuberculosis in The Gambia. *J Infect*. 2016;72(3):332-7.
42. Gray CM, Katamba A, Narang P, Giraldo J, Zamudio C, Joloba M, et al. Feasibility and Operational Performance of Tuberculosis Detection by Loop-Mediated Isothermal Amplification Platform in Decentralized Settings: Results from a Multicenter Study. *J Clin Microbiol*. 2016;54(8):1984-91.
43. Shah M, Hanrahan C, Wang ZY, Dendukuri N, Lawn SD, Denkinger CM, et al. Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in HIV-positive adults. *Cochrane Database Syst Rev*. 2016(5):CD011420.
44. World Health Organization. *The use of lateral flow urine lipoarabinomannan assay for the diagnosis and screening of active tuberculosis in people living with HIV*. WHO/HTM/TB/2015.25. 2015.
45. Cox JA, Lukande RL, Kalungi S, Van Marck E, Van de Vijver K, Kambugu A, et al. Is Urinary Lipoarabinomannan the Result of Renal Tuberculosis? Assessment of the Renal Histology in an Autopsy Cohort of Ugandan HIV-Infected Adults. *PLoS One*. 2015;10(4):e0123323.
46. Sakamuri RM, Price DN, Lee M, Cho SN, Barry CE, 3rd, Via LE, et al. Association of lipoarabinomannan with high density lipoprotein in blood: implications for diagnostics. *Tuberculosis (Edinb)*. 2013;93(3):301-7.
47. Sada E, Aguilar D, Torres M, Herrera T. Detection of lipoarabinomannan as a diagnostic test for tuberculosis. *J Clin Microbiol*. 1992;30(9):2415-8.
48. Broger T, Sossen B, du Toit E, Kerkhoff AD, Schutz C, Ivanova Reipold E. Novel lipoarabinomannan point-of-care tuberculosis test for people with HIV: a diagnostic accuracy study. *Lancet Infect Dis*. 2019 Aug;19(8):852-861.
49. Lewis DK, Peters R, Schijffelen MJ, Joaki G, Walsh AL, Kublin JG, et al. Clinical indicators of mycobacteraemia in adults admitted to hospital in Blantyre. *Malawi Med J*. 2003;15(2):56-60.

50. Kerkhoff AD, Barr DA, Schutz C, Burton R, Nicol MP, Lawn SD, et al. Disseminated tuberculosis among hospitalised HIV patients in South Africa: a common condition that can be rapidly diagnosed using urine-based assays. *Sci Rep*. 2017;7(1):10931.

51. Metcalfe JZ, Everett CK, Steingart KR, Cattamanchi A, Huang L, Hopewell PC, et al. Interferon-gamma release assays for active pulmonary tuberculosis diagnosis in adults in low- and middle-income countries: systematic review and meta-analysis. *J Infect Dis*. 2011;204 Suppl 4:S1120-9.

52. World Health Organization. *Use of tuberculosis interferon-gamma release assays (IGRAs) in low- and middle-income countries: Policy statement.* WHO/HTM/TB/2011.18. 2011.

53. Nanjappa V, Thomas JK, Marimuthu A, Muthusamy B, Radhakrishnan A, Sharma R, et al. Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Res*. 2014;42(Database issue):D959-65.

54. Zhang H, Liu AY, Loriaux P, Wollscheid B, Zhou Y, Watts JD, et al. Mass spectrometric detection of tissue proteins in plasma. *Mol Cell Proteomics*. 2007;6(1):64-71.

55. Papadopoulos MC, Abel PM, Agranoff D, Stich A, Tarelli E, Bell BA, et al. A novel and accurate diagnostic test for human African trypanosomiasis. *Lancet*. 2004;363(9418):1358-63.

56. World Health Organization. *Guidelines for the treatment of malaria. Third edition.* ISBN: 978 92 4 154912 7. 2015.

57. Yonekita T, Ohtsuki R, Hojo E, Morishita N, Matsumoto T, Aizawa T, et al. Development of a novel multiplex lateral flow assay using an antimicrobial peptide for the detection of Shiga toxin-producing *Escherichia coli*. *J Microbiol Methods*. 2013;93(3):251-6.

58. www.globalsources.com. Dengue NS1 -IgM -IgG Rapid Test. Available from: https://www.globalsources.com/?source=GSOL_TopNav_OM [Accessed 26th September 2019].

59. Propper CR, Tester JT, Vail TL, Venkatraman N. *Rapid multiplex lateral flow assay device*. Patent number: US20130280698A1. 2013.

60. Corstjens PL, Tjon Kon Fat EM, de Dood CJ, van der Ploeg-van Schip JJ, Franken KL, Chegou NN, et al. Multi-center evaluation of a user-friendly lateral flow assay to determine IP-10 and CCL4 levels in blood of TB and non-TB cases in Africa. *Clin Biochem*. 2016;49(1-2):22-31.

61. Bio-Rad. *SELDI technology.* Available from: https://www.bio-rad.com/en-uk/category/seldi-technology?ID=5781b780-ddf1-4f80-8d32-58e308e33a60 [Accessed 26th September 2019].

62. World Health Organization. *World Malaria Report.* ISBN 978-92-4-156565-3. 2018.

63. Agranoff D, Fernandez-Reyes D, Papadopoulos MC, Rojas SA, Herbster M, Loosemore A, et al. Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. *Lancet*. 2006;368(9540):1012-21.

64. Liu JY, Jin L, Zhao MY, Zhang X, Liu CB, Zhang YX, et al. New serum biomarkers for detection of tuberculosis using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry. *Clin Chem Lab Med*. 2011;49(10):1727-33.

65. Zhang J, Wu X, Shi L, Liang Y, Xie Z, Yang Y, et al. Diagnostic serum proteomic analysis in patients with active tuberculosis. *Clin Chim Acta*. 2012;413(9-10):883-7.

66. Sandhu G, Battaglia F, Ely BK, Athanasakis D, Montoya R, Valencia T, et al. Discriminating active from latent tuberculosis in patients presenting to community clinics. *PLoS One*. 2012;7(5):e38080.

67. Liu J, Jiang T, Wei L, Yang X, Wang C, Zhang X, et al. The discovery and identification of a candidate proteomic biomarker of active tuberculosis. *BMC Infect Dis*. 2013;13:506.

68. Song SH, Han M, Choi YS, Dan KS, Yang MG, Song J, et al. Proteomic profiling of serum from patients with tuberculosis. *Ann Lab Med*. 2014;34(5):345-53.
69. Xu D, Li Y, Li X, Wei LL, Pan Z, Jiang TT, et al. Serum protein S100A9, SOD3, and MMP9 as new diagnostic biomarkers for pulmonary tuberculosis by iTRAQ-coupled two-dimensional LC-MS/MS. *Proteomics*. 2015;15(1):58-67.
70. Achkar JM, Cortes L, Croteau P, Yanofsky C, Mentinova M, Rajotte I, et al. Host Protein Biomarkers Identify Active Tuberculosis in HIV Uninfected and Co-infected Individuals. *EBioMedicine*. 2015;2(9):1160-8.
71. Chegou NN, Sutherland JS, Malherbe S, Crampin AC, Corstjens PL, Geluk A, et al. Diagnostic performance of a seven-marker serum protein biosignature for the diagnosis of active TB disease in African primary healthcare clinic attendees with signs and symptoms suggestive of TB. *Thorax*. 2016;71:785–794.
72. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, et al. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med*. 2013;10(10):e1001538.
73. Herberg JA, Kaforou M, Wright VJ, Shailes H, Eleftherohorinou H, Hoggart CJ, et al. Diagnostic Test Accuracy of a 2-Transcript Host RNA Signature for Discriminating Bacterial vs Viral Infection in Febrile Children. *JAMA*. 2016;316(8):835-45.
74. Putri FA, Burhan E, Nawas A, Soepandi PZ, Sutoyo DK, Agustin H, et al. Body mass index predictive of sputum culture conversion among MDR-TB patients in Indonesia. *Int J Tuberc Lung Dis*. 2014;18(5):564-70.
75. Wejse C, Olesen R, Rabna P, Kaestel P, Gustafson P, Aaby P, et al. Serum 25-hydroxyvitamin D in a West African population of tuberculosis patients and unmatched healthy controls. *Am J Clin Nutr*. 2007;86(5):1376-83.
76. Velayutham BV, Allaudeen IS, Sivaramakrishnan GN, Perumal V, Nair D, Chinnaiyan P, et al. Sputum culture conversion with moxifloxacin-containing regimens in the treatment of patients with newly diagnosed sputum-positive pulmonary tuberculosis in South India. *Clin Infect Dis*. 2014;59(10):e142-9.
77. Hesseling AC, Walzl G, Enarson DA, Carroll NM, Duncan K, Lukey PT, et al. Baseline sputum time to detection predicts month two culture conversion and relapse in non-HIV-infected patients. *Int J Tuberc Lung Dis*. 2010;14(5):560-
78. Liu Q, Lu P, Martinez L, Yang H, Lu W, Ding X, et al. Factors affecting time to sputum culture conversion and treatment outcome of patients with multidrug-resistant tuberculosis in China. *BMC Infect Dis*. 2018;18(1):114.
79. Atif M, Sulaiman SA, Shafie AA, Ali I, Asif M, Babar ZU. Treatment outcome of new smear positive pulmonary tuberculosis patients in Penang, Malaysia. *BMC Infect Dis*. 2014;14:399.
80. Agbor AA, Bigna JJ, Billong SC, Tejiokem MC, Ekali GL, Plottel CS, et al. Factors associated with death during tuberculosis treatment of patients co-infected with HIV at the Yaounde Central Hospital, Cameroon: an 8-year hospital-based retrospective cohort study (2006-2013). *PLoS One*. 2014;9(12):e115211.
81. European and Developing Countries Clinical Trials Partnership. *ScreenTB: Evaluation of host biomarker-based point-of-care tests for targeted screening for active TB.* Available from: http://screen-tb.eu/ [Accessed 26th September 2019].

Chapter II

1. Chegou NN, Sutherland JS, Malherbe S, Crampin AC, Corstjens PL, Geluk A, et al. Diagnostic performance of a seven-marker serum protein biosignature for the diagnosis of active TB disease in African primary healthcare clinic attendees with signs and symptoms suggestive of TB. *Thorax*. 2016.
2. Hoggart, J, Clive. Parallel Regularised Regression Model Search for sparse biosignature discovery. 2018. Available from https://www.biorxiv.org/content/10.1101/355479v3 [accessed 1st September 2019].
3. Organization WH. *High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting.*; 2014. Report No.: WHO/HTM/TB/2014.18.

Chapter III

1. Chegou NN, Sutherland JS, Malherbe S, Crampin AC, Corstjens PL, Geluk A, et al. Diagnostic performance of a seven-marker serum protein biosignature for the diagnosis of active TB disease in African primary healthcare clinic attendees with signs and symptoms suggestive of TB. *Thorax*. 2016.
2. McQueen MJ, Hawken S, Wang X, Ounpuu S, Sniderman A, Probstfield J, et al. Lipids, lipoproteins, and apolipoproteins as risk markers of myocardial infarction in 52 countries (the INTERHEART study): a case-control study. *Lancet*. 2008;372(9634):224-33.
3. Ansari M, McKeigue PM, Skerka C, Hayward C, Rudan I, Vitart V, et al. Genetic influences on plasma CFH and CFHR1 concentrations and their role in susceptibility to age-related macular degeneration. *Hum Mol Genet*. 2013;22(23):4857-69.
4. Aroner SA, Yang M, Li J, Furtado JD, Sacks FM, Tjønneland A, et al. Apolipoprotein C-III and High-Density Lipoprotein Subspecies Defined by Apolipoprotein C-III in Relation to Diabetes Risk. *Am J Epidemiol.* 2017;186(6):736-44.
5. Ingenbleek Y, McCully KS. Vegetarianism produces subclinical malnutrition, hyperhomocysteinemia and atherogenesis. *Nutrition*. 2012;28(2):148-53.
6. Preas HL, Nylen ES, Snider RH, Becker KL, White JC, Agosti JM, et al. Effects of anti-inflammatory agents on serum levels of calcitonin precursors during human experimental endotoxemia. *J Infect Dis*. 2001;184(3):373-6.
7. Olpin SE, Price CP. High density lipoprotein and apolipoprotein A-I in acute phase response. *Atherosclerosis*. 1988;69(1):93-4.
8. Organization WH. *High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting.* Report No.: WHO/HTM/TB/2014.18. 2014.
9. Agranoff D, Fernandez-Reyes D, Papadopoulos MC, Rojas SA, Herbster M, Loosemore A, et al. Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. *Lancet*. 2006;368(9540):1012-21.
10. Achkar JM, Cortes L, Croteau P, Yanofsky C, Mentinova M, Rajotte I, et al. Host Protein Biomarkers Identify Active Tuberculosis in HIV Uninfected and Co-infected Individuals. *EBioMedicine*. 2015;2(9):1160-8.
11. Sandhu G, Battaglia F, Ely BK, Athanasakis D, Montoya R, Valencia T, et al. Discriminating active from latent tuberculosis in patients presenting to community clinics. *PLoS One*. 2012;7(5):e38080.

12. De Groote MA, Sterling DG, Hraha T, Russell TM, Green LS, Wall K, et al. Discovery and Validation of a Six-Marker Serum Protein Signature for the Diagnosis of Active Pulmonary Tuberculosis. *J Clin Microbiol.* 2017;55(10):3057-71.
13. Santos VS, Goletti D, Kontogianni K, Adams ER, Molina-Moya B, Dominguez J, et al. Acute phase proteins and IP-10 as triage tests for the diagnosis of tuberculosis: systematic review and meta-analysis. *Clin Microbiol Infect*. 2019;25(2):169-77.
14. Yoon C, Semitala FC, Atuhumuza E, Katende J, Mwebe S, Asege L, et al. Point-of-care C-reactive protein-based tuberculosis screening for people living with HIV: a diagnostic accuracy study. *Lancet Infect Dis*. 2017;17(12):1285-92.
15. Hoggart, CJ. Parallel Regularised Regression Model Search for sparse biosignature discovery. 2018. Available from https://www.biorxiv.org/content/10.1101/355479v3 [accessed 1st September 2019].
16. Organization WH. *High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting.*; 2014. Report No.: WHO/HTM/TB/2014.18.
17. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature*. 2010;466(7309):973-7.
18. Ripoche J, Day AJ, Harris TJ, Sim RB. The complete amino acid sequence of human complement factor H. *Biochem J*. 1988;249(2):593-602.

Chapter IV

1. Haralambous E, Dolly SO, Hibberd ML, Litt DJ, Udalova IA, O'Dwyer C, et al. Factor H, a regulator of complement activity, is a major determinant of meningococcal disease susceptibility in UK Caucasian patients. *Scand J Infect Dis*. 2006;38(9):764-71.
2. Davila S, Wright VJ, Khor CC, Sim KS, Binder A, Breunis WB, et al. Genome-wide association study identifies variants in the CFH region associated with host susceptibility to meningococcal disease. *Nat Genet*. 2010;42(9):772-6.
3. van der Maten E, Westra D, van Selm S, Langereis JD, Bootsma HJ, van Opzeeland FJ, et al. Complement Factor H Serum Levels Determine Resistance to Pneumococcal Invasive Disease. *J Infect Dis*. 2016;213(11):1820-7.
4. Parente R, Clark SJ, Inforzato A, Day AJ. Complement factor H in host defense and immune evasion. *Cell Mol Life Sci.* 2016.
5. Carroll MV, Lack N, Sim E, Krarup A, Sim RB. Multiple routes of complement activation by *Mycobacterium bovis BCG. Mol Immunol.* 2009;46(16):3367-78.
6. Abdul-Aziz M, Tsolaki AG, Kouser L, Carroll MV, Al-Ahdal MN, Sim RB, et al. Complement factor H interferes with *Mycobacterium bovis BCG* entry into macrophages and modulates the pro-inflammatory cytokine response. *Immunobiology*. 2016;221(9):944-52.
7. Holmes LV, Strain L, Staniforth SJ, Moore I, Marchbank K, Kavanagh D, et al. Determining the population frequency of the CFHR3/CFHR1 deletion at 1q32. *PLoS One.* 2013;8(4):e60352.
8. Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat Genet*. 2010;42(9):739-41.

9. Curtis J, Luo Y, Zenner HL, Cuchet-Lourenco D, Wu C, Lo K, et al. Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. *Nat Genet.* 2015;47(5):523-7.
10. Luo H, Zhu B, Gong L, Yang J, Jiang Y, Zhou X. The value of serum prealbumin in the diagnosis and therapeutic response of tuberculosis: a retrospective study. *PLoS One*. 2013;8(11):e79940.
11. Christie WW. Plasma Lipoproteins. 2019 [Available from: http://www.lipidhome.co.uk/lipids/simple/lipoprot/index.htm] [accessed 4th February 2019].
12. Uniprot. UniProtKB - P02656 (APOC3_HUMAN) 2019 [Available from: https://www.uniprot.org/uniprot/P02656.] [accessed 14th March 2019].
13. Noris M, Remuzzi G. Overview of complement activation and regulation. *Semin Nephrol.* 2013;33(6):479-92.
14. Wang C, Wei LL, Shi LY, Pan ZF, Yu XM, Li TY, et al. Screening and identification of five serum proteins as novel potential biomarkers for cured pulmonary tuberculosis. *Sci Rep*. 2015;5:15615.
15. Bajaj G, Rattan A, Ahmad P. Prognostic value of 'C' reactive protein in tuberculosis. *Indian Pediatr.* 1989;26(10):1010-3.
16. Hosp M, Elliott AM, Raynes JG, Mwinga AG, Luo N, Zangerle R, et al. Neopterin, beta 2-microglobulin, and acute phase proteins in HIV-1-seropositive and -seronegative Zambian patients with tuberculosis. *Lung.* 1997;175(4):265-75.
17. Lawn SD, Obeng J, Acheampong JW, Griffin GE. Resolution of the acute-phase response in West African patients receiving treatment for pulmonary tuberculosis. *Int J Tuberc Lung Dis.* 2000;4(4):340-4.
18. Ostrowski SR, Ravn P, Hoyer-Hansen G, Ullum H, Andersen AB. Elevated levels of soluble urokinase receptor in serum from mycobacteria infected patients: still looking for a marker of treatment efficacy. *Scand J Infect Dis*. 2006;38(11-12):1028-32.

Chapter V

1. Herberg JA, Kaforou M, Wright VJ, Shailes H, Eleftherohorinou H, Hoggart CJ, et al. Diagnostic Test Accuracy of a 2-Transcript Host RNA Signature for Discriminating Bacterial vs Viral Infection in Febrile Children. *JAMA.* 2016;316(8):835-45.
2. UniProt. UniProtKB - P01011 (AACT_HUMAN). 2019 [Available from: https://www.uniprot.org/uniprot/P01011#sequences] [accessed 15th May 2019].
3. Evans J. *Straightforward Statistics for the Behavioral Sciences.* Duxbury Press; 1996.
4. Plesner T, Wiik A. Demonstration of electrophoretic heterogeneity of serum beta 2-microglobulin in systemic lupus erythematosus and rheumatoid arthritis: evidence against autoantibodies to beta 2-microglobulin. *Scand J Immunol.* 1979;9(3):247-54.
5. Nissen MH, Thim L, Christensen M. Purification and biochemical characterization of the complete structure of a proteolytically modified beta-2-microglobulin with biological activity. *Eur J Biochem*. 1987;163(1):21-8.
6. Trenchevska O, Kamcheva E, Nedelkov D. Mass spectrometric immunoassay for quantitative determination of protein biomarker isoforms. *J Proteome Res*. 2010;9(11):5969-73.

7. Simonsen AH, Bahl JM, Danborg PB, Lindstrom V, Larsen SO, Grubb A, et al. Pre-analytical factors influencing the stability of cerebrospinal fluid proteins. *J Neurosci Methods*. 2013;215(2):234-40.
8. Butterfield LH, Potter DM, Kirkwood JM. Multiplex serum biomarker assessments: technical and biostatistical issues. *J Transl Med.* 2011;9:173.
9. Thompson S, Chesher D. Lot-to-Lot Variation. *Clin Biochem Rev*. 2018;39(2):51-60.
10. Pang S, Smith J, Onley D, Reeve J, Walker M, Foy C. A comparability study of the emerging protein array platforms with established ELISA procedures. *J Immunol Methods.* 2005;302(1-2):1-12.
11. Le Bastard N, Coart E, Vanderstichele H, Vanmechelen E, Martin JJ, Engelborghs S. Comparison of two analytical platforms for the clinical qualification of Alzheimer's disease biomarkers in pathologically-confirmed dementia. *J Alzheimers Dis*. 2013;33(1):117-31.
12. Wang LS, Leung YY, Chang SK, Leight S, Knapik-Czajka M, Baek Y, et al. Comparison of xMAP and ELISA assays for detecting cerebrospinal fluid biomarkers of Alzheimer's disease. *J Alzheimers Dis*. 2012;31(2):439-45.
13. Faull M, Ching SY, Jarmolowicz AI, Beilby J, Panegyres PK. Comparison of two methods for the analysis of CSF Aβ and tau in the diagnosis of Alzheimer's disease. *Am J Neurodegener Dis.* 2014;3(3):143-51.
14. Corstjens PL, Tjon Kon Fat EM, de Dood CJ, van der Ploeg-van Schip JJ, Franken KL, Chegou NN, et al. Multi-center evaluation of a user-friendly lateral flow assay to determine IP-10 and CCL4 levels in blood of TB and non-TB cases in Africa. *Clin Biochem.* 2016;49(1-2):22-31.

Chapter VI

1. Wejse C, Olesen R, Rabna P, Kaestel P, Gustafson P, Aaby P, et al. Serum 25-hydroxyvitamin D in a West African population of tuberculosis patients and unmatched healthy controls. *Am J Clin Nutr.* 2007;86(5):1376-83.
2. Putri FA, Burhan E, Nawas A, Soepandi PZ, Sutoyo DK, Agustin H, et al. Body mass index predictive of sputum culture conversion among MDR-TB patients in Indonesia. *The international journal of tuberculosis and lung disease*. 2014;18(5):564-70.
3. Velayutham BV, Allaudeen IS, Sivaramakrishnan GN, Perumal V, Nair D, Chinnaiyan P, et al. Sputum culture conversion with moxifloxacin-containing regimens in the treatment of patients with newly diagnosed sputum-positive pulmonary tuberculosis in South India. *Clin Infect Dis*. 2014;59(10):e142-9.
4. Hesseling AC, Walzl G, Enarson DA, Carroll NM, Duncan K, Lukey PT, et al. Baseline sputum time to detection predicts month two culture conversion and relapse in non-HIV-infected patients. *The international journal of tuberculosis and lung disease.* 2010;14(5):560-70.
5. Liu Q, Lu P, Martinez L, Yang H, Lu W, Ding X, et al. Factors affecting time to sputum culture conversion and treatment outcome of patients with multidrug-resistant tuberculosis in China. *BMC Infect Dis*. 2018;18(1):114.
6. Atif M, Sulaiman SA, Shafie AA, Ali I, Asif M, Babar ZU. Treatment outcome of new smear positive pulmonary tuberculosis patients in Penang, Malaysia. *BMC Infect Dis.* 2014;14:399.

7. Agbor AA, Bigna JJ, Billong SC, Tejiokem MC, Ekali GL, Plottel CS, et al. Factors associated with death during tuberculosis treatment of patients co-infected with HIV at the Yaounde Central Hospital, Cameroon: an 8-year hospital-based retrospective cohort study (2006-2013). *PloS one*. 2014;9(12):e115211.
8. Murray CJ, Ortblad KF, Guinovart C, Lim SS, Wolock TM, Roberts DA, et al. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet.* 2014;384(9947):1005-70.
9. Seddon JA, Chiang SS, Esmail H, Coussens AK. The Wonder Years: What Can Primary School Children Teach Us About Immunity to *Mycobacterium tuberculosis? Front Immunol*. 2018;9:2946.
10. Mitnick C, Bayona J, Palacios E, Shin S, Furin J, Alcantara F, et al. Community-based therapy for multidrug-resistant tuberculosis in Lima, Peru. *The New England journal of medicine*. 2003;348(2):119-28.
11. Leimane V, Riekstina V, Holtz TH, Zarovska E, Skripconoka V, Thorpe LE, et al. Clinical outcome of individualised treatment of multidrug-resistant tuberculosis in Latvia: a retrospective cohort study. *Lancet*. 2005;365(9456):318-26.
12. Van Lettow M, Kumwenda JJ, Harries AD, Whalen CC, Taha TE, Kumwenda N, et al. Malnutrition and the severity of lung disease in adults with pulmonary tuberculosis in Malawi. *The international journal of tuberculosis and lung disease.* 2004;8(2):211-7.
13. Rudolf F, Joaquim LC, Vieira C, Bjerregaard-Andersen M, Andersen A, Erlandsen M, et al. The Bandim tuberculosis score: reliability and comparison with the Karnofsky performance score. *Scandinavian journal of infectious diseases.* 2013;45(4):256-64.
14. Lonnroth K, Williams BG, Cegielski P, Dye C. A consistent log-linear relationship between tuberculosis incidence and body mass index. *Int J Epidemiol.* 2010;39(1):149-55.
15. Pepper DJ, Marais S, Wilkinson RJ, Bhaijee F, Maartens G, McIlleron H, et al. Clinical deterioration during antituberculosis treatment in Africa: incidence, causes and risk factors. *BMC Infect Dis*. 2010;10:83.
16. Macpherson P, Dimairo M, Bandason T, Zezai A, Munyati SS, Butterworth AE, et al. Risk factors for mortality in smear-negative tuberculosis suspects: a cohort study in Harare, Zimbabwe. *The international journal of tuberculosis and lung disease*. 2011;15(10):1390-6.
17. Ismail I, Bulgiba A. Predictors of death during tuberculosis treatment in TB/HIV co-infected patients in Malaysia. *PloS one*. 2013;8(8):e73250.
18. Pepper DJ, Schomaker M, Wilkinson RJ, de Azevedo V, Maartens G. Independent predictors of tuberculosis mortality in a high HIV prevalence setting: a retrospective cohort study. *AIDS Res Ther.* 2015;12:35.
19. Haarburger D, Bergström J, Pillay TS. Serum proteome changes following human immunodeficiency virus infection. *Clin Lab.* 2013;59(5-6):639-46.
20. Adedeji AL, Adenikinju RO, Ajele JO, Olawoye TL. Serum protein electrophoresis under effective control of HIV-1 disease progression. *EXCLI J.* 2014;13:761-71.
21. Pheiffer C, Carroll NM, Beyers N, Donald P, Duncan K, Uys P, et al. Time to detection of *Mycobacterium tuberculosis* in BACTEC systems as a viable alternative to colony counting. *The international journal of tuberculosis and lung disease*. 2008;12(7):792-8.
22. Stop TB Partnership. *Laboratory Diagnosis of Tuberculosis by Sputum Microscopy: The Handbook*.; 2013. Report No.: ISBN: 978-1-74243-602-9.

23. Chikamatsu K, Aono A, Kato T, Takaki A, Yamada H, Sasaki Y, et al. COBAS(R) TaqMan(R) MTB, smear positivity grade and MGIT culture; correlation analyses of three methods for bacillary quantification. *Journal of infection and chemotherapy*. 2016;22(1):19-23.

24. Hanrahan CF, Theron G, Bassett J, Dheda K, Scott L, Stevens W, et al. Xpert MTB/RIF as a measure of sputum bacillary burden. Variation by HIV status and immunosuppression. *American journal of respiratory and critical care medicine*. 2014;189(11):1426-34.

25. Olaru ID, Heyckendorf J, Grossmann S, Lange C. Time to culture positivity and sputum smear microscopy during tuberculosis therapy. *PloS one*. 2014;9(8):e106075.

26. Magee MJ, Kempker RR, Kipiani M, Gandhi NR, Darchia L, Tukvadze N, et al. Diabetes mellitus is associated with cavities, smear grade, and multidrug-resistant tuberculosis in Georgia. *The international journal of tuberculosis and lung disease.* 2015;19(6):685-92.

27. Rieder HL, Lauritsen JM, Naranbat N, Katamba A, Laticevschi D, Mabaera B. Quantitative differences in sputum smear microscopy results for acid-fast bacilli by age and sex in four countries. *The international journal of tuberculosis and lung disease.* 2009;13(11):1393-8.

28. Yone EW, Kengne AP, Moifo B, Kuaban C. Prevalence and determinants of extrapulmonary involvement in patients with pulmonary tuberculosis in a Sub-Saharan African country: a cross-sectional study. *Scandinavian journal of infectious diseases*. 2013;45(2):104-11.

29. Nglazi MD, Bekker LG, Wood R, Kaplan R. The impact of HIV status and antiretroviral treatment on TB treatment outcomes of new tuberculosis patients attending co-located TB and ART services in South Africa: a retrospective cohort study. *BMC Infect Dis*. 2015;15:536.

30. King L, Munsiff SS, Ahuja SD. Achieving international targets for tuberculosis treatment success among HIV-positive patients in New York City. *The international journal of tuberculosis and lung disease*. 2010;14(12):1613-20.

31. Yone É, Kuaban C, Kengne AP. Impact of HIV infection on the evolution of tuberculosis among adult patients in Yaounde, Cameroon. *Rev Pneumol Clin*. 2012;68(6):338-44.

32. Waitt CJ, Squire SB. A systematic review of risk factors for death in adults during and after tuberculosis treatment. *The international journal of tuberculosis and lung disease.* 2011;15(7):871-85.

33. Stop TB Partnership. *Laboratory Diagnosis of Tuberculosis by Sputum Microscopy: The Handbook*.; 2013. Report No.: ISBN: 978-1-74243-602-9.

34. World Health Organization. BMI classification. 2009. [Available from: https://web.archive.org/web/20090418181049/http:/www.who.int/bmi/index.jsp?introPage=intro_3.html] [accessed 14th February 2019].

35. Office for National Statistics. Overview of the UK population: November 2018 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/november2018] [accessed 14th February 2019].

36. World Health Organization. *World Report on Ageing and Health*. 2015. Report No.: ISBN 978 92 4 156504 2.

37. Rajagopalan S, Yoshikawa TT. Tuberculosis in the elderly. *Z Gerontol Geriatr*. 2000;33(5):374-80.

38. Gabay C, Kushner I. Acute-phase proteins and other systemic responses to inflammation. *N Engl J Med*. 1999;340(6):448-54.

39. Verbon A, Juffermans N, Van Deventer SJ, Speelman P, Van Deutekom H, Van Der Poll T. Serum concentrations of cytokines in patients with active tuberculosis (TB) and after treatment. *Clinical and experimental immunology.* 1999;115(1):110-3.

40. O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MP. The immune response in tuberculosis. *Annu Rev Immunol.* 2013;31:475-527.

41. Price NM, Farrar J, Tran TT, Nguyen TH, Tran TH, Friedland JS. Identification of a matrix-degrading phenotype in human tuberculosis in vitro and in vivo. *J Immunol*. 2001;166(6):4223-30.

42. Belton M, Brilha S, Manavaki R, Mauri F, Nijran K, Hong YT, et al. Hypoxia and tissue destruction in pulmonary TB. *Thorax*. 2016;71(12):1145-53.

43. Olpin SE, Price CP. High density lipoprotein and apolipoprotein A-I in acute phase response. *Atherosclerosis.* 1988;69(1):93-4.

44. Lacorte JM, Beigneux A, Parant M, Chambaz J. Repression of apoC-III gene expression by TNFalpha involves C/EBPdelta/NF-IL6beta via an IL-1 independent pathway. *FEBS Lett.* 1997;415(2):217-20.

45. Gebremicael G, Amare Y, Challa F, Gebreegziabxier A, Medhin G, Wolde M, et al. Lipid Profile in Tuberculosis Patients with and without Human Immunodeficiency Virus Infection. *Int J Chronic Dis*. 2017;2017:3843291.

46. Russell DG. Who puts the tubercle in tuberculosis? *Nat Rev Microbiol*. 2007;5(1):39-47.

47. Kim MJ, Wainwright HC, Locketz M, Bekker LG, Walther GB, Dittrich C, et al. Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism. *EMBO molecular medicine*. 2010;2(7):258-74.

48. Sposito AC, Carmo HR, Barreto J, Sun L, Carvalho LSF, Feinstein SB, et al. HDL-Targeted Therapies During Myocardial Infarction. *Cardiovasc Drugs Ther*. 2019;33(3):371-81.

49. Parihar SP, Guler R, Khutlang R, Lang DM, Hurdayal R, Mhlanga MM, et al. Statin therapy reduces the *mycobacterium tuberculosis* burden in human macrophages and in mice by enhancing autophagy and phagosome maturation. The *Journal of infectious diseases*. 2014;209(5):754-63.

50. Dutta NK, Bruiners N, Pinn ML, Zimmerman MD, Prideaux B, Dartois V, et al. Statin adjunctive therapy shortens the duration of TB treatment in mice. *J Antimicrob Chemother*. 2016;71(6):1570-7.

51. Kim MC, Yun SC, Lee SO, Choi SH, Kim YS, Woo JH, et al. Association between Tuberculosis, Statin Use, and Diabetes: A Propensity Score-Matched Analysis. *Am J Trop Med Hyg*. 2019;101(2):350-6.

52. Emsley J, White HE, O'Hara BP, Oliva G, Srinivasan N, Tickle IJ, et al. Structure of pentameric human serum amyloid P component. *Nature*. 1994;367(6461):338-45.

53. Xiang W, Cox N, Gomer RH. Identification of compounds that decrease numbers of Mycobacteria in human macrophages in the presence of serum amyloid P. *J Leukoc Biol*. 2017;102(3):857-69.

54. Bogdan C. Nitric oxide and the immune response. *Nat Immunol.* 2001;2(10):907-16.

55. Parente R, Clark SJ, Inforzato A, Day AJ. Complement factor H in host defense and immune evasion. *Cell Mol Life Sci.* 2016.

56. Abdul-Aziz M, Tsolaki AG, Kouser L, Carroll MV, Al-Ahdal MN, Sim RB, et al. Complement factor H interferes with *Mycobacterium bovis BCG* entry into macrophages and modulates the pro-inflammatory cytokine response. *Immunobiology*. 2016;221(9):944-52.

57. Vijayan AL, Vanimaya, Ravindran S, Saikant R, Lakshmi S, Kartik R, et al. Procalcitonin: a promising diagnostic marker for sepsis and antibiotic therapy. *J Intensive Care.* 2017;5:51.

58. Rasmussen TA, Sogaard OS, Camara C, Andersen PL, Wejse C. Serum procalcitonin in pulmonary tuberculosis. *The international journal of tuberculosis and lung disease*. 2011;15(2):251-6, i.

59. Wang W, Knovich MA, Coffman LG, Torti FM, Torti SV. Serum ferritin: Past, present and future. *Biochim Biophys Acta*. 2010;1800(8):760-9.

60. Sivakolundu S, Mannela UD, Jain S, Srikantam A, Peri S, Pandey SD, et al. Serum iron profile and ELISA-based detection of antibodies against the iron-regulated protein HupB of *Mycobacterium tuberculosis* in TB patients and household contacts in Hyderabad (Andhra Pradesh), India. *Trans R Soc Trop Med Hyg*. 2013;107(1):43-50.

61. Strandin T, Hepojoki J, Laine O, Makela S, Klingstrom J, Lundkvist A, et al. Interferons Induce STAT1-Dependent Expression of Tissue Plasminogen Activator, a Pathogenicity Factor in Puumala Hantavirus Disease. *The Journal of infectious diseases.* 2016;213(10):1632-41.

62. Haidaris PJ. Induction of fibrinogen biosynthesis and secretion from cultured pulmonary epithelial cells. *Blood*. 1997;89(3):873-82.

63. Robson SC, White NW, Aronson I, Woollgar R, Goodman H, Jacobs P. Acute-phase response and the hypercoagulable state in pulmonary tuberculosis. *Br J Haematol.* 1996;93(4):943-9.

64. Sack GH, Jr. Serum amyloid A - a review. *Mol Med.* 2018;24(1):46.

65. Jiang TT, Shi LY, Wei LL, Li X, Yang S, Wang C, et al. Serum amyloid A, protein Z, and C4b-binding protein beta chain as new potential biomarkers for pulmonary tuberculosis. *PloS one*. 2017;12(3):e0173304.

66. Ganter U, Arcone R, Toniatti C, Morrone G, Ciliberto G. Dual control of C-reactive protein gene expression by interleukin-1 and interleukin-6. *EMBO J*. 1989;8(12):3773-9.

67. Patel DN, King CA, Bailey SR, Holt JW, Venkatachalam K, Agrawal A, et al. Interleukin-17 stimulates C-reactive protein expression in hepatocytes and smooth muscle cells via p38 MAPK and ERK1/2-dependent NF-kappaB and C/EBPbeta activation. *J Biol Chem.* 2007;282(37):27229-38.

68. Oni T, Burke R, Tsekela R, Bangani N, Seldon R, Gideon HP, et al. High prevalence of subclinical tuberculosis in HIV-1-infected persons without advanced immunodeficiency: implications for TB screening. *Thorax*. 2011;66(8):669-73.

69. do Nascimento CO, Hunter L, Trayhurn P. Regulation of haptoglobin gene expression in 3T3-L1 adipocytes by cytokines, catecholamines, and PPARgamma. *Biochemical and biophysical research communications*. 2004;313(3):702-8.

70. Kasvosve I, Gomo ZA, Mvundura E, Moyo VM, Saungweme T, Khumalo H, et al. Haptoglobin polymorphism and mortality in patients with tuberculosis. *Int J Tuberc Lung Dis.* 2000;4(8):771-5.

71. Grange JM, Kardjito T, Setiabudi I. A study of acute-phase reactant proteins in Indonesian patients with pulmonary tuberculosis. *Tubercle.* 1984;65(1):23-39.

72. Zahedi K, Whitehead AS. Regulation of mouse serum amyloid P gene expression by cytokines in vitro. *Biochim Biophys Acta.* 1993;1176(1-2):162-8.

73. Green JA, Thi Hong Chau T, Farrar JJ, Friedland JS, Thwaites GE. CNS infection, CSF matrix metalloproteinase concentrations, and clinical/laboratory features. *Neurology.* 2011;76(6):577-9.

74. Seddon J, Kasprowicz V, Walker NF, Yuen HM, Sunpath H, Tezera L, et al. Procollagen III N-terminal propeptide and desmosine are released by matrix destruction in pulmonary tuberculosis. *The Journal of infectious diseases*. 2013;208(10):1571-9.

75. Gordon GM, Ledee DR, Feuer WJ, Fini ME. Cytokines and signaling pathways regulating matrix metalloproteinase-9 (MMP-9) expression in corneal epithelial cells. *J Cell Physiol*. 2009;221(2):402-11.

76. Hrabec E, Strek M, Zieba M, Kwiatkowska S, Hrabec Z. Circulation level of matrix metalloproteinase-9 is correlated with disease severity in tuberculosis patients. *The international journal of tuberculosis and lung disease*. 2002;6(8):713-9.
77. Banks RE, Forbes MA, Storr M, Higginson J, Thompson D, Raynes J, et al. The acute phase protein response in patients receiving subcutaneous IL-6. *Clinical and experimental immunology*. 1995;102(1):217-23.
78. Luo H, Zhu B, Gong L, Yang J, Jiang Y, Zhou X. The value of serum prealbumin in the diagnosis and therapeutic response of tuberculosis: a retrospective study. *PloS one.* 2013;8(11):e79940.
79. Pio R, Garcia J, Corrales L, Ajona D, Fleischhacker M, Pajares MJ, et al. Complement factor H is elevated in bronchoalveolar lavage fluid and sputum from patients with lung cancer. *Cancer Epidemiol Biomarkers Prev.* 2010;19(10):2665-72.
80. Brooimans RA, van der Ark AA, Buurman WA, van Es LA, Daha MR. Differential regulation of complement factor H and C3 production in human umbilical vein endothelial cells by IFN-gamma and IL-1. *J Immunol.* 1990;144(10):3835-40.
81. Azzurri A, Sow OY, Amedei A, Bah B, Diallo S, Peri G, et al. IFN-gamma-inducible protein 10 and pentraxin 3 plasma levels are tools for monitoring inflammation and disease activity in Mycobacterium tuberculosis infection. *Microbes Infect.* 2005;7(1):1-8.
82. Kim SJ, Kim HI, Lee YH, Kim SK. Production of tumor necrosis factor-alpha by alveolar macrophages from patients with pulmonary tuberculosis. *J Korean Med Sci*. 1991;6(1):45-53.
83. Alatas F, Alatas O, Metintas M, Ozarslan A, Erginel S, Yildirim H. Vascular endothelial growth factor levels in active pulmonary tuberculosis. *Chest*. 2004;125(6):2156-9.
84. Bell LCK, Noursadeghi M. Pathogenesis of HIV-1 and *Mycobacterium tuberculosis* co-infection. *Nat Rev Microbiol.* 2018;16(2):80-90.
85. Chegou NN, Sutherland JS, Malherbe S, Crampin AC, Corstjens PL, Geluk A, et al. Diagnostic performance of a seven-marker serum protein biosignature for the diagnosis of active TB disease in African primary healthcare clinic attendees with signs and symptoms suggestive of TB. *Thorax*. 2016.
86. Achkar JM, Cortes L, Croteau P, Yanofsky C, Mentinova M, Rajotte I, et al. Host Protein Biomarkers Identify Active Tuberculosis in HIV Uninfected and Co-infected Individuals. *EBioMedicine.* 2015;2(9):1160-8.
87. Lesosky M, Rangaka MX, Pienaar C, Coussens AK, Goliath R, Mathee S, et al. Plasma Biomarkers to Detect Prevalent or Predict Progressive Tuberculosis Associated With Human Immunodeficiency Virus-1. *Clin Infect Dis*. 2019;69(2):295-305.
88. Sutherland JS, Young JM, Peterson KL, Sanneh B, Whittle HC, Rowland-Jones SL, et al. Polyfunctional CD4(+) and CD8(+) T cell responses to tuberculosis antigens in HIV-1-infected patients before and after anti-retroviral treatment. *J Immunol.* 2010;184(11):6537-44.
89. Cavaillon JM. Cytokines and macrophages. *Biomed Pharmacother*. 1994;48(10):445-53.
90. Juffermans NP, Verbon A, van Deventer SJ, van Deutekom H, Speelman P, van der Poll T. Tumor necrosis factor and interleukin-1 inhibitors as markers of disease activity of tuberculosis. *American journal of respiratory and critical care medicine*. 1998;157(4 Pt 1):1328-31.
91. Tsao TC, Li L, Hsieh M, Liao S, Chang KS. Soluble TNF-alpha receptor and IL-1 receptor antagonist elevation in BAL in active pulmonary TB. *Eur Respir J*. 1999;14(3):490-5.

92. Schutz C, Barr D, Andrade BB, Shey M, Ward A, Janssen S, et al. Clinical, microbiologic, and immunologic determinants of mortality in hospitalized patients with HIV-associated tuberculosis: A prospective cohort study. *PLoS medicine*. 2019;16(7):e1002840.

93. Nouhin J, Pean P, Madec Y, Chevalier MF, Didier C, Borand L, et al. Interleukin-1 receptor antagonist, a biomarker of response to anti-TB treatment in HIV/TB co-infected patients. *J Infect.* 2017;74(5):456-65.

94. Sigal GB, Segal MR, Mathew A, Jarlsberg L, Wang M, Barbero S, et al. Biomarkers of Tuberculosis Severity and Treatment Effect: A Directed Screen of 70 Host Markers in a Randomized Clinical Trial. *EBioMedicine.* 2017;25:112-21.

95. Pankla R, Buddhisa S, Berry M, Blankenship DM, Bancroft GJ, Banchereau J, et al. Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. *Genome Biol*. 2009;10(11):R127.

96. NHS. C-reactive protein (CRP) test. 2018. [Available from: https://www.nhs.uk/conditions/blood-tests/types/] [accessed 26th October 2018].

97. Pepys MB, Baltz ML. Acute phase proteins with special reference to C-reactive protein and related proteins (pentaxins) and serum amyloid A protein. *Adv Immunol.* 1983;34:141-212.

98. Mendy J, Togun T, Owolabi O, Donkor S, Ota MO, Sutherland JS. C-reactive protein, Neopterin and Beta2 microglobulin levels pre and post TB treatment in The Gambia. *BMC Infect Dis*. 2016;16:115.

99. Wilson D, Moosa MS, Cohen T, Cudahy P, Aldous C, Maartens G. Evaluation of Tuberculosis Treatment Response With Serial C-Reactive Protein Measurements. *Open Forum Infect Dis*. 2018;5(11):ofy253.

100. Cudahy PGT, Warren JL, Cohen T, Wilson D. Trends in C-Reactive Protein, D-Dimer, and Fibrinogen during Therapy for HIV-Associated Multidrug-Resistant Tuberculosis. *Am J Trop Med Hyg*. 2018;99(5):1336-41.

101. Yoon C, Semitala FC, Atuhumuza E, Katende J, Mwebe S, Asege L, et al. Point-of-care C-reactive protein-based tuberculosis screening for people living with HIV: a diagnostic accuracy study. *Lancet Infect Dis*. 2017;17(12):1285-92.

102. Brown J, Clark K, Smith C, Hopwood J, Lynard O, Toolan M, et al. Variation in C - reactive protein response according to host and mycobacterial characteristics in active tuberculosis. *BMC Infect Dis*. 2016;16:265.

103. Chang JC, Wysocki A, Tchou-Wong KM, Moskowitz N, Zhang Y, Rom WN. Effect of *Mycobacterium tuberculosis* and its components on macrophages and the release of matrix metalloproteinases. *Thorax.* 1996;51(3):306-11.

104. Shapiro SD, Campbell EJ, Kobayashi DK, Welgus HG. Immune modulation of metalloproteinase production in human macrophages. Selective pretranslational suppression of interstitial collagenase and stromelysin biosynthesis by interferon-gamma. *J Clin Invest*. 1990;86(4):1204-10.

105. Elkington PT, Emerson JE, Lopez-Pascua LD, O'Kane CM, Horncastle DE, Boyle JJ, et al. *Mycobacterium tuberculosis* up-regulates matrix metalloproteinase-1 secretion from human airway epithelial cells via a p38 MAPK switch. *J Immunol.* 2005;175(8):5333-40.

106. Ong CWM, Fox K, Ettorre A, Elkington PT, Friedland JS. Hypoxia increases neutrophil-driven matrix destruction after exposure to *Mycobacterium tuberculosis. Scientific Reports*. 2018;8.

107. Ordonez AA, Tasneen R, Pokkali S, Xu ZY, Converse PJ, Klunk MH, et al. Mouse model of pulmonary cavitary tuberculosis and expression of matrix metalloproteinase-9. *Dis Model Mech*. 2016;9(7):779-88.

108. Chegou NN, Black GF, Kidd M, van Helden PD, Walzl G. Host markers in QuantiFERON supernatants differentiate active TB from latent TB infection: preliminary report. *BMC Pulm Med*. 2009;9:21.

109. Chegou NN, Sutherland JS, Namuganga AR, Corstjens PL, Geluk A, Gebremichael G, et al. Africa-wide evaluation of host biomarkers in QuantiFERON supernatants for the diagnosis of pulmonary tuberculosis. *Scientific reports*. 2018;8(1):2675.

110. Djoba Siawaya JF, Beyers N, van Helden P, Walzl G. Differential cytokine secretion and early treatment response in patients with pulmonary tuberculosis. *Clinical and experimental immunology*. 2009;156(1):69-77.

111. Kübler A, Luna B, Larsson C, Ammerman NC, Andrade BB, Orandle M, et al. *Mycobacterium tuberculosis* dysregulates MMP/TIMP balance to drive rapid cavitation and unrestrained bacterial proliferation. *J Pathol.* 2015;235(3):431-44.

112. Chakraborti S, Mandal M, Das S, Mandal A, Chakraborti T. Regulation of matrix metalloproteinases: an overview. *Mol Cell Biochem*. 2003;253(1-2):269-85.

113. Jenner W. A Treatise on Tuberculosis, the Constitutional Origin of Consumption and Scrofula. Br *Foreign Med Chir Rev.* 1853;12(24):443-52.

114. El-Sony AI, Mustafa SA, Khamis AH, Sobhi S, Enarson DA, Baraka OZ, et al. Symptoms in patients attending services for diagnosis of pulmonary tuberculosis in Sudan. *The international journal of tuberculosis and lung disease*. 2003;7(6):550-5.

115. Nunn P, Gicheha C, Hayes R, Gathua S, Brindle R, Kibuga D, et al. Cross-sectional survey of HIV infection among patients with tuberculosis in Nairobi, Kenya. *Tubercle and lung disease*. 1992;73(1):45-51.

116. Paton NI, Ng YM. Body composition studies in patients with wasting associated with tuberculosis. *Nutrition* 2006;22(3):245-51.

117. Karyadi E, Dolmans WM, West CE, Van Crevel R, Nelwan RH, Amin Z, et al. Cytokines related to nutritional status in patients with untreated pulmonary tuberculosis in Indonesia. *Asia Pac J Clin Nutr.* 2007;16(2):218-26.

118. Range NS, Malenganisho W, Temu MM, Changalucha J, Magnussen P, Krarup H, et al. Body composition of HIV-positive patients with pulmonary tuberculosis: a cross-sectional study in Mwanza, Tanzania. *Ann Trop Med Parasitol*. 2010;104(1):81-90.

119. Tsukaguchi K, Yoneda T, Yoshikawa M, Narita N, Enoki Y, Miyazaki R, et al. Interaction between interleukin-1 and tumor necrosis factor productions by peripheral blood monocytes and nutritional disturbance in active pulmonary tuberculosis. *Kekkaku.* 1991;66(7):477-84.

120. Waitt CJ, Peter KBN, White SA, Kampmann B, Kumwenda J, Heyderman RS, et al. Early deaths during tuberculosis treatment are associated with depressed innate responses, bacterial infection, and tuberculosis progression. *The Journal of infectious diseases*. 2011;204(3):358-62.

121. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature*. 2010;466(7309):973-7.

122. Janssen S, Schutz C, Ward AM, Huson MA, Wilkinson RJ, Burton R, et al. Hemostatic Changes Associated With Increased Mortality Rates in Hospitalized Patients With HIV-Associated Tuberculosis: A Prospective Cohort Study. *J Infect Dis*. 2017;215(2):247-58.

123. Matsuyama W, Hashiguchi T, Matsumuro K, Iwami F, Hirotsu Y, Kawabata M, et al. Increased serum level of vascular endothelial growth factor in pulmonary tuberculosis. *American journal of respiratory and critical care medicine*. 2000;162(3 Pt 1):1120-2.

124. Collazos J, Martinez E, Mayo J, Rinon M. Response of lymphocyte subsets in patients under treatment for tuberculosis. *Eur J Clin Microbiol Infect Dis*. 2000;19(8):623-6.
125. Tsao TC, Chen CH, Hong JH, Hsieh MJ, Tsao KC, Lee CH. Shifts of T4/T8 T lymphocytes from BAL fluid and peripheral blood by clinical grade in patients with pulmonary tuberculosis. *Chest*. 2002;122(4):1285-91.
126. Jadoon SM, Moin S, Ahmed TA, Bashir MM, Jadoon S. Smear-negative pulmonary tuberculosis and lymphocyte subsets. *J Coll Physicians Surg Pak.* 2004;14(7):419-22.
127. Davoudi S, Rasoolinegad M, Younesian M, Hajiabdolbaghi M, Soudbakhsh A, Jafari S, et al. CD4+ cell counts in patients with different clinical manifestations of tuberculosis. *Braz J Infect Dis*. 2008;12(6):483-6.
128. Fahey JL, Taylor JM, Detels R, Hofmann B, Melmed R, Nishanian P, et al. The prognostic value of cellular and serologic markers in infection with human immunodeficiency virus type 1. *The New England journal of medicine.* 1990;322(3):166-72.
129. Lang W, Perkins H, Anderson RE, Royce R, Jewell N, Winkelstein W, Jr. Patterns of T lymphocyte changes with human immunodeficiency virus infection: from seroconversion to the development of AIDS. *J Acquir Immune Defic Syndr*. 1989;2(1):63-9.
130. Wood R, Lawn SD. Challenges facing lipoarabinomannan urine antigen tests for diagnosing HIV-associated tuberculosis. *Expert Rev Mol Diagn.* 2012;12(6):549-51.
131. Cohen LA, Gutierrez L, Weiss A, Leichtmann-Bardoogo Y, Zhang DL, Crooks DR, et al. Serum ferritin is derived primarily from macrophages through a nonclassical secretory pathway. *Blood.* 2010;116(9):1574-84.
132. McKenzie SW, Means RT, Jr. Extreme hyperferritinemia in patients infected with human immunodeficiency virus is not a highly specific marker for disseminated histoplasmosis. *Clin Infect Dis*. 1997;24(3):519-20.
133. Arranz Caso JA, Garcia Tena J, Llorens MM, Moreno R. High serum ferritin concentration in an AIDS patient with miliary tuberculosis. *Clin Infect Dis*. 1997;25(5):1263-4.
134. Ayoade F, Cotelingam J, Joel Chandranesan AS. Disseminated *Mycobacterium avium-Intracellulare* Complex Infection Presenting With Disseminated Intravascular Coagulation in an AIDS Patient. *J Investig Med High Impact Case Rep.* 2017;5(4):2324709617740904.
135. Skogmar S, Schön T, Balcha TT, Sturegård E, Jansson M, Björkman P. Plasma Levels of Neopterin and C-Reactive Protein (CRP) in Tuberculosis (TB) with and without HIV Coinfection in Relation to CD4 Cell Count. *PLoS One*. 2015;10(12):e0144292.
136. Lawn SD, Rudolph D, Wiktor S, Coulibaly D, Ackah A, Lal RB. Tuberculosis (TB) and HIV infection are independently associated with elevated serum concentrations of tumour necrosis factor receptor type 1 and beta2-microglobulin, respectively. *Clin Exp Immunol.* 2000;122(1):79-84.
137. World Health Organization. *Global tuberculosis report 2018.*; 2018. Report No.: WHO/CDS/TB/2018.20.
138. Kollmann TR, Levy O, Montgomery RR, Goriely S. Innate immune function by Toll-like receptors: distinct responses in newborns and the elderly. *Immunity*. 2012;37(5):771-83.
139. Uniprot. UniProtKB - P00738 (HPT_HUMAN) 2019 [Available from: https://www.uniprot.org/uniprot/P00738] [accessed 24th July 2019].

140. Langlois MR, Delanghe JR. Biological and clinical significance of haptoglobin polymorphism in humans. *Clin Chem*. 1996;42(10):1589-600.

141. Nash D, Tymejczyk O, Gadisa T, Kulkarni SG, Hoffman S, Yigzaw M, et al. Factors associated with initiation of antiretroviral therapy in the advanced stages of HIV infection in six Ethiopian HIV clinics, 2012 to 2013. *J Int AIDS Soc*. 2016;19(1):20637.

142. Gupta-Wright A, Fielding K, Wilson D, van Oosterhout JJ, Grint D, Mwandumba HC, et al. Tuberculosis in hospitalised patients with HIV: clinical characteristics, mortality, and implications from the STAMP trial. *Clin Infect Dis*. 2019. Pii: ciz1133.

Chapter VII

1. Dye C, Glaziou P, Floyd K, Raviglione M. Prospects for tuberculosis elimination. *Annu Rev Public Health*. 2013;34:271-86.

2. Chegou NN, Sutherland JS, Malherbe S, Crampin AC, Corstjens PL, Geluk A, et al. Diagnostic performance of a seven-marker serum protein biosignature for the diagnosis of active TB disease in African primary healthcare clinic attendees with signs and symptoms suggestive of TB. *Thorax.* 2016; 71:785-794.

3. de Groote MA, Sterling DG, Hraha T, Russell TM, Green LS, Wall K, et al. Discovery and Validation of a Six-Marker Serum Protein Signature for the Diagnosis of Active Pulmonary Tuberculosis. *J Clin Microbiol*. 2017; 55:3057-3071.

4. Gliddon HD, Howes PD, Kaforou M, Levin M, Stevens MM. A nucleic acid strand displacement system for the multiplexed detection of tuberculosis-specific mRNA using quantum dots. *Nanoscale* 2016; 8:10087-95.

5. Maetzdorf J, McEwen G, Weiner J 3rd, Tian S, Lader E, Schriek U, et al. Concise gene signature for point-of-care classification of tuberculosis. *EMBO Mol Med*. 2016; 8:86-95.

6. Roe JK, Thomas N, Gil E, Best K, Tsaliki E, Morris-Jones S, et al. Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. *JCI Insight.* 2016; 1:e87238.

7. Laux da Costa L, Delcroix M, Dalla Costa ER, Prestes IV, Milano M, Francis SS, et al. A real-time PCR signature to discriminate between tuberculosis and other pulmonary diseases. *Tuberculosis (Edinb).* 2015 Jul;95(4):421-5.

8. Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. *PLoS Med.* 2019 Apr 23;16(4):e1002786.

9. Warsinske HC, Rao AM, Moreira FMF, Santos PCP, Liu AB, Scott M, et al. Assessment of Validity of a Blood-Based 3-Gene Signature Score for Progression and Diagnosis of Tuberculosis, Disease Severity, and Treatment Response. *JAMA Netw Open*. 2018 Oct 5;1(6):e183779.

10. Singhania A, Verma R, Graham CM, Lee J, Tran T, Richardson M, et al. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. *Nat Commun*. 2018 Jun 19;9(1):2308.

11. Weiner J 3rd, Parida SK, Maertzdorf J, Black GF, Repsilber D, Telaar A, et al. Biomarkers of inflammation, immunosuppression and stress with active disease are revealed by metabolomic profiling of tuberculosis patients. *PLoS One*. 2012;7(7):e40221.

12. Feng S, Du YQ, Zhang L, Zhang L, Feng RR, Liu SY. Analysis of serum metabolic profile by ultra-performance liquid chromatography-mass spectrometry for biomarkers discovery: application in a pilot study to discriminate patients with tuberculosis. *Chin Med J (Engl).* 2015 Jan 20;128(2):159-68.

13. Lau SK, Lee KC, Curreem SO, Chow WN, To KK, Hung IF, et al. Metabolomic Profiling of Plasma from Patients with Tuberculosis by Use of Untargeted Mass Spectrometry Reveals Novel Biomarkers for Diagnosis. *J Clin Microbiol*. 2015 Dec;53(12):3750-9.

14. Kedia K, Wendler JP, Baker ES, Burnum-Johnson KE, Jarsberg LG, Stratton KG, et al. Application of multiplexed ion mobility spectrometry towards the identification of host protein signatures of treatment effect in pulmonary tuberculosis. *Tuberculosis (Edinb)*. 2018 Sep;112:52-61.

15. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature.* 2010 Aug 19;466(7309):973-7.

16. Bloom CI, Graham CM, Berry MP, Wilkinson KA, Oni T, Rozakeas F, et al. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. *PLoS One*. 2012;7(10):e46191.

17. Thompson EG, Du Y, Malherbe ST, Shankar S, Braun J, Valvo J, et al. Host blood RNA signatures predict the outcome of tuberculosis treatment. *Tuberculosis (Edinb).* 2017 Dec;107:48-58.

18. Darboe F, Mbandi SK, Naidoo K, Yende-Zuma N, Lewis L, Thompson EG, et al. Detection of Tuberculosis Recurrence, Diagnosis and Treatment Response by a Blood Transcriptomic Risk Signature in HIV-Infected Persons on Antiretroviral Therapy. *Front Microbiol*. 2019 Jun 26;10:1441.

19. Mahapatra S, Hess AM, Johnson JL, Eisenach KD, DeGroote MA, Gitta P, et al. A metabolic biosignature of early response to anti-tuberculosis treatment. *BMC Infect Dis* 2014; 14: 53.

20. Penn-Nicholson A, Hraha T, Thompson EG, Sterling D, Mbandi SK, Wall KM, et al. Discovery and validation of a prognostic proteomic signature for tuberculosis progression: A prospective cohort study. *PLoS Med*. 2019 Apr 16;16(4):e1002781.

21. Cliff JM, Kaufmann SH, McShane H, van Helden P, O'Garra A. The human immune response to tuberculosis and its treatment: a view from the blood. *Immunol Rev.* 2015 Mar;264(1):88-102.

22. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet.* 2016 Jun 4;387(10035):2312-2322.

23. Blankley S, Graham CM, Levin J, Turner J, Berry MP, Bloom CI, et al. A 380-gene meta-signature of active tuberculosis compared with healthy controls. *Eur Respir J.* 2016 Jun;47(6):1873-6.

24. Duan L, Gan H, Arm J, Remold HG. Cytosolic phospholipase A2 participates with TNF-alpha in the induction of apoptosis of human macrophages infected with Mycobacterium tuberculosis H37Ra. *J Immunol.* 2001 Jun 15;166(12):7469-76.

25. Gebremicael G, Kassa D, Alemayehu Y, Gebreegziaxier A, Kassahun Y, van Baarle D, et al. Gene expression profiles classifying clinical stages of tuberculosis and monitoring treatment responses in Ethiopian HIV-negative and HIV-positive cohorts. *PLoS One*. 2019 Dec 10;14(12):e0226137.

26. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, et al. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med*. 2013;10(10):e1001538.

27. Tadokera R, Meintjes GA, Wilkinson KA, Skolimowska KH, Walker N, Friedland JS, et al. Matrix metalloproteinases and tissue damage in HIV-tuberculosis immune reconstitution inflammatory syndrome. *Eur J Immunol*. 2014 Jan;44(1):127-36.

28. Duffy FJ, Thompson EG, Scriba TJ, Zak DE. Multinomial modelling of TB/HIV co-infection yields a robust predictive signature and generates hypotheses about the HIV+TB+ disease state. *PLoS One*. 2019 Jul 15;14(7):e0219322.

29. Esmail H, Lai RP, Lesosky M, Wilkinson KA, Graham CM, Horswell S, et al. Complement pathway gene activation and rising circulating immune complexes characterize early disease in HIV-associated tuberculosis. *Proc Natl Acad Sci U S A*. 2018 Jan 30;115(5):E964-E973.

30. Silva CAM, Graham B, Webb K, Ashton LV, Harton M, Luetkemeyer AF, et al. A pilot metabolomics study of tuberculosis immune reconstitution inflammatory syndrome. *Int J Infect Dis*. 2019 Jul;84:30-38.

31. Dupnik KM, Bean JM, Lee MH, Jean Juste MA, Skrabanek L, Rivera V, et al. Blood transcriptomic markers of *Mycobacterium tuberculosis* load in sputum. *Int J Tuberc Lung Dis*. 2018 Aug 1;22(8):950-958.

32. Verma S, Du P, Nakanjako D, Hermans S, Briggs J, Nakiyingi L, et al. Tuberculosis in advanced HIV infection is associated with increased expression of IFNγ and its downstream targets. *BMC Infect Dis.* 2018 May 15;18(1):220.

APPENDICES

1. Introduction

Permission was obtained to re-use the following figures and images in the Introduction (Chapter I) that were under Copyright:

i) **Figure 1: Recent and predicted trends in global TB incidence (from Dye *et al*, '13);**

ii) **Figure 2: Examples of multiplexed lateral flow immunoassay (LFA) formats.** Left: detection of three serogroups of Shiga toxin-producing *Escherichia coli* by multiplexed LFA in series

iii) **Figure 6. Classification criteria used for cohort study by the African-European TB Consortium (AE-TBC, from Chegou *et al*, '16).**

The permissions for each are shown in turn. Fonts have been changed to Arial 10 for consistency, but otherwise the formatting has not been changed.

2. Correspondence

i) Figure 1: Recent and predicted trends in global TB incidence (from Dye et al, '13);

Order Date
29-Apr-2020
Order license ID
1031694-1
ISSN
1545-2093
Type of Use
Republish in a thesis/dissertation
Publisher
ANNUAL REVIEWS
Portion
Chart/graph/table/figure

LICENSED CONTENT

Publication Title
Annual review of public health
Date
01/01/1980
Language
English
Country
United States of America
Rightsholder
Annual Reviews, Inc.
Publication Type
e-Journal
URL
http://arjournals.annualreviews.org/loi/publhealth

REQUEST DETAILS

Portion Type
Chart/graph/table/figure
Number of charts / graphs / tables / figures requested
1

Format (select all that apply)
Print, Electronic
Who will republish the content?
Academic institution
Duration of Use
Life of current edition
Lifetime Unit Quantity
Up to 499
Rights Requested
Main product
Distribution
Worldwide
Translation
Original language of publication
Copies for the disabled?
Yes
Minor editing privileges?
No
Incidental promotional use?
No
Currency
GBP

NEW WORK DETAILS

Title
Host serum protein biomarkers of tuberculosis in sub-Saharan Africa
Instructor name
Dr Thomas C Morris
Institution name
Imperial College London
Expected presentation date
2020-04-30

ADDITIONAL DETAILS

Order reference number
N/A
The requesting person / organization to appear on the license
Dr Thomas Morris

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)
Figure 1 a)
Editor of portion(s)
Jonathan E. Fielding
Volume of serial or monograph
34
Page or page range of portion
272
Title of the article/chapter the portion is from
Prospects for Tuberculosis Elimination
Author of portion(s)
Christopher Dye, Philippe Glaziou, Katherine Floyd, and Mario Raviglione
Issue, if republishing an article from a serial
N/A
Publication date of portion
2012-12-14

ii) *Figure 2. Examples of multiplexed lateral flow immunoassay (LFA) formats. Left: detection of three serogroups of Shiga toxin-producing* Escherichia coli *by multiplexed LFA in series*

The following notification of acceptance was received by email:

ELSEVIER LICENSE

TERMS AND CONDITIONS
Dec 02, 2019

This Agreement between Dr. Thomas Morris ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4720991119693
License date	Dec 02, 2019
Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Microbiological Methods
Licensed Content Title	Development of a novel multiplex lateral flow assay using an antimicrobial peptide for the detection of Shiga toxin-producing Escherichia coli
Licensed Content Author	Taro Yonekita, Ryuji Ohtsuki, Eri Hojo, Naoki Morishita, Takashi Matsumoto, Tomoyasu Aizawa, Fumiki Morimatsu
Licensed Content Date	Jun 1, 2013
Licensed Content Volume	93
Licensed Content Issue	3
Licensed Content Pages	6
Start Page	251
End Page	256
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	Host serum protein biomarkers of tuberculosis in sub-Saharan Africa.
Institution name	Imperial College London
Expected presentation date	Dec 2019
Portions	Figure 2
Requestor Location	Dr. Thomas Morris 7 Kenwood Road Leicester, LE2 3PL United Kingdom Attn: Dr. Thomas Morris
Publisher Tax ID	GB 494 6272 12
Total	0.00 GBP

iii) Figure 6. Classification criteria used for cohort study by the African-European TB Consortium (AE-TBC, from Chegou et al, '16).

The following letter of acceptance was received by email:

BMJ PUBLISHING GROUP LTD. LICENSE

TERMS AND CONDITIONS

Dec 02, 2019

This Agreement between Dr. Thomas Morris ("You") and BMJ Publishing Group Ltd. ("BMJ Publishing Group Ltd.") consists of your license details and the terms and conditions provided by BMJ Publishing Group Ltd. and Copyright Clearance Center.

License Number

4717601114652

License date

Nov 28, 2019

Licensed Content Publisher

BMJ Publishing Group Ltd.

Licensed Content Publication

Thorax

Licensed Content Title

Diagnostic performance of a seven-marker serum protein bio-signature for the diagnosis of active TB disease in African primary healthcare clinic attendees with signs and symptoms suggestive of TB

Licensed Content Author

Novel N Chegou, Jayne S Sutherland, Stephanus Malherbe, Amelia C Crampin, Paul L A M Corstjens, Annemieke Geluk, Harriet Mayanja-Kizza, Andre G Loxton, Gian van der Spuy, Kim Stanley, Leigh A Kotzé, Marieta van der Vyver, Ida Rosenkrands, Martin Kidd, Paul D van Helden, Hazel M Dockrell, Tom H M Ottenhoff, Stefan H E Kaufmann, Gerhard Walzl

Licensed Content Date

Sep 1, 2016

Licensed Content Volume 71

Licensed Content Issue 9

Type of Use: Dissertation/Thesis

Requestor type: Individual

Format: Print and electronic

Portion: Figure/table/extract

Number of figure/table/extracts: 1

Description of figure/table/extracts: Table 1

Will you be translating? No

Circulation/distribution: 8

Title of your thesis / dissertation: Host serum protein biomarkers of tuberculosis in sub-Saharan Africa.

Expected completion date: Dec 2019

Estimated size(pages): 150

Requestor Location:

Dr. Thomas Morris

7 Kenwood Road

Leicester, LE2 3PL

United Kingdom

Attn: Dr. Thomas Morris

Publisher Tax ID: GB674738491

Total: 0.00 GBP