# When the Pathogen Hides: Using Host Biology to Diagnose Paediatric Tuberculosis




**Nathan Mudrak**

**St. Peter's College**

**University of Oxford**

**A thesis submitted for the degree of**

***Doctor of Philosophy in Paediatrics***

**Trinity 2026**

# Abstract

Tuberculosis (TB) is a substantial contributor to global childhood morbidity and mortality, yet current, pathogen-based diagnostics fail to detect many cases of paediatric TB. Host-based diagnostics, which measure disease-associated changes in the host's own biology, offer a promising alternative. Here, I evaluate various approaches to the host-based diagnosis of TB using samples from the Ugandan NOD-pedFEND cohort, a diagnostic cohort study which enrolled children under five with presumptive TB, along with a small number of healthy controls.

I first evaluated the prototype Cepheid Host Response Cartridge (HRC), which measures a transcriptomic signature in capillary blood. The HRC showed poor diagnostic performance, with an area under the curve (AUC) of between 0.55-0.56 for detecting microbiologically confirmed TB. No meaningful performance gains were seen using alternative scoring methodologies, longitudinal testing, or in subgroup analyses.

Next, I investigated sixteen published transcriptomic signatures of TB using a NanoString assay performed on RNA-stabilized whole blood. While top-performing signatures exhibited higher AUCs (maximum 0.70) compared to the HRC, diagnostic accuracy was still relatively poor. A bespoke signature developed from our data performed similarly, with a cross-validated AUC of 0.66. Exploratory analyses, however, suggested improved performance in several subgroups and when utilizing the change in signature scores after two weeks of treatment, although limited sample sizes warrant caution.

Finally, I evaluated the ability of a panel of 52 plasma protein biomarkers to diagnose paediatric TB. While previously reported signatures performed suboptimally (AUCs from

0.64-0.70), one protein—sICAM-1—performed well individually, with an AUC of 0.76 against an extended microbiological reference standard and a sensitivity of 53% when set to 91% specificity. Additionally, detectable levels of IL-6 universally decreased after two weeks of treatment in individuals with microbiologically confirmed, drug-sensitive TB (with IFN-γ performing similarly), suggesting potential utility for treatment monitoring or retrospective diagnosis.

Taken together, these results demonstrate both the challenges and the continued potential of host-based diagnostics for paediatric TB, identifying numerous avenues for future research.

# Acknowledgements

The work described herein simply would not have been possible without the support and concerted efforts of so many.

First, I want to thank our study participants, who graciously volunteered their time and energy to participate in our study, with minimal direct benefit to themselves. It is truly such a privilege to be entrusted by them to pursue this research.

Next, I must thank the entire NOD-pedFEND study team. To enrol 678 children into an intensive, longitudinal study despite numerous hurdles—including a global pandemic and an Ebola outbreak—is truly remarkable, and a testament to true grit. This study was a truly monumental undertaking, one which required the efforts of so many clinical, regulatory, and scientific staff, across the US, UK, Switzerland, and Uganda. To be a part of this global network has truly been such an honour.

To Dr. Rinn Song, thank you for granting me the opportunity to work on this study, in this fascinating field. I am forever inspired by the magnitude and ambition of the studies you have helped to establish.

To Dr. Nisreen Khambati, I'm so grateful to have embarked upon this DPhil journey alongside you. Your camaraderie throughout this experience has been so deeply appreciated; I will always treasure our cafeteria lunches or coffee and dinner meetups.

To Dr. Else Bijker, you were always someone I could rely upon. Thank you so much for agreeing to join my supervision team. I am so grateful to have been the recipient of your scientific insights and guidance throughout my DPhil, especially your lightning-fast turnarounds on reviewing a thesis chapter.

To Dr. Dominic Kelly, thank you for stepping up into a supervisory role you didn't necessarily expect, despite your busy schedule. From day one, you were always invested in my success, and you were critical to helping me get grounded at the Oxford Vaccine Group.

To Dr. Padmini Salgame, I will never forget how you welcomed me into your laboratory with open arms, eventually leading to two of my three thesis projects. Your support and guidance have been invaluable, and it has truly been such a privilege to learn from you. I couldn't have done this without you.

To Dr. Francesca Basile, I will be forever grateful for how you took me under your wing. From helping me transition to DPhil life during your time as interim PI to the invaluable mentorship you've given me as a supervisor, the last three years just wouldn't have been possible without you. It has also just been such a joy working with you—I will leave my time at Oxford with so many fond memories of the laughs and stories shared.

Dr. Laura Olbrich—it was such a pleasure working with you on our narrative review, and the numerous other projects ongoing. I always felt I could turn to you if I needed guidance, and just that availability has meant the world.

I am also deeply grateful to the Marshall Aid Commemoration Commission and the British government for supporting my studies at Oxford, a life-changing experience that would've otherwise been impossible. I am also grateful to the National Institutes of Health for funding this research, as well as to Cepheid for their donation of the Host Response Cartridges used for our testing.

I've also been blessed with a rich support network, filled with truly some of the most wonderful people. To members of the Salgame lab, especially Pam, Heidi, Vaishnavi, and Chris, I've had such a blast with you all. Thanks for the lunches, group costumes, and for always having my back. A lab experience is defined by who you work with, and you guys made it truly special.

Outside of work, I couldn't have done this without some amazing friends. Kyrolos, Kyra, Emma, Peter, Michael, Bea, Ricky, Leslie, Max, and everyone else I can't name—thank you for making my life so much richer, and for being the people I could always lean on when times got tough. I am inspired by you all, and I can't wait to see what you do next.

To Mom and Oma, everything I am and everything I do is because of you. I have been so incredibly lucky to be raised by two incredible people, role models who I genuinely look up to each and every day. Thank you for guiding me, trusting me when I want to make a leap, and for supporting me through all the ups and downs. To Ben and Kathryn, thank you for being the best siblings a guy could ask for. To my partner, Chris—you are my rock. Despite often lying an ocean away, you have stood by my side for this entire degree, and my life has been made so much better for it. I'll never forget how much it meant to have your support. I love you all.

It truly takes a village, and I am so deeply grateful to mine.

# Publications

The following authored manuscripts, related to the work presented herein, were published over the course of my DPhil:

- Laura Olbrich*, **Nathan Mudrak***, Ceri Evans, Padmini Salgame, Jerrold Ellner, Andrew Prendergast, Rinn Song, Untangling the complex relationship between HIV-exposure and tuberculosis in children: a narrative review, *The Lancet Global Health,* Volume 14, Issue 5, (2026), E817-E828.

- Nisreen Khambati*, Kanyange Angel Moureen*, Francesca Wanda Basile, Rudie Desravines, **Nathan Mudrak,** Rebecca Post, Emmanuel Nasinghe, Rutvi Upadhyay, Sandra Ruth Babirye, Stephannie Nabuduwa, Germine Nakayita, Farag Kakyama Luwambya, Malik Koire, Allen Nabisere, John Paul Lubega, Rose Nabirye, Margaretha de Vos, Adam Penn-Nicholson, Emily Douglass, Prossy Mbekeeka, Derek Armstrong, H Simon Schaaf, Megan Palmer, Eric Wobudeya, Soyeon Kim, Morten Ruhwald, David Alland, Susan E Dorman, Jerrold Ellner, Moses Joloba, Adeodata R Kekitiinwa, Else Margreet Bijker, Grace Paul Kisitu, Rinn Song, and the NOD-pedFEND Consortium, Oral Swab Testing With Xpert MTB/RIF Ultra for the Diagnosis of Tuberculosis in Children Aged <5 Years in Uganda: An Exploratory Interim Analysis of Diagnostic Accuracy in the NOD-pedFEND Cohort, *Open Forum Infectious Diseases,* Volume 12, Issue 4, (2025), ofaf206.

- Rinn Song*, Else Margreet Bijker*, Grace Kisitu*, Emily Douglass*, Emmanuel Nasinghe, Francesca Wanda Basile, Nisreen Khambati, Adam Penn-Nicholson, Morten Ruhwald, Soyeon Kim, **Nathan Mudrak,** Nandini Dendukuri, César Ugarte-Gil, Padmini Salgame, David Alland, Susan E. Dorman, Jerrold Ellner, Moses Joloba and Adeodata Kekitiinwa for the NOD-pedFEND consortium, Novel and optimized diagnostics for pediatric TB in endemic countries: NOD-pedFEND study protocol. *BMC Pediatrics,* Volume 25, Article 647 (2025).

* Denotes joint first authorship.

# Statement of Contributions

The work presented in this thesis originates from the NOD-pedFEND study, which was a large, multi-institutional undertaking, with many individuals contributing to the study's success. While I was one such individual helping to coordinate study logistics, many others were integral to its operations. In Uganda, these include Angel Kanyange and Sandra Babirye, who were critical for data entry and general study operations. Stephannie Nabuduwa was crucial for biorepository maintenance and sample shipments. Emmanuel Nasinghe led day-to-day management of the study in Uganda. Local supervision was provided by Dr. Moses Joloba (Makerere University) and Dr. Adeodata R. Kekitiinwa (Baylor College of Medicine Children's Foundation - Uganda), with support from Dr. Grace Kisitu (Baylor College of Medicine Children's Foundation - Uganda). Further, the clinical teams at Baylor Uganda Children's Clinical Centre of Excellence and the Nalufenya Children's Hospital were crucial to the study's success.

In the US, credit is due to the Principal Investigators of the parent FEND and NOD studies—Dr. Jerrold Ellner, Dr. David Alland, Dr. Padmini Salgame (all from Rutgers University), and Dr. Susan Dorman (Medical University of South Carolina)—for providing the funding, vision, and oversight necessary for the study to move forward. April Borkman provided exceptional coordination across the FEND studies, and the DOMC, specifically Rutvi Upadhyay, provided data management services. Additionally, Dr. Adam Penn-Nicholson, Dr. Margaretha De Vos (Foundation for Innovative New Diagnostics) and Dr. Derek Armstrong (Johns Hopkins University) provided critical insights, including on technology selection and scouting. Further, while all statistics presented in this thesis are

my own work, Frontier Science, particularly Dr. Soyeon Kim, Rebecca Post, and Rudie Desravines, provided invaluable statistical guidance for the study writ large. FEND was funded by the NIH under grant number #U01AI152084, and NOD under grant number #R01AI152159.

During my time here, the Oxford team, which led the NOD-pedFEND study, consisted of myself, Dr. Francesca Basile, Dr. Else Bijker, and Dr. Nisreen Khambati, and was formally supervised by Dr. Rinn Song. Credit is particularly due to Dr. Francesca Basile, who took on the role of interim Oxford PI for nearly a year. Dr. Nisreen Khambati played an integral role in study coordination and day-to-day clinical oversight. Oversight was also supported by Dr. Else Bijker, who provided valuable scientific guidance throughout the study.

Outside of these contributions to the study as a whole, there are also contributions specific to my thesis which should be highlighted.

For chapter one, the discussion of tuberculosis in children who are HIV-exposed but uninfected originates from an article published in *The Lancet Global Health* which I co-authored with Dr. Laura Olbrich (LMU Munich), who introduced me to the project. The writing and figures presented here on that topic originate from my sections of the manuscript, however, though her feedback on my work was certainly invaluable.

For chapter two, the NOD-pedFEND study procedures described were designed prior to my joining the team. Additionally, the algorithmic method for assigning case classifications was developed by myself and another DPhil student, Nisreen Khambati, with input from the remainder of the Oxford team. Further, cross-validation of our classification method

was provided by Frontier Science, specifically Rebecca Post and Rudie Desravines, in close collaboration with me and Dr. Khambati. Calculation of the “disease severity” variable was led by Dr. Francesca Basile and was also cross-validated by Frontier Science.

For chapter three, host response cartridge testing was performed on-site in Uganda by clinical staff, with results then analysed by me. Cepheid also played a core role through the donation of test cartridges and the loaning of an Xpert instrument, as well as in providing technical support for their technology.

For chapters four and five, laboratory work and analysis was conducted by me in the lab of Dr. Padmini Salgame, who also was the primary supervisor of the work. Brian Nyiro provided valuable training and initial oversight on the procedures used for the NanoString work (chapter four), which were harmonized with protocols and gene sets initially developed by Dr. Vaishnavi Kaipilyawar (Rutgers University) in the Salgame lab. SomaScan results were provided by SomaLogic, Inc., using samples I prepared.

Dr. Else Bijker, Dr. Francesca Basile, and Dr. Padmini Salgame provided invaluable feedback on this thesis. The contents presented herein have not been submitted for any other degree programme and, unless explicitly stated to the contrary, all work is my own.

# Table of Contents

*Abstract* ..... *2*

*Acknowledgements* ..... *4*

*Publications* ..... *7*

*Statement of Contributions* ..... *8*

*List of Tables* ..... *16*

*List of Figures* ..... *17*

*List of Abbreviations and Acronyms* ..... *19*

*Chapter 1: Introduction* ..... *21*

**1.1 Tuberculosis in Children: A Primer** ..... **21**

Tuberculosis—A Persistent Challenge ..... 21

Paediatric Tuberculosis ..... 21

Available Diagnostic Tests for Paediatric TB Disease ..... 23

Available Tests of TB Infection ..... 25

The Paediatric TB Diagnostic Dilemma ..... 27

**1.2 The Host Response to Paediatric Tuberculosis and its Promise for TB Diagnosis** ..... **30**

The Pathophysiology of Mtb Infection and the Host Immune Response ..... 30

**1.3 HIV-Exposure as a Critical Factor in the Host Response** ..... **33**

The Growing Challenge of HIV-Exposure ..... 33

TB Exposure in CHEU ..... 34

TB Infection in CHEU ..... 35

TB Disease in CHEU ..... 36

**1.4 Designing the Ideal Diagnostic** ..... **37**

**1.5 Conclusion** ..... **41**

*Chapter 2: Methods* ..... *43*

**2.1 Introduction** ..... **43**

**2.2 Design and Setting of the NOD-pedFEND Study** ..... **43**

Study Setup and Enrolment ..... 43

Description of Study Visits ..... 45

Healthy Controls ..... 47

Sample Collection ..... 48

**2.3 Ethical considerations** ..... **49**

**2.4 Study Monitoring and Data Management** ..... **50**

**2.5 Case Classification Approach** ..... **51**

**2.6 Reference Standard Definitions ..... 53**

**2.7 Additional Definitions ..... 56**

**2.8 Statistical Approaches ..... 58**

**2.9 NOD-pedFEND Cohort Characteristics ..... 59**

***Chapter 3: Performance of the Host-Response Cartridge in the NOD-pedFEND Cohort ..... 64***

**Introduction ..... 64**

**3.1 Principle of the Host Response Cartridge ..... 64**

**3.2 Reported Performance of the HRC ..... 66**

HRC in Paediatric TB ..... 67

Use as a Treatment Monitoring Tool ..... 68

**3.3 Methods ..... 69**

HRC Testing ..... 69

Data Cleaning ..... 70

Data Analysis ..... 71

**Results ..... 73**

**3.4 The HRC Cohort ..... 73**

**3.5 Diagnostic Performance of HRC Testing Performed at Baseline ..... 75**

Median HRC Scores ..... 77

Diagnostic Performance of the HRC ..... 79

Subgroup-Specific Performance ..... 82

**3.6 Longitudinal Changes to the HRC Score after TB Treatment ..... 88**

Longitudinal Changes as a Retrospective Diagnostic ..... 93

**3.7 Correlation Between HRC Scoring Approaches ..... 93**

**3.8 Alternative Scoring Methodologies ..... 95**

Diagnostic Performance—Individual Genes ..... 95

Diagnostic Performance—Pairwise Scoring ..... 99

Diagnostic Performance—Longitudinal Changes in Alternative Scoring Methods ..... 100

Unsupervised clustering with t-SNE ..... 101

**3.9 Discussion ..... 102**

Discrepancies in HRC Performance across Paediatric Studies ..... 102

Subgroup-Specific Performance ..... 106

Validating HRC Technology ..... 107

Longitudinal Changes in TB Scores ..... 108

Alternative Scoring Methodologies ..... 109

**3.10 Conclusion ..... 110**

*Chapter 4: Validation of Reported RNA Signatures of the Host Response to Tuberculosis in the NOD-pedFEND Cohort ..... 111*

**Introduction ..... 111**

**4.1 Reported Transcriptomic Signatures of TB ..... 111**
Signatures of TB Disease ..... 116
Signatures of TB in Individuals Living With HIV ..... 117
Prognostic Signatures for TB ..... 118

**4.2 Use of NanoString nCounter Technology for the Validation of RNA Signatures ..... 121**
Principles of the NanoString nCounter System ..... 122

**4.3 Approaches to Signature Development and Evaluation ..... 123**
Machine Learning Models for Signature Development and Scoring ..... 124
Statistical Methods for TB Signature Discovery ..... 125
Prevention of Overfitting ..... 126
Gene Set Scoring Methods ..... 126
Experimental Aims ..... 127

**4.4 Methods ..... 128**
Sample Selection ..... 128
RNA Extraction ..... 130
NanoString nCounter Measurements and Data Processing ..... 131
Signature Scoring and Analysis ..... 132
Derivation of a Novel Signature ..... 133

**Results ..... 134**

**4.5 Overall Performance of RNA Extraction and the NanoString Cartridge ..... 134**
Sample Characteristics ..... 134

**4.6 Clustering Analysis of Baseline Gene Expression ..... 140**

**4.7 Performance of Previously Reported RNA Signatures at Baseline ..... 143**
Distribution of Signature Scores ..... 143
Diagnostic Performance of TB Signatures ..... 146
Subgroup Analysis ..... 151

**4.8 Ability of Longitudinal Changes to Signature Scores to Retrospectively Identify TB Disease ..... 154**

**4.9 Development of a Novel Transcriptomic Signature from Assessed Genes ..... 158**

**4.10 Correlation Between NanoString- and HRC-Derived TB Signature Scores ..... 161**

**4.11 Discussion ..... 165**
Subgroup-Specific Performance of Signatures ..... 166
Influence of Low-Expression Genes ..... 168
Comparison with HRC ..... 169
Novel Signature Development ..... 171

Future Work and Next Steps ..... 172
A Note on Priorities and Experimental Design ..... 173

**4.12 Conclusion ..... 175**

***Chapter 5: Proteomic Signatures of Paediatric TB in Blood Plasma ..... 177***

**Introduction ..... 177**

**5.1 Reported Proteomic Signatures of Tuberculosis ..... 177**
Approaches to Proteomic Signature Development ..... 177
Existing Proteomic Signatures of Paediatric TB ..... 178

**5.2 Meso Scale Discovery V-Plex Assays for the Validation of Proteomic Signatures ..... 181**

**5.3 Methods ..... 183**
Sample Selection ..... 183
Plasma preparation and MSD V-Plex Protocol ..... 183
Data Processing ..... 185
Statistical Analysis and Machine Learning ..... 187

**Results ..... 190**

**5.4 Performance of Meso Scale Assays in NOD-pedFEND Samples ..... 190**
Sample Characteristics ..... 190
Technical Validation of the MSD Assay ..... 191

**5.5 Differences in Individual Plasma Biomarkers in Children with TB ..... 196**

**5.6 Diagnostic Accuracy of an Ensemble Model Composed of All Tested Proteins ..... 202**

**5.7 Diagnostic Performance of Previously Reported Proteomic Signatures of TB ..... 206**

**5.8 Subgroup-Specific Performance of Proteomic Signatures ..... 207**

**5.9 The Use of Longitudinal Changes in Biomarker Levels to Assist in Retrospective TB Diagnosis ..... 210**

**5.10 Confirmation of sICAM-1 Performance using SomaScan Data ..... 215**

**5.11 Discussion ..... 217**
sICAM-1 ..... 218
Biological Significance of (s)ICAM-1 ..... 219
Longitudinal and Subgroup-Specific Performance of sICAM-1 ..... 220
Diagnostic Performance of CRP ..... 221
Longitudinal Changes in Protein Biomarkers May Differentiate TB Statuses ..... 222
Limitations ..... 224

**5.12 Conclusion ..... 225**

***Chapter 6: Discussion and Future Directions ..... 227***

**6.1 Paediatric TB—An Ongoing Challenge ..... 227**

**6.2 Summary of Work Presented ..... 228**
Performance of the Host Response Cartridge ..... 228
Transcriptomic Signatures of TB ..... 229
Proteomic Biomarkers of TB ..... 231
In Summary ..... 232
**6.3 Limitations of the Work and Continued Challenges for the Field ..... 233**
Low Rates of Microbiological Confirmation ..... 233
The Proportion of Children with Unconfirmed TB ..... 234
Imperfect Reference Standards ..... 237
Geographic Limitations ..... 239
**6.4 Future Directions ..... 240**
Suggested Next Steps for the Field: Improving the Ground Truth Used for Diagnostic Studies ..... 240
Additional Paths Worth Exploring ..... 242
In Summary ..... 242
***Chapter 7: References ..... 244***
***Appendices ..... 257***

# List of Tables

Chapter 1, Table 1 ..... 29
Chapter 1, Table 2 ..... 40
Chapter 1, Table 3 ..... 40
Chapter 2, Table 1 ..... 60
Chapter 3, Table 1 ..... 76
Chapter 3, Table 2 ..... 82
Chapter 3, Table 3 ..... 84
Chapter 3, Table 4 ..... 89
Chapter 3, Table 5 ..... 93
Chapter 3, Table 6 ..... 96
Chapter 3, Table 7 ..... 99
Chapter 3, Table 8 ..... 100
Chapter 3, Table 9 ..... 101
Chapter 3, Table 10 ..... 104
Chapter 4, Table 1 ..... 113
Chapter 4, Table 2 ..... 136
Chapter 4, Table 3 ..... 138
Chapter 4, Table 4 ..... 148
Chapter 4, Table 5 ..... 150
Chapter 4, Table 6 ..... 151
Chapter 4, Table 7 ..... 158
Chapter 4, Table 8 ..... 165
Chapter 5, Table 1 ..... 179
Chapter 5, Table 2 ..... 184
Chapter 5, Table 3 ..... 190
Chapter 5, Table 4 ..... 195
Chapter 5, Table 5 ..... 203
Chapter 5, Table 6 ..... 207
Chapter 5, Table 7 ..... 208
Appendix 1 ..... 257
Appendix 2 ..... 258
Appendix 3 ..... 261

# List of Figures

Chapter 1, Figure 1 ..... 35
Chapter 2, Figure 1 ..... 47
Chapter 2, Figure 2 ..... 48
Chapter 2, Figure 3 ..... 56
Chapter 3, Figure 1 ..... 74
Chapter 3, Figure 2 ..... 78
Chapter 3, Figure 3 ..... 80
Chapter 3, Figure 4 ..... 85
Chapter 3, Figure 5 ..... 87
Chapter 3, Figure 6 ..... 88
Chapter 3, Figure 7 ..... 90
Chapter 3, Figure 8 ..... 91
Chapter 3, Figure 9 ..... 92
Chapter 3, Figure 10 ..... 94
Chapter 3, Figure 11 ..... 97
Chapter 3, Figure 12 ..... 98
Chapter 3, Figure 13 ..... 102
Chapter 4, Figure 1 ..... 123
Chapter 4, Figure 2 ..... 130
Chapter 4, Figure 3 ..... 141
Chapter 4, Figure 4 ..... 141
Chapter 4, Figure 5 ..... 143
Chapter 4, Figure 6 ..... 145
Chapter 4, Figure 7 ..... 147
Chapter 4, Figure 8 ..... 149
Chapter 4, Figure 9 ..... 153
Chapter 4, Figure 10 ..... 154
Chapter 4, Figure 11 ..... 156
Chapter 4, Figure 12 ..... 157
Chapter 4, Figure 13 ..... 160
Chapter 4, Figure 14 ..... 162
Chapter 4, Figure 15 ..... 163
Chapter 4, Figure 16 ..... 164
Chapter 5, Figure 1 ..... 189
Chapter 5, Figure 2 ..... 193

Chapter 5, Figure 3 ..... 199
Chapter 5, Figure 4 ..... 200
Chapter 5, Figure 5 ..... 201
Chapter 5, Figure 6 ..... 204
Chapter 5, Figure 7 ..... 205
Chapter 5, Figure 8 ..... 209
Chapter 5, Figure 9 ..... 214
Chapter 5, Figure 10 ..... 216
Chapter 5, Figure 11 ..... 217

# List of Abbreviations and Acronyms

| | |
|---|---|
| ACS | Adolescent cohort study |
| AF | Aetiologic fraction |
| AI | Artificial intelligence |
| AIDS | Acquired immunodeficiency syndrome |
| ANOVA | Analysis of variance |
| AUC | Area under the curve |
| BCG | Bacillus Calmette-Guérin |
| BH | Benjamini-Hochberg |
| BMI | Body mass index |
| CHEU | Children who are HIV-exposed but uninfected |
| CLHIV | Children living with HIV |
| CHU | Children who are HIV-unexposed |
| CI | Confidence interval |
| CRP | C-reactive protein |
| CRS | Composite reference standard |
| CSF | Cerebrospinal fluid |
| Ct | Cycle threshold |
| CV | Coefficient of variation |
| CXR | Chest X-ray |
| DERA | Deoxyribose-phosphate aldolase |
| DRF | Distributed random forest |
| DUSP3 | Dual specificity phosphatase 3 |
| EDTA | Ethylenediaminetetraacetic acid |
| ELISA | Enzyme-linked immunosorbent assays |
| EPTB | Extrapulmonary tuberculosis |
| FEND | Feasibility of Novel Diagnostics for TB in Endemic Countries |
| GA | Gastric aspirate |
| GBM | Gradient boosting machine |
| GBP5 | Guanylate binding protein 5 |
| GCP | Good clinical practice |
| GSVA | Gene set variation analysis |
| HC | Healthy control |
| HHC | Household contact |
| HIV | Human immunodeficiency virus |
| HRC | Host response cartridge |
| IGRA | Interferon-gamma release assay |
| IQR | Interquartile range |
| KLF2 | Krüppel-like factor 2 |
| LAM | Lipoarabinomannan |
| LDA | Linear discriminant analysis |
| LJ | Lowenstein-Jensen |

| | |
|---|---|
| LLOQ | Lower limit of quantification |
| MAM | Moderate acute malnutrition |
| MGIT | Mycobacterial growth indicator tube |
| MRS | Microbiological reference standard |
| MSD | Meso Scale Discovery |
| Mtb | Mycobacterium tuberculosis |
| MUAC | Mid-upper arm circumference |
| NAAT | Nucleic acid amplification test |
| NIH | National Institutes of Health |
| NOD | Novel and Optimized Diagnostics for Paediatric TB |
| NPA | Nasopharyngeal aspirate |
| NTM | Nontuberculous mycobacteria |
| OD | Other disease |
| PBMC | Peripheral blood mononuclear cell |
| PCA | Principal component analysis |
| PLAGE | Pathway level analysis of gene expression |
| POC | Point of care |
| PPIA | Peptidylprolyl isomerase A |
| QC | Quality control |
| QFT | QuantiFERON-TB |
| qRT-PCR | Quantitative reverse transcription polymerase chain reaction |
| RFU | Relative fluorescent unit |
| RIPE | Rifampin, isoniazid, pyrazinamide, and ethambutol |
| ROC | Receiver operating characteristic |
| SAM | Severe acute malnutrition |
| SRS | Strict reference standard |
| STARD | Standards for reporting of diagnostic accuracy studies |
| SVM | Support vector machine |
| TB | Tuberculosis |
| TBD | TB disease |
| TBI | Tuberculosis infection |
| TBP | TATA-box binding protein |
| TDA | Treatment decision algorithm |
| TPP | Target product profile |
| TPT | Tuberculosis preventive therapy |
| t-SNE | t-distributed stochastic neighbour embedding |
| TST | Tuberculin skin test |
| WHO | World Health Organization |
| XRT | Extremely randomized tree |

# Chapter 1: Introduction

## 1.1 Tuberculosis in Children: A Primer

*Tuberculosis—A Persistent Challenge*

Tuberculosis (TB) continues to be the world's leading cause of death from infectious disease, killing an estimated 1.23 million people in 2024 alone [1]. This is despite the fact that the causative pathogen of TB—*Mycobacterium tuberculosis (Mtb)*—has been known for nearly 150 years, with antibiotic treatment for the disease first trialled in the mid-1940s [2]. Now, with an assortment of antibiotics available, all but the most extremely complex or drug-resistant TB cases should be curable, if treated early. Nevertheless, TB remains a global killer, particularly in low- and middle-income countries [3]. A key contributor to TB's persistent global mortality is the undertreatment of individuals with TB disease, which is frequently the result of missed diagnoses. Roughly 25% of adults and 50% of children with TB are never identified as having TB disease [4]. As a result, better identifying these individuals, and treating them accordingly, is crucial to lowering TB's global burden.

*Paediatric Tuberculosis*

The challenge of TB in children is particularly grim, with 174,000 children less than 15 years old having died from TB in the most recent year of reporting (2024) [1]. It is estimated that 96% of all children who die from TB were never given TB treatment, a sobering statistic largely driven by the 50% of children with TB who are never diagnosed with the disease [4]. While there are certainly additional contributors to the challenge of TB undertreatment in children, including the insufficient supply of antibiotics in endemic regions [5], improving

paediatric TB diagnosis is a clear approach to improving rates of TB treatment, and thereby reducing TB mortality. In children, however, diagnosing TB is notoriously difficult, a reflection of its paucibacillary nature and frequently non-specific clinical presentation.

While paediatric TB can develop months to years after initial exposure to the pathogen, the highest risk of disease occurs in the immediate months following infection, particularly in young children. In fact, a 2020 meta-analysis found 83% of cases in children under five to develop within 90 days of their initial work-up post-TB exposure [6]. Once developed, paediatric TB disease often presents with a variety of non-specific symptoms such as cough, fever, and malaise. Additionally, given TB's often delayed onset and chronic, "slow-burning" presentation (although the disease tends to develop more quickly in younger children), a child with TB can yield a fairly broad differential diagnosis, which can delay the ultimate identification and treatment of TB. Families affected by TB also often experience stigma, which can discourage them from seeking medical attention for their child, particularly in the early stages of disease [7]. However, once developed, and if untreated, TB disease is fatal in an estimated 44% of children under five, as it progresses to life-threatening complications such as meningitis, miliary TB, or respiratory failure [8, 9]. As a result, diagnosing the disease in its earlier stages, when it is most treatable, is critical. However, these are also the stages in which the disease may be least detectable, as the bacteria have yet to yield more specific symptomatology, radiographic indications of disease, or to replicate to a sufficient level to be detectable by microbiological testing.

*Available Diagnostic Tests for Paediatric TB Disease*

Given the frequently paucibacillary manifestation of paediatric TB, it is thought that less than a third of TB cases in children will be detected using an existing lab test [10]. As a result, the TB status of many children is not known with certainty. As such, the precise performance of TB diagnostic tests is difficult to quantify, as there is no absolute ground truth against which test results can be compared. Nevertheless, estimates of diagnostic performance have been made, using a variety of reference standards and modelling techniques.

For example, sputum smear microscopy, the traditional laboratory test for TB, is known to have particularly low sensitivity in children, estimated to be around 0.5% [11]. Moreover, as young children typically have difficulty spontaneously expectorating sputum, a sample often cannot even be obtained. Culture-based diagnostic methods, in contrast, have long been considered the gold standard of paediatric TB diagnosis. Using a solid- or liquid-based culture media, such as Löwenstein-Jensen (LJ) or Mycobacterial Growth Indicator Tube (MGIT) media, respectively, a variety of specimens can be tested for TB, including gastric aspirate (GA), nasopharyngeal aspirate (NPA), sputum, or lymph node aspirates (in cases of extrapulmonary TB). However, even these approaches have limited diagnostic performance; on average, they are estimated to detect only 24% of all clinically diagnosed paediatric TB cases [10]. Newer nucleic acid amplification tests (NAATs), such as those run on the Xpert platform (Cepheid; Sunnyvale, CA, USA), however, have recently replaced culture as the World Health Organization (WHO)-recommended first-line diagnostic of choice for paediatric TB [12]. It should be noted, though, that this change was primarily

motivated by the improved time-to-result and reduced resource requirements of NAATs. The sensitivities of culture and the highest-sensitivity NAAT—Xpert MTB/RIF Ultra (Ultra)—are likely similar [13]. This is difficult to definitively conclude, however, as the high specificities of both testing approaches make it unclear which test is incorrect when there is a discrepancy in results between the two approaches.

For children living with Human Immunodeficiency Virus (HIV), testing for lipoarabinomannan (LAM) is also frequently performed to assist in TB diagnosis, as the performance of culture and NAAT is typically lower in those living with HIV [14]. LAM is a glycolipid which forms a component of the cell wall in mycobacteria species, including *Mtb*. To date, only one LAM test is commercially available: a lateral-flow test marketed as Determine LAM (Abbott Laboratories; Abbott Park, Illinois, USA), formerly known as Alere Determine TB LAM. The test, which detects LAM present in urine samples from individuals with TB, was endorsed by the WHO in 2019 for the testing of children living with HIV (CLHIV), following data which showed the test's elevated sensitivity in HIV-positive adults, results which were generalized to—and later confirmed in—children [15].

Chest radiography also plays a key role in the diagnosis of TB, particularly in the absence of microbiological confirmation of disease. While some findings on chest x-rays (CXR)—such as hilar lymphadenopathy, airway compression, or a miliary pattern—are highly suggestive of TB disease, particularly in the appropriate clinical context, these are present in a minority of TB cases. For the majority of children with TB, chest radiographs will demonstrate non-specific findings such as effusion or consolidation, or even no abnormal findings at all [16]. Further, while artificial intelligence (AI)-based tools for CXR

interpretation are rapidly advancing, CXRs currently rely on human interpretation, which is highly variable [17]. Additionally, CXRs require radiography equipment and a trained reader, both of which can be in short supply in settings with high TB burdens [18].

Similarly, resource constraints can limit the accessibility of microbiological testing. Last year, only 8 out of the 30 countries with the highest TB burden reported that more than 50% of their TB diagnostic sites had access to the WHO-recommended first-line NAATs [1]. Globally, only 54% of all individuals (including adults) newly diagnosed with TB in 2024 received a first-line NAAT as their initial test, as advised by WHO guidelines [1, 12]. The existing state of paediatric TB diagnostics, then, is hallmarked by generally poorly performing tests, which are inadequately accessible to individuals with TB disease.

*Available Tests of TB Infection*

An alternative approach to combatting TB is to identify and treat it during the "TB infection" stage, after initial exposure to the pathogen but before it develops into disease. While tests of TB infection exist, they also face challenges.

Tests for TB infection can be broadly split into two categories: the tuberculin skin test (TST) and interferon-gamma release assays (IGRAs), such as QuantiFERON-TB Gold Plus (QIAGEN; Venlo, The Netherlands) and T-SPOT.TB (Revvity; Massachusetts, USA). The TST consists of a suspension of tuberculin (a mix of several *Mtb* antigens), which is injected intradermally into the forearm, with any resultant induration, the result of a type IV hypersensitivity reaction, quantified two to three days later. The measured induration is compared to established thresholds, which can be adjusted to account for the pre-test

probability of TB infection, weakened immune responses, or the likelihood of TB progressing to disease (particularly if used for prophylactic treatment). If the measured induration exceeds the threshold, the test is positive for TB infection. The TST is reported to identify infection in between 64-86% of children with TB disease [19]. The TST, however, is constrained by the need for a follow-up visit to measure the induced induration, as well as its reliance upon human measurement of induration, which has been shown to be variable [20].

An alternative, newer class of tests for TB infection are the IGRAs, which measure the release of IFN-γ from a blood sample stimulated with *Mtb* antigens. These tests remove the requirement for a follow-up visit or the error-prone measurement of induration and are believed to have slightly higher sensitivity than TSTs, at about 84% in individuals with TB disease overall [21]. However, IGRAs are generally thought to be inferior to TST in children under five years old, with only 45% of TST+ children under five also found to be positive on an IGRA, according to a recent systematic review [22].

Moreover, the information gained from any test of TB infection is limited in many endemic regions, given the existing prevalence of TB there. In urban southern Uganda, for instance, tests of TB infection are positive in about a third of individuals present in high-TB-burden communities there, despite testing occurring irrespective of prior exposure history [23]. Despite testing positive, though, most of these community members will not immediately develop TB disease [24]. In fact, even globally, around a quarter of the world's population is estimated to be infected with TB (of which roughly 6% are children) [25, 26]. Given the cost of TB infection testing, along with the current impracticality of treating all

individuals positive for infection, TB cannot currently be combatted through a strategy solely reliant on testing and treating TB infections. As such, tests that identify individuals in greatest need of TB treatment—those with TB disease—are critical.

*The Paediatric TB Diagnostic Dilemma*

Given the low sensitivity and limited availability of diagnostic tests for TB disease, then, many children's TB diagnosis must be made on clinical grounds alone. This reliance on clinical expertise—access to which also varies—can lead to missed opportunities for TB treatment if a TB diagnosis is missed. Conversely, it can also come at the cost of overtreatment of children without true TB disease, a caution that reflects the often-fatal consequences of a missed TB diagnosis [8]. Additionally, overtreatment further reduces availability of an already limited medication supply. Moreover, TB treatment is not risk-free; potential side effects of even the standard rifampin, isoniazid, pyrazinamide, and ethambutol (RIPE) regimen include peripheral neuropathy, retrobulbar neuritis, and joint pain [27]. As such, improving a treating clinician's ability to distinguish true TB cases from other diseases with similar symptom profiles would substantially benefit not just their own patients but also global TB control efforts.

Recent developments have demonstrated that improved methods of diagnosis are possible, however. The ability to perform NAATs on new sample types—particularly nasopharyngeal aspirates and stool— has made molecular testing more accessible to the large proportion of children under five who are unable to spontaneously expectorate sputum [28]. Further, a new platform for NAATs, Truenat (Molbio Diagnostics; Verna, Goa, India), has been endorsed by the WHO for use in children, which may increase test

accessibility [12]. Finally, updated WHO recommendations for concurrent TB testing in children—in which NAAT testing of stool is performed along with NAAT testing of a respiratory sample and a LAM test, if indicated—have also demonstrated improved TB detection rates [29, 30]. With this concurrent testing approach, in which a child is considered positive for TB if any of the 2-3 tests performed are positive, sensitivity for detecting paediatric TB disease is estimated to reach 30% and 29%, with a specificity of 83% and 99%, in children living with and without HIV, respectively [31]. While this diagnostic performance represents an improvement from culture's estimated sensitivity of 24%, it is simultaneously clear that the performance of diagnostic tests for paediatric TB (summarized in **Table 1**) remains suboptimal.

**Table 1.** An overview of the classes of laboratory tests commonly used for the diagnosis of paediatric TB, along with their reported diagnostic performance and the specimen types for which they are endorsed by the WHO. Tests used exclusively for the determination of drug resistance are not included. Extrapulmonary TB (EPTB) samples consist of lymph node aspirate or biopsy; pleural, peritoneal, pericardial, synovial, or cerebrospinal fluid; blood (in CLHIV); or urine specimens. Abbreviations: Cerebrospinal Fluid (CSF), Gastric Aspirate (GA), Nasopharyngeal Aspirate (NPA), Nucleic Acid Amplification Test (NAAT), Löwenstein-Jensen (LJ), Mycobacterial Growth Indicator Tube (MGIT).

| | **Commonly Used Approaches/Technologies** | **Specimen Types** | **Estimated Sensitivity/Specificity** | **Additional Notes** |
|---|---|---|---|---|
| Smear Microscopy | Acid-fast staining under light microscopy, or fluorochrome staining with fluorescence microscopy. | Most commonly, sputum (induced or spontaneous); otherwise, GA, NPA, and all EPTB samples besides blood and CSF. | ~7%/>90%; 1% sensitivity in children <5 years old [11, 32] | Cannot differentiate between non-tuberculosis mycobacteria (NTMs) and *Mtb* |
| *Mtb* Culture | Solid-based culture (on LJ or Middlebrook media) or liquid-based culture (using MGIT or similar technologies) | | ~24%/approaching ~100% [10] | NAATs are now generally preferred, in lieu of culture or smear microscopy |
| Low complexity NAATs (LC-NAATs) | Xpert or Truenat platforms, others under development | For both: sputum (induced or spontaneous). For Xpert only: GA, NPA, stool, and all EPTB samples. | With a concurrent testing approach, ~30%/~90%. Performance can vary depending on sample type/population. | Higher complexity NAATs are also available, but not outside of specialized laboratories. |
| Chest X-Rays (CXRs) | Portable and non-portable options available. AI for CXR interpretation under development. | N/A | ~45%/93%, though the balance between sensitivity and specificity can be adjusted [33]. | Prone to substantial inter-reader variability in interpretation [17]. Lateral views increase sensitivity by ~15% |

## 1.2 The Host Response to Paediatric Tuberculosis and its Promise for TB Diagnosis

Due to the paucibacillary nature of paediatric TB, diagnostics which rely upon detection of the pathogen itself—whether directly (as in culture), or indirectly, as when detecting LAM or nucleic acids from TB—are inherently challenged by the low bacterial load. However, the systemic immune response to the pathogen's presence can display larger, detectable modulations in response to infection. As such, it represents a promising target for next-generation TB diagnostics: technologies which capitalize upon the host response to *Mtb* to assist in its diagnosis.

*The Pathophysiology of Mtb Infection and the Host Immune Response*

Excluding congenital TB infections, children are exposed to tuberculosis through infected aerosols, which are inhaled into the alveoli [34]. Once settled in the lung parenchyma, bacterial replication is accompanied by a localized inflammatory response wherein alveolar macrophages migrate into the lung and phagocytose bacilli, with neutrophils recruited in an attempt to surround and contain the infection, forming the early stages of what is known as a Ghon focus. Simultaneously, bacilli and dendritic cells carrying *Mtb* antigens drain into adjacent lymph nodes, with antigen presentation in the nodes activating an acquired immune response. Combined, the Ghon focus and impacted lymph nodes can create a radiographically detectable lesion known as the Ghon complex [35].

Entrance of *Mtb* into the lymphatics can have numerous eventual outcomes, including the initiation of disseminated disease, such as miliary TB or TB meningitis, sufficient induction of acquired immunity to contain disease within only the lymphatic and pulmonary systems (pulmonary TB), further control of infection to prevent disease while retaining chronic infection (TB infection), to even bacterial clearance without treatment [9]. Further, the boundary between disease states can be blurred, with progression or regression thought to be determined by the host immune response. Ultimately, disease outcomes lie along a spectrum from fully disseminated disease to full immune control and clearance of infection, making strict classification of disease status difficult.

Intriguingly, how TB manifests along this spectrum seems to be strongly influenced by age, with children under five more likely to experience disseminated disease after infection but less likely to experience what is considered “adult-type” cavitary pulmonary disease, instead developing a more paucibacillary presentation. In contrast, children age 5-10 years have been observed to have a lowered likelihood of either manifestation of TB disease, while those over 10 are more apt to develop the aforementioned “adult-type” cavitary disease [36]. While other variables such as the level of TB exposure and environmental influences have been investigated as potential contributors to these changing manifestations of disease, it is widely thought that immune system maturation largely contributes to this age-specific epidemiology [9].

Returning to the developing Ghon complex, it is clear that the immune response influences disease presentation, though specific correlates of protection remain unknown. In adults, however, a robust Th1 immune response is thought to further encourage

granuloma formation by recruiting additional macrophages and neutrophils, the latter of which may undergo necrosis to form a niche within which *Mtb* can persist [9]. In contrast, children under five, whose maturing immune systems are biased towards a Th2 and Treg T-cell phenotype, are thought to contain the bacteria less vigorously, leading to the higher likelihood of disseminated disease [9]. However, if the infection is contained, weaker inflammatory responses also attenuate the caseous necrosis characteristic of the adult anti-TB response, denying *Mtb* the niche it can use to replicate. This would contribute to young children's lowered odds for cavitation formation and shift disease towards a more paucibacillary presentation, if it so develops. With children between five and ten at lowered risk of both these outcomes, they are thought to be in a "Goldilocks" immune state, with an immune response sufficiently vigorous to contain infection but not so vigorous as to create the niches in which *Mtb* can persist [36].

Given that many host biomarkers commonly studied, such as IFN-γ or IL-4, are classically associated with either a Th1, Th2, or other immune phenotype, differences in the broad immune landscape in children versus adults may directly translate to inferior diagnostic performance of biomarkers developed in adults. Host-based diagnostics, then, must either be sufficiently general to measure an aspect of the host response to *Mtb* which is conserved across all age groups, or must test for distinct responses in distinct ages. Other covariates may also need to be considered, including sex, disease presentation, and co-infections, which have each been hypothesized to alter host immunity *[9].* However, with the vast majority of TB cases likely to induce some measurable immune response,

host-based diagnostics are a promising avenue for diagnosing paucibacillary cases of TB disease, where current methods for directly detecting the pathogen itself are inadequate.

## 1.3 HIV-Exposure as a Critical Factor in the Host Response

The host response to *Mtb* can also be influenced by environmental factors, including those present during early immune development. HIV exposure is one such factor which may need to be accounted for in the development of host-based diagnostics. While HIV is often considered in the context of HIV infection, it is not as often considered for children who are HIV exposed but uninfected, despite evidence suggesting that they also may experience immune alterations. Given this, what follows is a short overview of how HIV exposure can influence the host response to *Mtb*.

*The Growing Challenge of HIV-Exposure*

With the swift roll-out of antiretroviral prophylaxis for preventing the vertical transmission of HIV, there is a growing number of children for whom HIV infection was successfully averted [37]. While these children may have otherwise been infected with HIV, they were instead just *exposed* to HIV *in utero*. Hereafter, these children will be referred to as children who are HIV-exposed but uninfected (CHEU). There is growing appreciation that, while CHEU have better outcomes than those CLHIV [38], they are still at risk of worse outcomes than children who are HIV-unexposed (CHU) [39], particularly outcomes of infectious disease [40]. This is likely true for tuberculosis as well, with early studies demonstrating an increased risk for TB exposure, infection, and disease. I address each here in turn.

*TB Exposure in CHEU*

Mothers living with HIV are known to have higher rates of TB infection than those living without HIV and are also at higher risk of progressing to TB disease during the peripartum period [41]. CHEU, therefore, are understood to have higher rates of congenital tuberculosis [42, 43]. With an increased likelihood of maternal TB exposure that persists beyond birth, however, we expect to also see higher rates of TB exposures through the aerosol route, which is borne out in the data. A Russian study showed a 3.7x higher risk of TB exposure for CHEU that was significantly greater than that experienced by healthy controls ($p<0.001$) [44]. Similarly, the IMPAACT P1041 study, while not enrolling CHU, observed a rate of TB exposure that was similar between CHEU (at 13%) and CLHIV (12%), who are also known to have higher rates of TB exposure [45]. These increased rates of TB exposures demonstrate a critical question that should be considered throughout the remaining chapter: are disparate outcomes primarily the result of a biological impact of HIV exposure, or are they simply due to confounding variables correlated with HIV exposure, such as socioeconomic status or access to healthcare? While studies attempt to control for these external factors to search for biological underpinnings of the poor outcomes observed in CHEU, it is difficult to distinguish between the likely multifactorial influences driving the worsened health outcomes of CHEU, which are summarized in **Figure 1**.

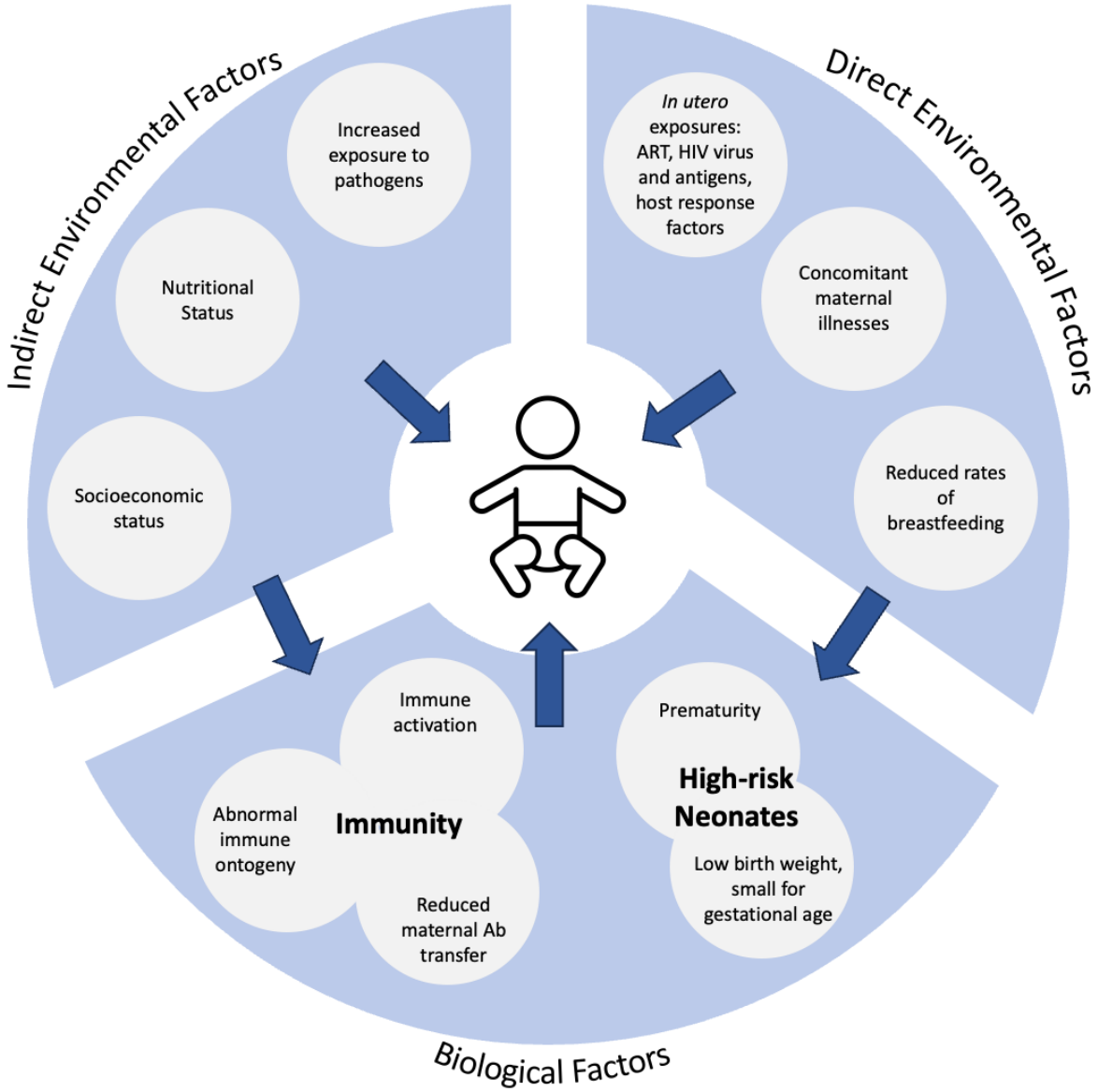



**Figure 1.** The multifactorial influences on CHEU which may drive altered TB outcomes. "Direct Environmental Factors" are those which directly follow from maternal HIV infection; in contrast, "Indirect Environmental Factors" are only correlated with HIV exposure. These factors may act independently or through biological factors, such as those listed.

*TB Infection in CHEU*

CHEU are also thought to be at elevated risk of TB infection. It is unclear, however, whether this is due to increased rates of TB exposure, easier establishment of TB infection, reduced ability to clear infection, or some combination thereof. It could also be that tests

of infection exhibit increased sensitivity or decreased specificity in CHEU, which would cause an increase in infection statistics but not in the underlying prevalence of infection, though this is purely speculative. Nonetheless, the impact of HIV-exposure on infection rates is quantifiable and can be measured by either IFN-γ release assays such as QuantiFERON-TB Gold (QFT), or tuberculin skin testing (TST). A sub-study of the IMPAACT P1041 cohort showed that QFT positivity in CHEU actually exceeded that of CLHIV at both measured timepoints (14% vs 12% at enrolment, and 24% vs 15% at follow-up, for CHEU and CLHIV, respectively), notable considering that CLHIV are thought themselves to be at a greater risk of TB infection than CHU [45]. In a study conducted in rural Uganda, CHEU had 21.2 (95% confidence interval (CI) 2.2-204.7) times higher adjusted odds of TB infection than CHU [46], as measured by TST and QFT. In a Kenyan cohort, CHEU had a 3.1 (95% CI 1.2-8.5)-times higher rate of TST positivity than CHU [47].

*TB Disease in CHEU*

This increased rate of TB infection aligns with reports of greater incidence of TB disease among CHEU. The South-Africa-based PERCH study found *Mtb* to contribute to a larger portion of pneumonia cases in CHEU (aetiologic fraction (AF) 12%, 95% credible interval (CrI) 6-19%) than in CHU (AF 8%, 95% CrI 4.5-13.8%) [48]. IMPAACT P1041 also found a rate of TB disease in CHEU higher than that reported in other studies (41 cases per 1000 child-years in CHEU [45], versus 9 per 1000 in CHU reported in a separate study based in South Africa [49]); however, differences in study design make these values difficult to directly compare. Conversely, a South African birth cohort study found the proportion of CHEU in all those diagnosed with TB disease to be lower than the proportion of CHEU overall, at

16/98 (16%) and 228/1143 (20%), respectively (the remaining children diagnosed with TB were all CHU) [50].

Early evidence also supports an increased risk of progression to TB disease after infection in CHEU, with an overall progression rate of less than 10% in Bacillus Calmette-Guérin (BCG)-vaccinated CHU compared with 20-40% of BCG-vaccinated CHEU [38, 39, 51-54]. However, the relative contributions to TB disease incidence of this increased risk of both infection and progression have not been sufficiently disentangled. As such, while it does still appear as if CHEU have a higher incidence and risk of progressing to TB disease than CHU, more research is needed to both cement and explicate this finding. Questions about decreased BCG effectiveness in CHEU, and the immune modulations altered effectiveness may reveal, are also pertinent but are too extensive to be discussed herein. It is hopefully apparent, however, that CHEU represent a cohort of growing, but underrecognized, importance for paediatric TB research. With a higher pre-test probability of disease, as well as potential immune changes induced by HIV-exposure, host-based diagnostics for TB may require dedicated evaluation and optimization for this cohort. More broadly, a diagnostic's effectiveness is considerably influenced by the biological and epidemiological context into which it is deployed, making such considerations critical.

## 1.4 Designing the Ideal Diagnostic

Many factors contribute to the performance of a diagnostic assay. As mentioned, the population in which the diagnostic is used is critical, both for calibrating a test to the specific biology of, for example, individuals who are HIV-exposed, but also for optimally

balancing sensitivity and specificity for use in that cohort. The importance of this balance is demonstrated by Bayes' theorem (given in **Equation 1.1**), which calculates, in the context of diagnostics, the odds of having a disease given one's test result, the accuracy of the test used, and one's pre-test probability of disease, which is influenced by disease prevalence.

$$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)}$$

**Equation 1.1**. Bayes' theorem. Here, P(A) represents the probability of having the disease being tested for, P(B) represents the probability of the test being positive, P(A|B) represents the probability of having the disease given a positive test, and P(B|A) is the probability of having a positive test, if the tested individual has the disease (sensitivity).

From the equation, it becomes apparent that tests used in areas with low TB prevalence should typically prioritize specificity over sensitivity, as a higher likelihood of test positivity (given by P(B)) is more likely to produce false positive results when the pre-test probability of disease (given by P(A)) is low [55]. Conversely, regions with higher TB burdens can strike a different balance between sensitivity and specificity, in reflection of the fact that fewer true negative cases will typically be tested. Considering the prevalence and the pre-test probability of disease, then, is critical for deciding upon the optimal diagnostic for each unique use case.

Additionally, the accessibility, affordability, and acceptability of novel diagnostics are important considerations, particularly in the development of entirely novel diagnostic approaches. For example, NAATs performed on oral swabs, a less-invasive sample type currently under investigation, are primarily researched not because they demonstrate better diagnostic performance, but because the sample type is more easily collected [56].

If, it is argued, a less-invasive sample type yields a 10% drop in sensitivity, but allows 50% more individuals to be tested, that approach can still be superior from a public health perspective, as it detects the largest number of cases [57]. Of course, the increased number of false positives which can emerge from increased testing volume must also be considered; however, it is generally accepted that cheaper, more accessible tests may yield greater population-level case detection even with lower individual-level sensitivity and specificity than less accessible counterparts with higher individual-level performance [58]. Additionally, testing approaches which are more accepted by the individuals offered testing (and, in the case of children, their families) are themselves of greater value if they enable greater numbers of individuals to consent to being tested for TB. The balance between these implementation considerations and the technical, diagnostic performance of TB tests is core to the design of optimally positioned diagnostic tests.

In reflection of the need for a set of target characteristics to guide TB test manufacturers, the WHO developed target product profiles (TPPs) for TB screening and diagnostic tests (detailed in **Tables 2** and **3**), which consider many of the above factors [58, 59]. While not designed specifically for the paediatric setting, the TPPs are a helpful benchmark against which the reported performance of novel diagnostics can be compared, particularly when considered in conjunction with setting-specific needs and preferences.

**Tables 2 and 3.** Evaluation criteria from the WHO's Target Product Profiles (TPPs) for a TB screening and diagnostic test which are most relevant to the results presented in subsequent chapters. This table is adapted directly from the cited WHO publications. Abbreviations: Point-of-Care (POC).

**Table 2: TPP for a Diagnostic Test [58]**

| | **Minimal Requirements** | **Optimal Requirements** |
|---|---|---|
| ***Non-sputum, low-complexity assay*** | 80% Sensitivity/>98% Specificity | ≥ 95% Sensitivity/>98% Specificity |
| ***Non-sputum, near-POC assay*** | 75% Sensitivity/>98% Specificity | |
| ***Non-sputum, POC assay*** | 65% Sensitivity/>98% Specificity | |
| ***Non-valid result rate*** | <5% | <3% |
| ***Treatment monitoring capability*** | Not required | Preferable |
| ***Sample Type*** | "Sputum or non-sputum samples that are not more complex to obtain than sputum". | "Self-collectable clinical specimens". |
| ***Time-to-result*** | <60 minutes | <15 minutes |

**Table 3: TPP for a Screening Test [59]**

| | **Minimal Requirements** | **Optimal Requirements** |
|---|---|---|
| ***High-sensitivity, high-specificity screening test*** | 90% Sensitivity/80% Specificity | 95% Sensitivity/95% Specificity |
| ***High-sensitivity screening test*** | 90% Sensitivity/60% Specificity | 95% Sensitivity/85% Specificity |
| ***High-specificity screening test*** | 60% Sensitivity/98% Specificity | 70% Sensitivity/98% Specificity |
| ***Sample Type*** | "Requires easy-to-specimen type obtain samples, including non-physical samples such as imaging and sound". | "Does not require a physical sample, or requires an easy-to-obtain sample that is self-collectible". |
| ***Time-to-result*** | <30 minutes | <5 minutes |

## 1.5 Conclusion

Dr. Martin H. Fischer is attributed as saying, “Diagnosis is not the end, but the beginning of practice” [60]. For the 50% of children with paediatric TB who go undiagnosed, lacking a diagnosis removes even the chance of accessing life-saving treatment. And with the field’s current reliance on clinical diagnosis for the >50% of TB cases which are not detected by current microbiological tests, treatment opportunities may be missed, or not even be available, if treatment is given to misdiagnosed children instead. It is clear that we urgently need improved diagnostic technologies to help reduce paediatric TB mortality.

Given the low sensitivity of microbiological testing to detect paediatric TB, host-based diagnostics represent a promising new frontier for paediatric TB diagnosis. A key challenge to host-based diagnostics, however, is accounting for other factors that influence the host response, including co-infections, age, sex, or other environmental factors. Whether a universal signature agnostic to these influences can be developed, or if signatures must (or can) be made sufficiently specific for these factors, has yet to be determined. Nevertheless, several approaches are under active investigation, including transcriptomic, proteomic, and metabolomic signatures of the host response to TB disease.

In this thesis, I will explore transcriptomic and proteomic approaches for the diagnosis of paediatric TB disease. First, I report on the diagnostic performance of a prototype host-based diagnostic developed by Cepheid using a modified signature first reported by Sweeney et al. [61]: the Xpert Host Response Cartridge. Subsequently, I investigate the diagnostic performance of other known transcriptomic signatures of TB through targeted gene expression analysis of RNA-stabilized whole blood by NanoString. Finally, I report on

the performance of various proteomic signatures of disease, using blood plasma collected from the same cohort. In sum, here I investigate multiple distinct approaches to the use of host-based biomarkers to diagnose TB disease in children and, from this, outline paths forward for future study.

# Chapter 2: Methods

## 2.1 Introduction

My DPhil work utilized samples collected during the NOD-pedFEND study to evaluate the diagnostic performance of the Cepheid Host Response Cartridge, as well as to explore various transcriptomic and proteomic signatures of TB. While methods specific to each chapter's results are detailed in the relevant chapter, what follows is a general overview of the NOD-pedFEND study itself, as well as the reference standards used in subsequent results chapters. Additionally, this chapter presents the general characteristics of all children enrolled into the Uganda-based NOD-pedFEND study, the full cohort from which a subset of study participants was selected for the work presented in a given chapter. The characteristics of each subset of selected participants are detailed in the relevant chapter.

## 2.2 Design and Setting of the NOD-pedFEND Study

*Study Setup and Enrolment*

The NOD-pedFEND study was a non-interventional paediatric TB diagnostic cohort study based in Uganda and Peru [62]. NOD-pedFEND is the result of the integration of two protocols: Novel and Optimized Diagnostics for Paediatric TB (NOD), which focused on the development of novel diagnostic assays for paediatric TB, and the paediatric arm of Feasibility of Novel Diagnostics for TB in Endemic Countries (FEND), which centred around the validation of prototype diagnostics and test combinations. Together, both studies

established the same well-characterized paediatric cohort, that of NOD-pedFEND, to enable the pursuit of both studies' aims. To do this, NOD-pedFEND recruited children under five with presumptive TB from four study sites: two in southern Uganda and two in Lima, Peru. Due to low study enrolment in Peru, however, only samples from Uganda were utilized for all analyses. In Uganda, the study actively enrolled participants from August 2021 to October 2024 and was based at two clinical sites: one in Kampala, at the Baylor Uganda Children's Clinical Centre of Excellence; and the other in Jinja, at the Nalufenya Children's Hospital. Both sites were tertiary hospitals, although the former was an urban academic medical facility primarily focused on paediatric and HIV/acquired immunodeficiency syndrome (AIDS) care, formed as a public-private partnership between Baylor College of Medicine in the United States and the Ugandan government. In contrast, Nalufenya Children's Hospital serves as a public paediatric referral hospital for a range of emergent conditions and covers a more rural and semi-urban catchment area. In both sites, after unsuccessfully undergoing a trial of antibiotic therapy, children entered the symptomatic study arm according to the following inclusion and exclusion criteria[1]:

Inclusion Criteria:

- Age less than five years at time of enrolment
- Written informed consent obtained from parent and/or legal guardian
- No treatment for TB disease or infection in the past six months
- Weight greater than 3kg
- Has one or more of the following:
    - Cough for more than two weeks that is refractory to antibiotic therapy; or
    - Weight loss, defined as an unexplained weight loss of greater than 5% compared to the highest weight recorded in the prior 3 months; or
    - Failure to thrive, defined as:

[1] Inclusion and exclusion criteria were largely copied from the NOD-pedFEND study protocol, which was not drafted by the author.

        - A clear deviation from a prior growth trajectory; or
        - Documented downward crossing of percentile lines in the preceding three months; or
        - A weight-for-age and/or weight-for-height z-score of ≤ -2 in the absence of information on previous growth trajectory
    - Unexplained fever (i.e. malaria excluded) for more than one week that is refractory to antibiotic therapy, which can be subjectively reported or objectively measured
    - Night sweats for greater than one week

Exclusion Criteria:

- Critical condition, such as hypovolemic shock or clinically relevant anaemia (tachypnoea, tachycardia)
- If study procedures seem like an undue risk to the child's life
- Expected unavailability for follow-up study visits (i.e. travel)
- Parents and/or legal guardian unwilling or unable to provide written consent.

*Description of Study Visits*

If participants met inclusion criteria, and all relevant parties continued to be amenable to their study participation, the participant was enrolled into the study. Upon enrolment, the participant was given an extensive two-day diagnostic workup for TB, which included the following microbiological tests:

Day One: Xpert Ultra, MGIT, and LJ culture of one nasopharyngeal aspirate sample, along with a Determine LAM test[2].

Day Two: Xpert Ultra, MGIT, and LJ culture of an additional nasopharyngeal aspirate sample, along with a gastric aspirate sample.

Day One or Two: Xpert Ultra of a stool sample.

[2] Of note, Determine LAM tests were evaluated both prospectively and retrospectively on a frozen urine sample, with a participant only considered LAM positive if they were positive on both tests. For CLHIV, the results of the prospective test were reported back to the clinician. Retrospective tests which were positive in a CLHIV would also have been reported back to the treating clinician if the original prospective test performed had been negative; however, this scenario did not apply for any NOD-pedFEND participants.

Along with these standard microbiological tests, blood samples were collected at both the day one and, for children who weighed more than six kilograms, the day two visit. Exceptions to the standard collection schedule could be made to stay within maximum volume limits for blood draws.

Participants were also asked to return to the clinic two weeks (± four days), and two months (± seven days) after their two-day baseline visit. Blood samples were also collected during the week-two visit, as further detailed below. Further, participants were able to request an unscheduled visit at any time, from which standard visit data, further described below, was also collected. Finally, participants were asked to come in for a final visit six months (± 14 days) after their baseline visit or at the end of their TB treatment regimen, whichever was later.

At these visits, participants underwent a comprehensive physical examination, medication review, and medical history assessment, which included a review of symptoms and documentation of any known TB contacts at baseline. If any contacts to individuals with a chronic cough were reported, these were noted with follow-up at study conclusion to identify if those individuals had been diagnosed with TB. Additionally, at the two-week visit, a tuberculin skin test (TST) was performed, which the child returned to clinic 48-72 hours later to have read (with ± 2 hours of buffer on either side of the allowable time window). Study procedures for the symptomatic arm are summarized in **Figure 1.** A list of variables which could be recorded by the site team over the course of the study for each participant is provided in **Appendix 1**.

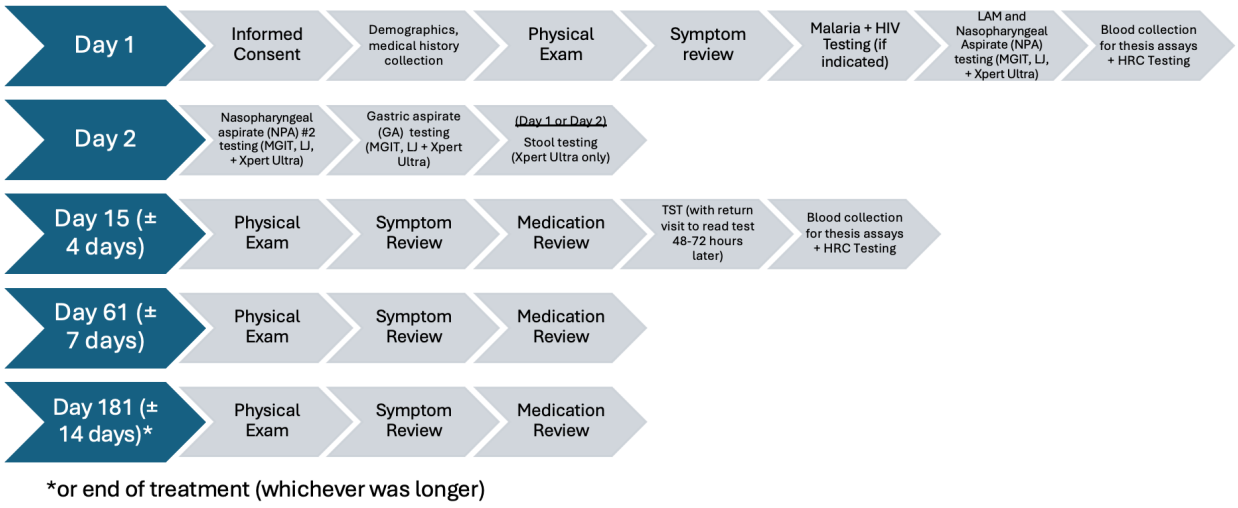



**Figure 1.** A summary of the relevant procedures performed for children enrolled into the symptomatic arm of NOD-pedFEND.

*Healthy Controls*

At the Kampala study site, asymptomatic participants were also recruited to serve as healthy controls. For enrolment, healthy controls were matched, to the greatest extent possible, to an enrolled symptomatic child with microbiologically confirmed TB based on age, sex, and HIV status. Inclusion criteria for healthy controls were as follows:

1) No history of symptoms of TB in the prior six months.
2) No history of known TB exposure.

Exclusion criteria for healthy controls were the same as those used for the symptomatic arm.

Like participants in the symptomatic arm, healthy controls were scheduled for a minimum of four visits: one at baseline, one two weeks (± four days) later, one two months (± seven days) after the baseline visit, and one six months (± fourteen days) after the baseline visit. While children in the healthy control arm were not given microbiological testing, they were given a comprehensive symptom review, medication review, and

physical exam at each visit. Additionally, a range of samples, including the whole blood in PAXgene tubes used for this thesis, were collected at the baseline visit, during which a TST test was also administered and read 48-72 hours (± 2 hours) later. Study procedures for the healthy control arm are summarized in **Figure 2.**

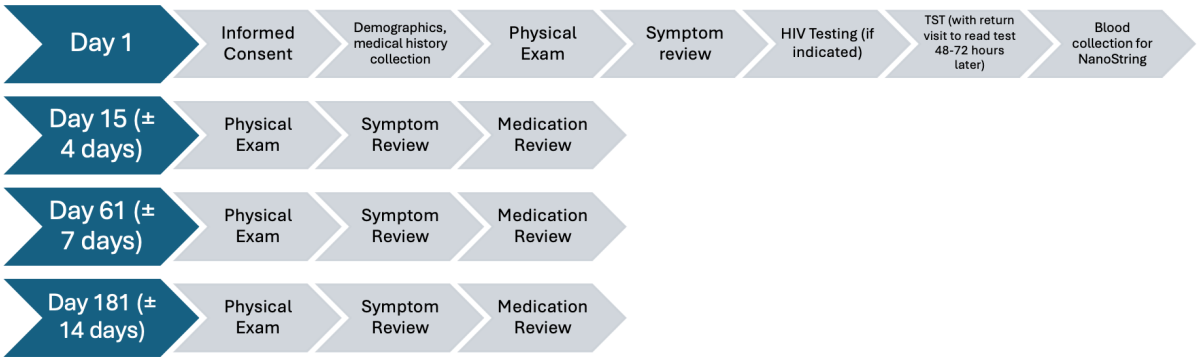



**Figure 2.** A summary of the relevant procedures performed for children enrolled into the healthy control arm of NOD-pedFEND.

*Sample Collection*

All samples utilized for the analyses presented in this thesis were collected at either the first day of a participant's baseline visit, or at the visit scheduled two weeks (± four days) after. Three sample types were utilized: capillary blood (for Host Response Cartridge testing); whole blood, collected in PAXgene tubes (for transcriptomic work); and blood plasma, collected first in ethylenediaminetetraacetic acid (EDTA) tubes (for measurements of protein biomarkers). While capillary blood samples were immediately tested, PAXgene and plasma samples were frozen at -80° C after processing and were shipped frozen to Rutgers University prior to their analysis. For both plasma and PAXgene tubes, processing first consisted of simply inverting the tube 10 times. For PAXgene tubes, this was followed by incubation at room temperature for 2 to 72 hours, aliquoting, and storing aliquots first at

-20° C for 24 hours before freezing them long-term at -80° C. For the blood plasma samples utilized, inversion was followed by refrigeration at 2-8° C until ready for centrifugation, with centrifugation performed for 10 minutes at 1500xg, also at the refrigerated temperature. Plasma was then taken from the top of the EDTA tube and aliquoted, after which it was immediately stored at -80° C. Aliquots of both PAXgene and plasma samples were made to enable on-site testing of future blood-based diagnostic prototypes. The remaining samples were eventually shipped to Rutgers University on dry ice for testing; on receipt, samples were immediately placed back into a freezer set to -80° C.

Of note, QFT samples were also obtained and stored at -80° C as two components: the plasma and cell pellet which emerged from centrifuging the samples at 2500xg. While outside the scope of this thesis, these samples remain available for further analysis.

## 2.3 Ethical considerations

Study design, procedures, and informed consent protocols were approved by relevant authorities in Uganda, the United Kingdom, and the United States. In Uganda, local ethical approval was received by the School of Biomedical Sciences Research Ethics Committee (SBS-REC-845), as well as the Uganda National Council for Science and Technology (HS1341ES). Ethical approval was also obtained from the Oxford Tropical Research Ethics Committee (OxTREC 9—21), as well as the Rutgers University IRB (Pro2020000750). All research activities adhered to the Declaration of Helsinki [63].

Additionally, as NOD-pedFEND was a non-interventional study, risk to participants was minimized. Site clinicians retained full discretion over treatment decisions at all times,

and these decisions did not affect a participant's continued enrolment. Participants were able to leave the study at any time and were informed as such prior to enrolment in the study. Of note, clinicians did have access to results from WHO-endorsed microbiological tests and CXRs performed as part of the study and were permitted to use them, in accordance with standard practice, in their clinical decision making. In contrast, investigational test results were not disclosed to site clinicians. All such procedures were approved and continually reviewed by the aforementioned review boards.

## 2.4 Study Monitoring and Data Management

Study monitoring and data management were led by the study sponsor, Rutgers University, and performed before, after, and throughout the study's duration in a mix of on-site and virtual visits. Monitoring was conducted to ensure study procedures conformed to the approved protocol and Good Clinical Practice (GCP) guidelines. Items verified during study monitoring included procedures for laboratory testing and informed consent, appropriate record-keeping of critical study documents, accuracy of collected data, documentation of protocol deviations, and continued adequacy of study personnel and facilities. On-site or virtual trainings were also conducted for relevant personnel upon receiving all required approvals of protocol revisions or amendments.

To record all collected data, participants were assigned an anonymized study ID, with only the local study sites securely storing the key necessary for deanonymization. All primary study data were first collected on paper worksheets, referencing the participant's assigned ID. Worksheets were later entered and stored in a password-protected online

REDCap database by on-site study personnel [64]. Data entered into REDCap were regularly compared against these paper worksheets, with all discrepancies resolved. Data cleaning and validation were performed continuously over the course of the study, through dataset finalization. When clinical variables required computation (as in the case of weight-for-length Z scores, clinical case classifications, etc.), the variables we calculated were validated against those computed using the same dataset by an external statistical team—from the Frontier Science Foundation—with all discrepancies resolved prior to dataset finalization. Analysis was performed using this finalized dataset of all clinical study data, which had been cleaned and locked to further edits.

## 2.5 Case Classification Approach

Study endpoints and case classifications were based on the National Institutes of Health (NIH)'s consensus statement on the Clinical Case Definitions for Classification of Intrathoracic Tuberculosis in Children [65], operationalized to NOD-pedFEND's clinical context with the addition of an "unclassifiable" group and the utilization of stool testing in the definition of "confirmed TB". In short, participants were classified as "confirmed TB" when microbiological confirmation of disease was obtained, "unconfirmed TB" when microbiological confirmation was not obtained, but clinical history or chest radiography were suggestive of a TB diagnosis, "unlikely TB" when the child lacked both microbiological confirmation or clinical history/chest radiography suggestive of TB, and "unclassifiable" when an insufficient volume of data was collected for an enrolled child such that a definitive classification was unable to be determined for them, or if the child was started

on tuberculosis preventative therapy (TPT) before the end of study visit. The full, detailed criteria used for case classification were as follows:

1) Confirmed TB
    a. The participant was positive on at least one of the following microbiological tests, performed at the two-day baseline visit:
        i. Xpert Ultra of stool, NPA (day 1), NPA (day 2), or GA.
        ii. MGIT culture of NPA (day 1), NPA (day 2), or GA.
        iii. LJ culture of NPA (day 1), NPA (day 2), or GA.

2) Unconfirmed TB
    a. The participant:
        i. Did not meet criteria for confirmed TB; and
        ii. Was not given TPT before the end of study visit; and
        iii. If untreated, did not have resolution of the four tracked symptoms (detailed below) at the end of study visit; and
        iv. If untreated, attended the end of study visit or died prior to that visit; and
        v. Had any two of the following:
            1. Any of the four tracked symptoms of tuberculosis, further detailed below (given NOD-pedFEND's inclusion criteria, all symptomatic children received this point.)
            2. Chest radiograph suggestive of tuberculosis
            3. Reported close exposure to an individual with TB
            4. Positive TST result
            5. Response to treatment, defined as:
                a. Resolution of all of the four tracked symptoms which were present at baseline at either the visit closest to two months from treatment initiation or visit held at the end of treatment (with all symptoms needing to have been resolved at that visit).

3) Unlikely TB
    a. The participant:
        i. Did not meet criteria for confirmed or unconfirmed TB; and
        ii. Was not given TPT before the end of study visit; and
        iii. Had one of the following:
            1. If untreated, resolution of all the four tracked symptoms at the end of study visit; or
            2. Irrespective of treatment status, all of the following:
                a. An interpretable chest radiograph
                b. Adequate follow-up data, defined as at least one visit held ten days or later after baseline for untreated

participants, and two months or later for treated participants.

4) Unclassifiable
    a. The participant did not meet criteria for confirmed, unconfirmed, or unlikely TB, which should only be due to one of the following:
        i. Missing data with the potential to have changed the participant's final case classification (if one were to be otherwise assigned); or
        ii. The participant was started on TPT prior to the end of study visit.

The four tracked symptoms of tuberculosis utilized for the definitions given above were as follows:

- Cough
- Failure to thrive/weight loss, defined as:
    - A weight-for-age or weight-for-length/height z score of ≤ -2, calculated using the WHO's child growth standards [66].
- Fever
- Fatigue or lethargy

In the case of cough, fever, or fatigue, the presence or absence of symptoms was determined from reported history given by the participant's parent or caregiver. In the case of failure to thrive/weight loss, objective anthropometric measurements recorded by the site clinician were utilized. Further, the treating clinician also had to agree that a response to treatment had been demonstrated.

## 2.6 Reference Standard Definitions

From the case classifications determined via the above definitions, reference standards were constructed against which the performance of novel diagnostics could be assessed. Each standard defined a set of "true positive" cases and "true negative" controls that the results of an evaluated diagnostic were compared against. Five standards were

utilized: the strict reference standard (SRS), a modified strict reference standard (mSRS), the microbiological reference standard (MRS), the extended microbiological reference standard (eMRS), and the composite reference standard (CRS). Each standard is composed of different combinations of the study cohort, with some participants considered reference standard positive, others reference standard negative, and the remainder reference standard unclassifiable. While standards primarily utilize the NIH case definitions described in **Section 2.5**, they may also consider other factors; for example, whether a CLHIV was positive on a LAM test. I discuss each standard, and how they are determined, in turn.

The SRS is designed to examine diagnostic performance in the children from whom TB status was known with the greatest certainty: those with confirmed TB (defined as SRS+) and those with unlikely TB (defined as SRS-). As such, all children with unconfirmed TB were considered SRS unclassifiable (notated as “SRSun”). Of note, as NIH case classifications do not currently consider LAM test results in the assignment of case definitions, it is possible for a CLHIV with a positive LAM test to be classified as unlikely TB, as was the case for one child in NOD-pedFEND. As a result, and given that LAM positivity in a CLHIV would itself be sufficient to start TB treatment in most cases [12], the mSRS was constructed. Its only difference from the SRS is the classification of the single LAM+ CLHIV with unlikely TB, who moves from SRS- to mSRS unclassifiable.

The MRS, in contrast, focuses evaluations on the ability of a diagnostic to identify microbiologically confirmed TB. As such, only individuals with confirmed TB were classified as MRS+. Individuals who did not have confirmed TB, and who had results available from at

least six of the microbiological tests used for identifying confirmed TB, were all classified as MRS-. Notably, this MRS- group includes those CLHIV who were LAM+ but who were not classified as confirmed TB. As a result, almost all children with unconfirmed and unlikely TB classifications are considered MRS-, with only those without available results from at least six microbiological tests being classified as MRS unclassifiable (MRSun). Relatedly, the eMRS mirrors a participant's MRS classification, except that all CLHIV who are LAM+ are classified as eMRS+. This standard better reflects the fact that a positive LAM test in a CLHIV is typically considered sufficient laboratory evidence to diagnose TB in these children [12], at the expense of the slightly lower specificity associated with LAM testing [67].

Finally, the CRS was constructed to best reflect which children are currently diagnosed with TB, incorporating children with characteristics consistent with clinically diagnosed TB in the CRS+ group. This standard is largely a consequence of the fact that many children with TB may remain negative on currently available microbiological tests; in other words, many children with unconfirmed TB may truly have TB, although this is known with less certainty. At the expense of lower specificity in the "true positive" group, then, the CRS classifies all children with confirmed and unconfirmed TB as CRS+, with only children with unlikely TB being CRS-, and unclassifiable children being "CRSun". A graphic depicting the differences between these reference standards is shown in **Figure 3.**

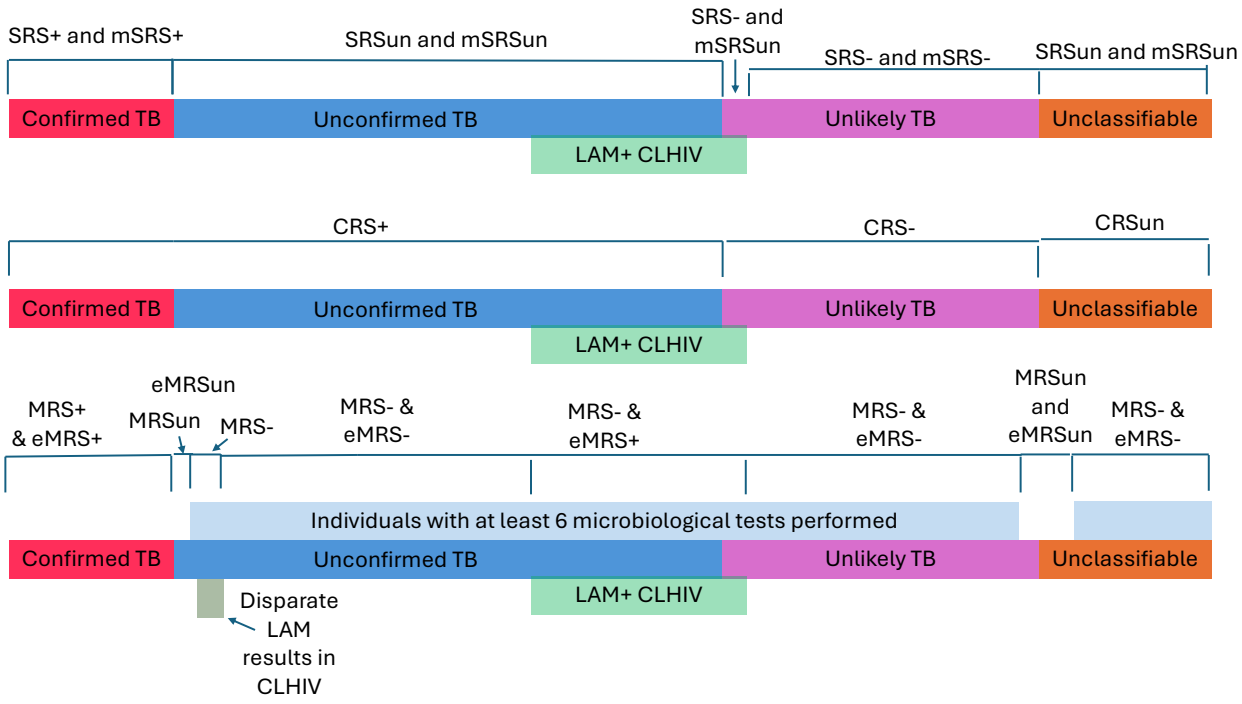



**Figure 3.** A graphic detailing the specific reference standards utilized in NOD-pedFEND, and the participants which were selected in the construction of each standard. Microbiological tests include all tests utilized for determination of the "confirmed TB" classification. The letters "un" next to any standard signifies participants who were unclassifiable by that reference standard. Abbreviations: Microbiological Reference Standard (MRS), Composite Reference Standard (CRS), Strict Reference Standard (SRS), Extended Microbiological Reference Standard (eMRS), Modified Strict Reference Standard (mSRS), Unclassifiable (un), Lipoarabinomannan (LAM), Children Living with HIV (CLHIV).

## 2.7 Additional Definitions

To enable subgroup analyses, each participant also had their malnutrition status and disease severity calculated using pre-determined algorithms. Malnutritional status was calculated in alignment with WHO definitions, utilizing anthropometric indices (based on weight and either age or height/length), the measured middle-upper-arm circumference (MUAC), and the presence (or absence) of nutritional oedema. The criteria used for

determining malnutrition status were as follows, determined using the WHO's child growth standards [66]:

> Severe acute malnutrition (SAM): Weight-for-length/height z-score < -3, or MUAC < 115mm, or presence of nutritional oedema.
>
> Moderate acute malnutrition (MAM): Weight-for-length/height z-score ≥ -3 and < -2, or MUAC ≥ 115mm and < 125mm.
>
> No acute malnutrition: Weight-for-length/height z-score ≥ -2 and MUAC ≥ 125mm, and no nutritional oedema.
>
> Malnutrition status unknown: All remaining children who meet none of the above definitions, typically due to missing data.

Additionally, disease severity was calculated to align with the criteria for "severe TB" published by the WHO for the purpose of determining TB treatment length. The determination of disease severity took into account nutritional status, Xpert semi-quantitative results, chest radiography, symptom presentation, and whether TB treatment was initiated. The classification of TB severity[3] was determined as follows:

Severe TB:

1) The child was started on treatment for TB disease (not TPT); and
2) The child has any of the following:
    a. Radiologically severe TB, defined as:
        i. Two CXR readers agreeing on a definition of "severe TB"; or
        ii. Two CXR readers agreeing on the presence of one of the following radiographic findings:
            1. Cavitation, miliary pattern, airway compression, complex lymph node disease, pneumothorax, vertebral collapse, paravertebral abscess, expansile pneumonia, lung collapse, complex effusion, multi-lobar disease, or hyperinflation.
    b. A semi-quantitative Xpert value of medium or above, on at least one sample
    c. SAM

[3] The criteria for severe TB were largely adapted to NOD-pedFEND by Dr. Francesca Basile and are only paraphrased by the author here.

d. A "severity symptom", defined as:
    i. Any of fever >39° C, asymmetrical and persistent wheezing, severe pallor, restlessness, irritability, profound lethargy, respiratory distress, vomiting, severe dehydration, unable to eat and drink, stridor, $O_2$ saturation <90%, seizures, unconsciousness, neck stiffness, or a bulging fontanelle.

Non-severe TB:

1) The child was started on treatment for TB disease (not TPT); and
2) The child has all of the following:
    a. Known nutritional status, and no SAM
    b. No radiologically severe TB, as defined above
    c. No severity symptoms, as defined above
    d. No semi-quantitative Xpert values of medium or above, with at least one Xpert result available.
    e. Radiologically non-severe TB, defined as:
        i. Two CXR readers agreeing on a definition of "normal CXR" or "non-severe CXR"; or
        ii. Two CXR readers agreeing on the presence of one of the following radiographic findings:
            1. Ghon focus, interstitial pneumonia, perihilar infiltrates, simple lymphadenopathy, simple effusions.

If the above criteria are not met, disease severity for that child is considered to be unclassifiable.

## 2.8 Statistical Approaches

Unless otherwise stated, comparisons between group medians were performed using non-parametric methods. For comparisons between two groups, the Wilcoxon rank-sum test was used. For comparisons between 3 or more groups, the Kruskal-Wallis rank-sum test was used for comparisons of continuous variables, and the Pearson's chi-squared test with simulated p-values (with 2000 replicates) was used for categorical variables. The

confidence interval for sensitivity and specificity calculations was determined using bootstrapping with 2000 replicates. When receiver operating characteristic (ROC) curves were constructed, the AUC was calculated with reference standard positive and negative individuals only; reference standard unclassifiable children were not included in construction of the ROC curve and subsequent AUC calculation. Statistical methods specific to an individual chapter are discussed in the respective chapter.

## 2.9 NOD-pedFEND Cohort Characteristics

Overall, 1,328 children were screened in Uganda for potential inclusion in NOD-pedFEND. Of these children, 678 (51%) were eventually enrolled: 183 (27%) from the Jinja site and 495 (73%) from the Kampala site. 640 of these children were enrolled into the symptomatic arm, while 38 were enrolled into the healthy control arm. Healthy controls were only enrolled in Kampala. 470/640 (73%) of symptomatic children were started on treatment for TB disease, with an additional 24/640 (4%) of symptomatic children started on tuberculosis preventative therapy (TPT). General participant demographics, differentiated by each participant's ultimate case classification, are detailed in **Table 1.**

**Table 1.** An overview of characteristics of the full cohort enrolled into NOD-pedFEND at the Uganda study sites. For continuous variables, p values were determined using the non-parametric Kruskal-Wallis test across the five case classifications. For categorical variables, Pearson's $\chi^2$ test with Monte Carlo (simulated) p-values (n=2,000 replicates) was used. Case classifications were determined as outlined in **Section 2.5**. Abbreviations: Lipoarabinomannan (LAM), Tuberculosis Preventive Treatment (TPT).

| | **Confirmed TB (n=43)** | **Unconfirmed TB (n=383)** | **Unlikely TB (n=144)** | **Unclassifiable (n=70)** | **Healthy Controls (n=38)** | **P Value** |
|---|---|---|---|---|---|---|
| Age (in months) | 21 (13-36) | 17 (12-34) | 20 (13-40) | 19 (13-35.8) | 23 (12.2-40.8) | 0.29 |
| **Sex Distribution** | | | | | | 0.54 |
| % Female | 23/43 (53%) | 181/383 (47%) | 73/144 (51%) | 28/70 (40%) | 20/38 (53%) | |
| % Male | 20/43 (47%) | 202/383 (53%) | 71/144 (49%) | 42/70 (60%) | 18/38 (47%) | |
| **HIV Status** | | | | | | 0.02 |
| % HIV Positive | 7/43 (16%) | 65/383 (17%) | 5/144 (3%) | 13/70 (19%) | 4/38 (11%) | |
| % HIV Exposed-Uninfected | 7/43 (16%) | 67/383 (17%) | 26/144 (18%) | 6/70 (9%) | 7/38 (18%) | |
| % HIV Negative | 29/43 (67%) | 250/383 (65%) | 113/144 (78%) | 50/70 (71%) | 27/38 (71%) | |
| % HIV Status Unknown | 0/43 (0%) | 1/383 (0%) | 0/144 (0%) | 1/70 (1%) | 0/38 (0%) | |
| **Malnutrition Status** | | | | | | <0.001 |
| % with SAM | 15/43 (35%) | 151/383 (39%) | 46/144 (32%) | 29/70 (41%) | 0/38 (0%) | |
| % with MAM | 7/43 (16%) | 44/383 (11%) | 13/144 (9%) | 5/70 (7%) | 1/38 (3%) | |
| % Without Malnutrition | 19/43 (44%) | 186/383 (49%) | 81/144 (56%) | 32/70 (46%) | 36/38 (95%) | |
| % Malnutrition Status Unknown | 2/43 (5%) | 2/383 (1%) | 4/144 (3%) | 4/70 (6%) | 1/38 (3%) | |
| **Disease Severity** | | | | | | <0.001 |
| % with Severe Disease | 22/43 (51%) | 200/383 (52%) | 24/144 (17%) | 17/70 (24%) | 0/38 (0%) | |
| % with Non-Severe Disease | 16/43 (37%) | 157/383 (41%) | 3/144 (2%) | 4/70 (6%) | 0/38 (0%) | |
| % with Unclassifiable | 5/43 (12%) | 26/383 (7%) | 117/144 (81%) | 49/70 (70%) | 38/38 (100%) | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Disease Severity | | | | | | |
| **LAM Positivity** | | | | | | |
| % of all participants who are LAM+ | 9/43 (21%) | 103/383 (27%) | 23/144 (16%) | 6/70 (9%) | 0/38 (0%) | <0.001 |
| % of all participants who are LAM- | 21/43 (49%) | 209/383 (55%) | 82/144 (57%) | 44/70 (63%) | 0/38 (0%) | |
| % of all participants who are LAM unknown or unclassifiable | 13/43 (30%) | 71/383 (19%) | 39/144 (27%) | 20/70 (29%) | 38/38 (100%) | |
| % of all CLHIV with the NIH classification who are LAM+ | 2/7 (29%) | 14/65 (22%) | 1/5 (20%) | 1/13 (8%) | 0/4 (0%) | 0.004 (for differences in LAM results in CLHIV, measured across classifications) |
| **TST Results**[4] | | | | | | <0.001 |
| % TST+ | 17/43 (40%) | 98/383 (26%) | 8/144 (6%) | 7/70 (10%) | 3/38 (8%) | |
| % TST- | 23/43 (53%) | 267/383 (70%) | 123/144 (85%) | 37/70 (53%) | 31/38 (82%) | |
| % TST Status Unknown | 3/43 (7%) | 18/383 (5%) | 13/144 (9%) | 26/70 (37%) | 4/38 (11%) | |
| **Hospitalization Status** | | | | | | <0.001 |
| % Hospitalized at time of enrolment (rest enrolled as outpatient, unless otherwise noted) | 22/43 (51%) | 183/383 (48%) | 45/144 (31%) | 33/70 (47%), 2/70 (3%) hospitalization status unknown | 0/38 (0%) | |
| **Chest Radiography (CXR)**[4] | | | | | | <0.001 |
| % With CXR adjudicated as "Normal" | 17/43 (40%) | 174/383 (45%) | 95/144 (66%) | 33/70 (47%) | 0/38 (0%) | |

[4] Of note, a positive TST result or CXR suggestive of TB are each one of the possible conditions which could cause a child to be classified as "Unconfirmed TB".

| | | | | | | |
|---|---|---|---|---|---|---|
| % With CXR adjudicated as "Abnormal, Unlikely TB" | 13/43 (30%) | 134/383 (35%) | 45/144 (31%) | 18/70 (26%) | 0/38 (0%) | |
| % With CXR adjudicated as "Abnormal, Suggestive of TB" | 9/43 (21%) | 52/383 (14%) | 3/144 (2%) | 2/70 (3%) | 0/38 (0%) | |
| % With CXR missing or adjudicated as "unclassifiable" | 4/43 (9%) | 23/383 (6%) | 1/144 (1%) | 17/70 (24%) | 38/38 (100%) | |
| **Treatment Status by End of Study[5]** | | | | | | |
| Started on treatment for TB disease | 42/43 (98%) | 377/383 (98%) | 28/144 (19%) | 23/70 (33%) | 0/38 (0%) | <0.001 |
| Started on TPT only | 0/43 (0%) | 0/383 (0%) | 0/144 (0%) | 24/70 (34%) | 0/38 (0%) | |
| Untreated | 1/43 (2%)[6] | 6/383 (2%) | 116/144 (81%) | 22/70 (31%) | 38/38 (100%) | |
| Treatment status unknown | 0/43 (0%) | 0/383 (0%) | 0/144 (0%) | 1/70 (1%) | 0/38 (0%) | |

While the full cohort was not used in its entirety for any of the analyses detailed in subsequent chapters, it is presented as a reference against which the characteristics of NOD-pedFEND participants selected for particular analyses can be compared. Additionally, the data highlight some broad trends observed within the cohort. These

[5] Of note, treatment decisions could themselves influence or be related to a child's TB classification. For example, children in the healthy control arm were, by definition, untreated. Further, administration of TPT was sufficient to move a child to the "unclassifiable" category. Finally, symptom resolution in the absence of TB treatment categorized a child as "unlikely TB" while, conversely, symptom resolution after TB treatment fulfilled one of the possible criteria needed for a child to be classified as having "unconfirmed TB".

[6] One child with confirmed TB was only identified by Hain testing of a purported NTM performed ~5 months after their enrolment in the study had concluded. When exiting the study, the participant was reported to have had no remaining symptoms of TB. These revised microbiological results were communicated to the on-site study team for follow-up; however, as the child was not treated during the study period, they are listed as such here.

include a relatively low age at enrolment (median 18 months, interquartile range (IQR) 12-36), as well as a high proportion of malnutrition in the cohort, with nearly half of all participants with a known malnutrition status (310/628, 49%) having either moderate or severe acute malnutrition. The proportion of inpatient to outpatient enrolments was relatively even, with 44% (283/640) of participants enrolled into the study during an inpatient stay. Additionally, 14% (90/640) of all participants were CLHIV, with an additional 17% (106/640) being CHEU. Notable trends in presentation were also observed specifically amongst those children classified as confirmed TB, who had a substantial proportion of "normal" CXRs (17/43, 40%) and negative TST results (23/40 with a known result, 58%). Similarly, despite TST or CXR positivity being one of the means by which a child could be classified as having unconfirmed TB, only 98/383 (26%) and 52/383 (14%) of the children with unconfirmed TB had each, respectively.

# Chapter 3: Performance of the Host-Response Cartridge in the NOD-pedFEND Cohort

## Introduction

### 3.1 Principle of the Host Response Cartridge

Transcriptomic signatures of disease are being developed to help guide clinicians in the diagnosis and triage of a range of conditions, including sepsis, breast and prostate cancer, and Kawasaki disease [68-71]. Similarly, transcriptomic signatures of TB (further detailed in **Section 4.1**) have been developed, though none have, as of yet, been WHO-endorsed as a diagnostic technology. One such TB signature has been incorporated into a prototype diagnostic cartridge developed by the diagnostic company Cepheid (California, USA), which is dubbed the "Host-Response Cartridge" (HRC). We sought to evaluate the diagnostic performance of this technology in the NOD-pedFEND cohort.

The HRC utilizes a modified version of a signature originally developed by Sweeney et al. in 2016 to optimally distinguish between individuals with TB disease and TB infection [61]. The signature—composed of transcripts of the three genes Guanylate-binding protein 5 (*GBP5*), Dual-Specificity Phosphatase 3 (*DUSP3*), and Krüppel-like factor 2 (*KLF2*)—was developed using three publicly available datasets of gene expression microarrays performed on whole blood samples. Across the 3 datasets, 1023 samples were used to develop the signature in total: 296 from adults with TB disease, 236 from those with TB infection without disease (previously called "latent TB"), and 491 from those with other

diseases (pulmonary and non-pulmonary, including lung cancer, sarcoidosis, and a range of autoimmune diseases). The signature, when validated on 11 independent datasets (n=1345 samples in total), differentiated TB disease from TB infection with an AUC of 0.88 (95% CI 0.84-0.92).

The component genes of the signature are each implicated in host immunity. *KLF2* is a transcription factor which influences T Cell responses [72, 73], including through modulating CCR5 expression on CD4+ T Cells [74]. *DUSP3* is implicated in MAPK signalling and in upregulating innate immune pathways, particularly in macrophages and monocytes [75]. Finally, *GBP5* is an interferon-inducible GTPase which supports innate immunity by promoting assembly of the inflammasome and inhibiting viral envelope formation, and which is upregulated in neutrophils in patients with TB disease [76, 77].

Despite being based on the expression of only these three genes, the Sweeney signature's AUC of 0.88 in the 11 independent datasets reflected strong performance, supporting its potential for generalized application. Cepheid subsequently developed a prototype assay for measuring this signature using their existing Xpert platform. The assay, the aforementioned HRC, utilizes quantitative reverse transcription polymerase chain reactions (qRT-PCR) to generate a "TB Score" based upon the expression levels of the three component genes in the Sweeney signature, as measured by the cycle threshold (Ct) values of those genes in a capillary blood sample. Importantly, a lower TB score corresponds to a stronger indication of TB disease. The precise algorithm used to calculate the TB score is unknown, but it broadly aligns with the result of **Equation 3.1**.

$$\left(\frac{Ct_{GBP5} + Ct_{DUSP3}}{2} - Ct_{KLF2}\right)$$

**Equation 3.1** The equation used to estimate the *KLF2*-based TB score.

Recently, Cepheid updated the calculation of the TB score to utilize the expression of the housekeeping gene encoding for TATA-binding protein (*TBP)*, instead of *KLF2*. This change was purported to improve stability and performance across different sample collection methods, though the data used to make this determination are not publicly available. Unlike the *KLF2*-based TB score, however, the equation used to calculate the *TBP*-based score is known and is given in **Equation 3.2.** As the HRC prototype measures expression of both *TBP* and *KLF2*, along with *DUSP3* and *GBP5*, I performed this analysis of HRC performance using both the old (*KLF2*-based) and new (*TBP*-based) scoring methods.

$$\left(\frac{Ct_{GBP5} + Ct_{DUSP3}}{2} - Ct_{TBP}\right)$$

**Equation 3.2** The equation used to calculate the *TBP*-based TB score.

## 3.2 Reported Performance of the HRC

Evidence of the diagnostic accuracy of HRC has been mixed. Amongst incarcerated adults in Brazil, specificity of the HRC was 53% when set for 90% sensitivity in detecting

microbiologically confirmed disease (by sputum Xpert or culture), with sensitivity improving with higher bacterial loads [78]. A study amongst 6-83 year olds in Pakistan reported assay specificity of 32% when set to the same 90% sensitivity threshold and evaluated against the same reference standard; however, when restricting the analysis to participants between 10 and 24 years old, specificity rose to 82% [79], meeting the WHO's TPP for a non-sputum-based triage test [80]. However, in a cohort of 1262 participants aged 12-70 years across The Gambia, South Africa, Uganda, and Vietnam, specificity was 62.5% with sensitivity of 92.8% against a CRS [81], barely meeting the WHO criteria for a high-sensitivity screening test. In China, results were similar, with one study reporting sensitivity/specificity of 82.1%/65.6%, respectively, in differentiating TB disease from other pulmonary diseases in those over 12 years old [82], and another reporting 70% sensitivity and 75.8% specificity in those over 15 [83]. The heterogeneity of the results and populations studied merits further investigation to identify indications for use of the HRC, including dedicated studies in paediatric populations.

*HRC in Paediatric TB*

Few studies on HRC in exclusively paediatric populations have been performed, and only four, to my knowledge, published. The first report came from RaPaed-TB, which enrolled children under 15 years old in South Africa, Malawi, India, Tanzania, and Mozambique. In their cohort, a sensitivity cutoff of 90% yielded a specificity of 51.2% [84] against a SRS which only considered culture positivity, falling below the WHO's TPPs. However, as the test demonstrated similar performance across age, HIV, and malnutrition statuses, and utilized a sample that is relatively easy to collect, the authors concluded that

the test showed promise and warranted further exploration of paediatric-specific use cases.

Recently, the utility of HRC in differentiating WHO-defined severe TB from non-severe TB in children was also assessed as part of the COMBO study. Using a cutoff that achieved 91.7% sensitivity for detecting severe TB among children with confirmed disease, specificity was 62.5% [85]. In another cohort from the same study (with unclear participant overlap), setting the TB score cutoff to a sensitivity of 88.5% yielded a specificity of 30.3% against the MRS, with performance below that of a random classifier when evaluated against the CRS [86]. Notably, these results are the only paediatric-specific data published using the new TB score methodology which utilizes the *TBP* gene in lieu of *KLF2*. Finally, a recent study reported on HRC performance in 780 individuals ≤ 18 years old enrolled across four Chinese provinces. There, the researchers reported 39.8% specificity of the HRC when set to a sensitivity of 90% in identifying CRS+ participants versus healthy controls. Interestingly, however, performance was improved in children under 5, with 89.3% specificity at a sensitivity of 78.5% [87]. In this cohort, the ability of the TB score to distinguish TB cases was also improved in more severe TB cases, extrapulmonary TB, and when evaluated against a reference standard of culture-confirmed TB versus healthy controls.

*Use as a Treatment Monitoring Tool*

The utility of the HRC assay to evaluate responses to anti-tuberculosis treatment is also of interest. In the aforementioned Chinese study, TB scores significantly improved for participants after both one and three months of TB treatment [87]. In a prospective trial

enrolling adults in Moldova, Peru, and South Africa, however, only a 1.08-fold (IQR 0.54-1.52) increase in TB score was observed in the first month of treatment [88], though a 1.68-fold increase was observed (at a two month timepoint post-treatment) in the study by Li et al. [83]. Further, the study by Olbrich et al. on paediatric samples showed a statistically significant increase in TB score (Analysis of Variance (ANOVA) $p<0.0001$) in the months following treatment initiation in children with culture-confirmed TB [84]. Interestingly, while this signal was still observed in children with non-severe tuberculosis, it was not in children classified as "unlikely TB", indicating a potential use case for the HRC of informing retrospective TB disease classification for diagnostic studies.

Given the promise but uncertain performance of the Cepheid HRC in children, as well as the lack of dedicated studies of the technology in exclusively young children, we sought to evaluate the assay in our dedicated cohort of children under five years old with presumptive TB. To supplement the results reported by other groups, I analysed both the assay's diagnostic accuracy at defined cutoffs as well as its longitudinal performance after initiating TB treatment. To facilitate inter-study comparability, I present these analyses using both versions of the TB score.

## 3.3 Methods

*HRC Testing*

All testing of the HRC was performed prospectively and on-site by trained technicians. HRC testing was integrated into the NOD-pedFEND protocol mid-way through

the study, and all participants enrolled after the implementation date were given HRC testing per protocol.

Testing was performed on the first day of the baseline visit. In short, the participant's finger was pricked using a lance following standard practices for capillary blood collection, after which 100 µL of blood was collected in a minivette. The collected sample was loaded into the HRC within 15 minutes of collection and sent for testing at room temperature.

HRC tests were loaded onto the GeneXpert instrument and initiated within an hour from the time of blood collection. Once the test concluded, results were entered on a paper worksheet which would be later entered into REDCap. Additionally, the raw result files output by the GeneXpert instrument were saved. If invalid results were obtained, the test was repeated, if possible, which required a new sample that could be collected on the same or following day. In the case the repeat test was also invalid, the test was not repeated again. Raw result files for all tests, invalid and valid, were uploaded in batches to be converted into a readable format by Cepheid, which was then shared back with us for analysis.

*Data Cleaning*

The result files we received from Cepheid were those used for this chapter's analysis, as they contained individual gene Ct values in addition to *KLF2*- and *TBP*-based TB scores. Results for valid tests were only retained, however, if they matched those recorded on the paper worksheets at the time of testing, or if they did not match for a clear, explainable, reason (for instance, if one digit of a score was miswritten, but all other results

matched perfectly). In contrast, as it was later learned that invalid test results were inconsistently recorded on paper worksheets, all invalid results recorded by the GeneXpert instrument were included in the analysis unless they were unable to be confidently connected to a specific participant's sample (for example, if there was a typo in the entered participant identifier). As an additional quality control (QC) check, recorded blood collection times were compared to the recorded sample loading time and test initiation time. If the recorded times fell outside of the 15- and 60-minute windows, respectively, the result was not included in the analysis. Finally, test results were excluded if any other protocol deviations were noted, including if multiple tests with valid results were performed on the same participant sample.

*Data Analysis*

Compiled results were analysed in R v4.5.2 using tidyverse v2.0.0 for data wrangling and ggplot2 v4.0.3 for data visualization. pROC v1.19.0.1 was used for construction of the ROC curves, Rtsne v0.17 for the t-distributed Stochastic Neighbour Embedding (t-SNE) analysis, UpSetR v1.4.0 for the UpSet plot, and sjPlot v2.9.0 for regression visualizations.

For analysis of individual gene Ct values, gene expression was normalized against *TBP* expression by subtracting the *TBP* Ct value from the gene of interest's Ct value. TB scores and gene Ct values are presented as median values, with an accompanying IQR. Outside of the statistical methods detailed in **Section 2.8**, ANOVA was used in linear regression models to determine the significance of covariates in a review of HRC results from children with unconfirmed TB. The coefficient of determination ($R^2$) was used to measure correlations between test results.

For analysis of longitudinal changes in TB scores post-treatment, a subset of the HRC cohort was constructed which contained only those participants who had both a baseline and two-week HRC test performed, had started on TB treatment at baseline, and who did not have known drug-resistant TB. In this subset of the HRC cohort, the change in TB scores after two weeks was calculated by directly subtracting a participant's two-week HRC score from their baseline score, for both the *KLF2*- and *TBP*-based TB scores. The same calculation was performed for longitudinal changes in other investigated gene combinations. The Wilcoxon signed-rank test was used to determine the significance of changes to the TB scores over time. Additionally, AUCs were calculated for the diagnostic performance of the calculated difference in TB scores post-treatment initiation. A histogram of per-participant longitudinal changes in TB scores, coloured by NIH case classification, was also constructed to visually evaluate whether there appeared to be any bimodality in the distribution of longitudinal changes after TB treatment initiation within each case classification (which could potentially signify groups that did and did not respond to TB treatment).

Of note, given how the subsetting performed for the longitudinal analysis altered the proportion of NIH classifications included in certain reference standards (with differing proportions of confirmed and unconfirmed TB in the CRS+ group, for example), longitudinal analyses were performed using the NIH clinical classifications directly when possible, in lieu of the established reference standards.

# Results

## 3.4 The HRC Cohort

Of the 640 symptomatic participants, 430 (67%) were tested using the HRC. While 1 of the remaining 210 participants declined testing, the remaining 209 were all enrolled prior to the implementation of HRC testing into the NOD-pedFEND protocol. The resultant test records from NOD-pedFEND's Uganda sites, totalling 866 unique results from 430 participants, were recorded by the Xpert instrument for potential inclusion in this analysis. On review, 53 records were excluded, yielding 813 total test records for inclusion in the analysis (detailed in **Figure 1**). Reasons for exclusion included having two distinct, but valid, test results for a single sample without written explanation (n=8); valid test records not matching those manually entered in REDCap and therefore missing secondary verification (n=8); the listed sample tested not being capillary blood (n=3); a time from collection to cartridge loading which exceeded 15 minutes (n=3); and a time from collection to testing initiation which exceeded one hour (n=35). Two tests exceeded the maximum allowable duration from collection to both cartridge loading and testing initiation.

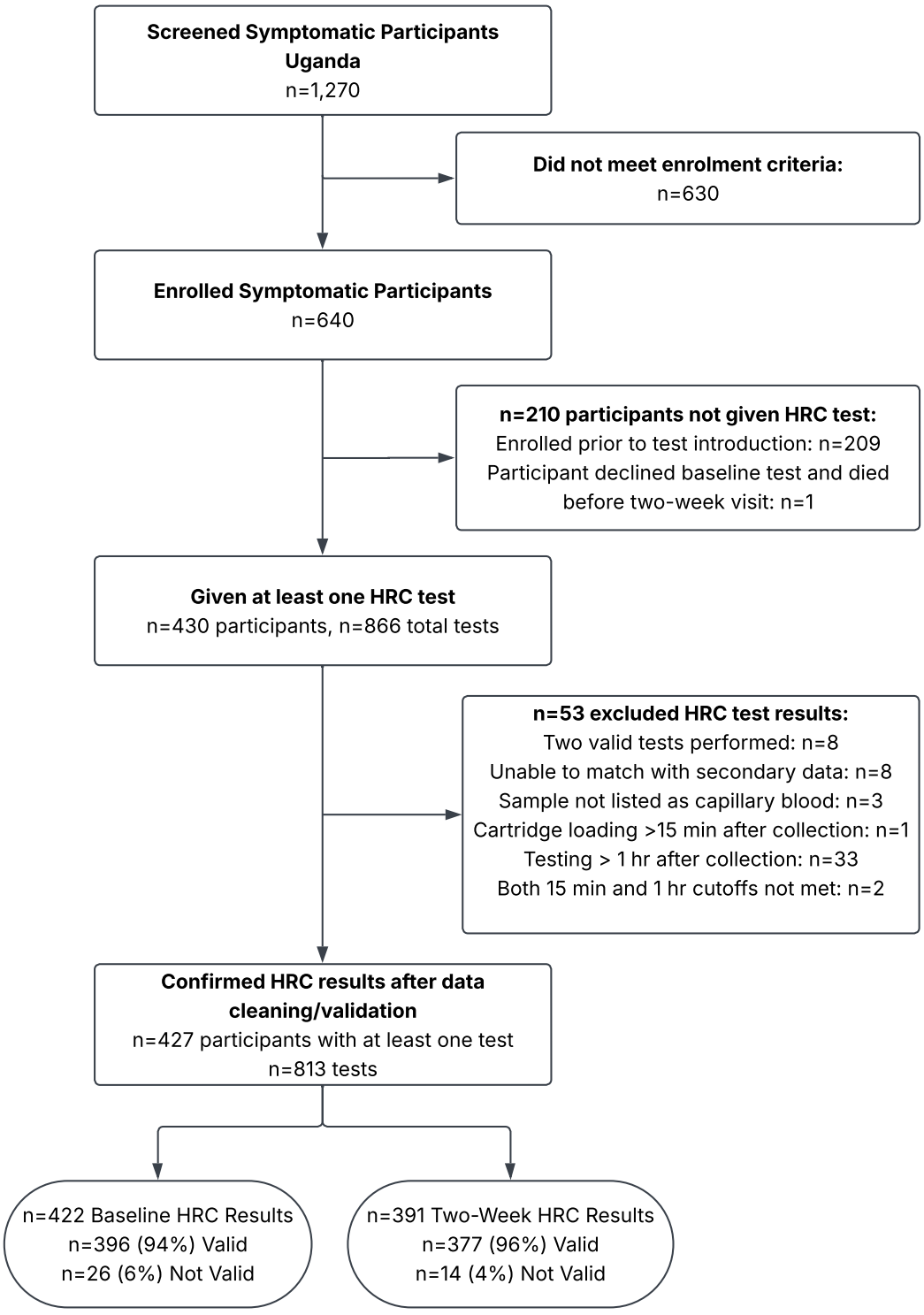



**Figure 1.** A diagram detailing the participants included in my analysis of the host response cartridge (HRC). Counts for total tests include both valid and invalid test results.

HRC testing was performed at the baseline and two-week visits. As HRC testing was introduced midway through the study period, the 813 tests included in this analysis were conducted on 427 children in total. Of these tests, 773/813 (95.1%) gave valid results: 396/422 (93.8%) at baseline and 377/391 (96.4%) at the two-week visit.

## 3.5 Diagnostic Performance of HRC Testing Performed at Baseline

400 children had HRC testing performed at baseline. Of these, 396 (99.0%) had valid test results, in part because non-valid tests were repeated, which resolved 24/26 (92.3%) of initial non-valid results. Characteristics of the 396 children, split by NIH case classification, are given in **Table 1**, along with their HRC results. Besides disease severity or *DUSP3* Ct values, no results or characteristics differed significantly between groups.

**Table 1.** Demographic info and host response cartridge (HRC) results of NOD-pedFEND participants with available and valid baseline HRC results. Data for a normalized *TBP* value is not given as normalization was performed against *TBP* expression. Significance between medians was determined using the Kruskal-Wallis rank sum test for continuous variables, and Pearson's chi-squared test with simulated p-values (with 2000 replicates) for categorical variables. Significance for categorical variables was calculated for the entire category across NIH classifications. "Unknown" values for categorical variables were excluded for Chi-squared tests. Interquartile ranges for median values are given in parentheses, with the 25$^{th}$ and 75$^{th}$ quartiles separated by a comma. Case classifications were determined as outlined in **Section 2.5**. Abbreviations: Severe Acute Malnutrition (SAM), Moderate Acute Malnutrition (MAM).

| | **Confirmed TB (n=19)** | **Unconfirmed TB (n=241)** | **Unlikely TB (n=88)** | **Unclassifiable TB status (n=48)** | **P Value (with/without unclassifiable)** |
|---|---|---|---|---|---|
| Age (in months) | 25 (15.5, 39) | 20 (12, 36) | 23 (13, 41.2) | 22.5 (13, 47) | 0.55, 0.45 |
| **Sex Distribution** | | | | | |
| % Female | 12/19 (63%) | 112/241 (46%) | 48/88 (55%) | 17/48 (35%) | 0.09, 0.20 |
| % Male | 7/19 (37%) | 129/241 (53.5%) | 40/88 (45%) | 31/48 (65%) | |
| **HIV Status** | | | | | |
| % HIV Positive | 4/19 (21%) | 36/241 (15%) | 5/88 (6%) | 7/48 (15%) | 0.12, 0.14 |
| % HIV Exposed-Uninfected | 2/19 (11%) | 42/241 (17%) | 14/88 (16%) | 3/48 (6%) | |
| % HIV Negative | 13/19 (68%) | 163/241 (68%) | 69/88 (78%) | 37/48 (77%) | |
| % HIV Status Unknown | 0/19 (0%) | 0/241 (0%) | 0/88 (0%) | 1/48 (2%) | |
| **Malnutrition Status** | | | | | |
| % with SAM | 7/19 (37%) | 76/241 (32%) | 27/88 (31%) | 14/48 (29%) | 0.35, 0.15 |
| % with MAM | 1/19 (5%) | 30/241 (12%) | 3/88 (3%) | 4/48 (8%) | |
| % Without Malnutrition | 9/19 (47%) | 133/241 (55%) | 55/88 (63%) | 27/48 (56%) | |
| % Malnutrition Status Unknown | 2/19 (11%) | 2/241 (1%) | 3/88 (3%) | 3/48 (6%) | |
| **Disease Severity** | | | | | |
| % with Severe Disease | 8/19 (42%) | 108/241 (45%) | 18/88 (20%) | 8/48 (17%) | <0.01, <0.01 |

| | | | | | |
|---|---|---|---|---|---|
| % with Non-Severe Disease | 8/19 (42%) | 115/241 (48%) | 3/88 (3%) | 4/48 (8%) | |
| % with Unclassifiable Disease Severity | 3/19 (16%) | 18/241 (7%) | 67/88 (76%) | 36/48 (75%) | |
| **HRC Data** | | | | | |
| Median (IQR) KLF2-based TB Score | 2.48 (1.73, 3.00) | 2.50 (1.98, 2.96) | 2.65 (2.11, 3.15) | 2.71 (2.19, 2.99) | 0.23, 0.16 |
| Median (IQR) TBP-based TB Score | -0.9 (-1.52, -0.5) | -0.9 (-1.35, -0.4) | -0.75 (-1.11, -0.4) | -0.7 (-1.25, -0.35) | 0.42, 0.26 |
| Median (IQR) normalized KLF2 Ct value | -3.3 (-3.4, -3.1) | -3.3 (-3.5, -3.1) | -3.4 (-3.7, -3.1) | -3.4 (-3.7, -3.1) | 0.36, 0.28 |
| Median (IQR) normalized DUSP3 Ct value | -0.3 (-0.8, 0.05) | 0 (-0.3, 0.5) | 0.1 (-0.2, 0.5) | -0.05 (-0.53, 0.4) | 0.04, 0.03 |
| Median (IQR) normalized GBP5 Ct value | -1.7 (-2.5, -0.65) | -1.8 (-2.6, -1.1) | -1.6 (-2.3, -1.1) | -1.5 (-2.03, -0.95) | 0.38, 0.51 |

*Median HRC Scores*

HRC results, besides the normalized *DUSP3* Ct value, did not significantly differ between case classifications (**Figure 2**). The normalized *DUSP3* Ct value was -0.30 (IQR -0.80, 0.05) for children with confirmed TB, 0.00 (IQR -0.30, 0.50) for children with unconfirmed TB, and 0.10 (IQR -0.20, 0.50) for children with unlikely TB. However, the median *KLF2*-based TB score, which includes *DUSP3* in its calculation, was similar across case classifications: 2.48 (IQR 1.73, 3.00) for the 19 children with confirmed TB, 2.50 (IQR 1.98, 2.96) for the 241 children with unconfirmed TB, 2.65 (IQR 2.11, 3.15) for the 88 with unlikely TB, and 2.71 (IQR 2.19, 2.99) for the 48 participants with unclassifiable TB status.

The median *TBP*-based TB score was -0.90 (IQR -1.52, -0.50) for confirmed TB, -0.90 (IQR -1.35, -0.40) for unconfirmed TB, -0.75 (IQR -1.11, -0.40) for unlikely TB, and -0.70 (IQR -1.25, -0.35) for participants with unclassifiable TB status. I also determined the median TB scores of CLHIV who were LAM+, a cohort which constituted one child with confirmed TB, nine with unconfirmed TB, and one with unlikely TB. For this group, the median *KLF2*-based TB score was 1.66 (IQR 1.54, 2.35), and *TBP*-based TB score was -1.55 (IQR -1.93, -1.08).

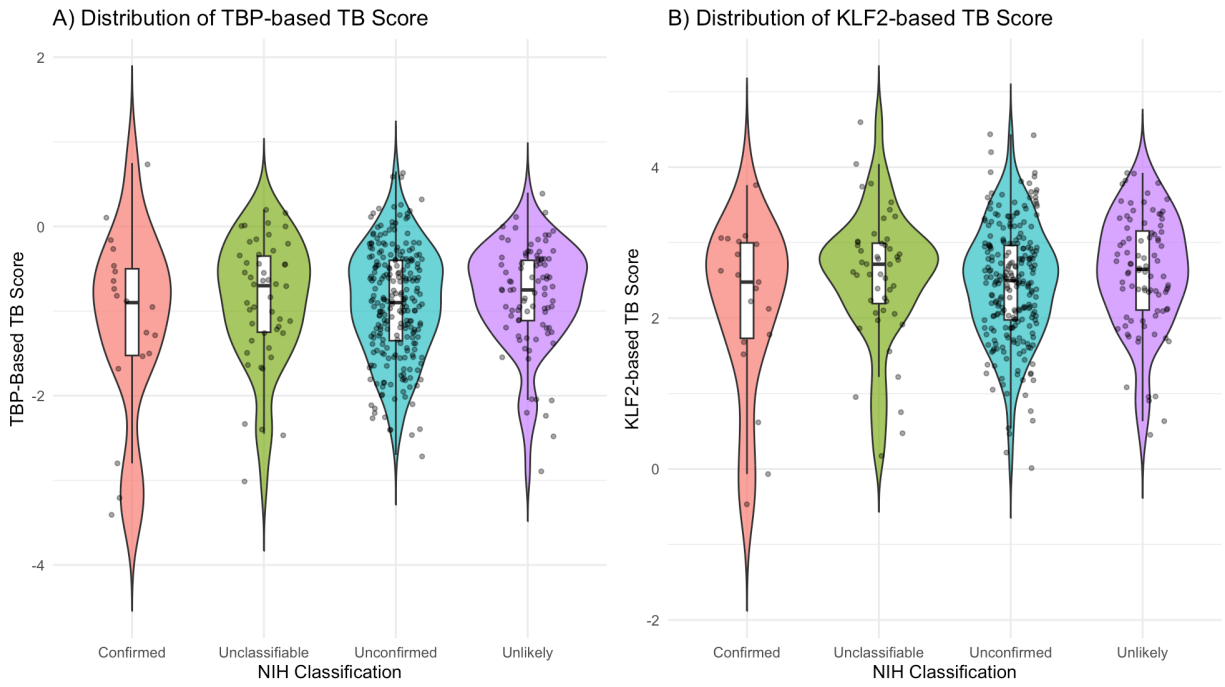



**Figure 2.** A violin plot showing the distribution of A) *TBP*- and B) *KLF2*-based TB scores at baseline, per NIH classification, for all participants with valid host response cartridge results. Case classifications were determined as outlined in **Section 2.5**.

*Diagnostic Performance of the HRC*

Performance of the HRC was evaluated against the microbiological and composite reference standards previously described in **Section 2.6**. Against the MRS, the *TBP*-based TB score yielded an AUC of 0.55 (95% Confidence Interval (CI) 0.39-0.69); using the *KLF2*-based TB score, AUC was 0.56 (95% CI 0.42-0.69). Against the CRS, results were similar, with the *TBP*-based TB score yielding an AUC of 0.56 (95% CI 0.49-0.62), and the *KLF2*-based TB score yielding an AUC of 0.57 (95% CI 0.50-0.63). CIs were calculated using 2000 bootstrap resamples. The ROC curves for the evaluations which yielded these AUCs are given in **Figure 3.**

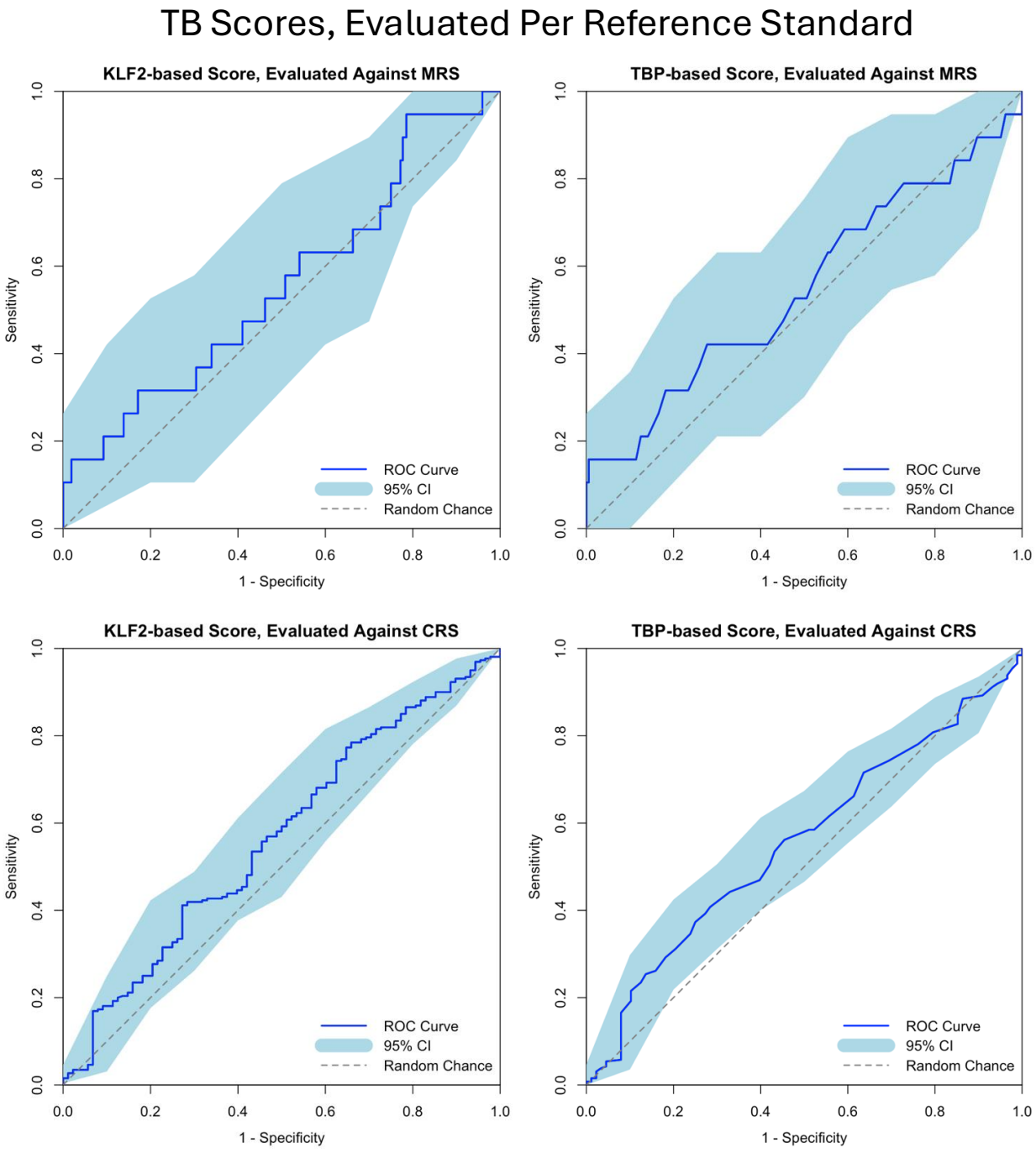



**Figure 3.** ROC curves for both the *KLF2*- and *TBP*-based TB scores, calculated using the microbiological (MRS) and composite reference standards (CRS). Participants with unknown statuses for a given reference standard were excluded from the plot. Reference standards were defined as detailed in **Section 2.6**.

Sensitivity and specificity were evaluated at multiple TB score cutoffs. When the *KLF2*-based TB score was set to the cutoff of 1.50 established for adults, this yielded a sensitivity and specificity of 16% and 91% against the MRS, and 10% and 93% against the CRS, respectively. A *TBP*-based TB score cutoff has not been similarly established; however, an optimal cutoff of -2.75 in our cohort was derived using the Youden's index, as applied to the microbiological reference standard. When set to this cutoff, the *TBP*-based TB score yielded a sensitivity and specificity of 16% and 100% against the MRS and 1% and 99% against the CRS, respectively. For the *KLF2*-based score, the optimal cutoff was calculated as 3.10 using the Youden's index applied to the MRS; using this cutoff, sensitivity and specificity was 95% and 22% against the MRS and 82% and 28% against the CRS, respectively. These diagnostic accuracy results, along with those calculated at alternative cutoff scores proposed by other studies, are summarized in **Table 2.**

**Table 2.** The diagnostic performance of the HRC's TB Scores when compared against both a microbiological reference standard and a composite reference standard. Confidence intervals were determined using bootstrapping with 2000 resamples. Reference standards are outlined in **Section 2.6**.

| | *Microbiological Reference Standard* | | *Composite Reference Standard* | |
|---|---|---|---|---|
| | **Sensitivity** | **Specificity** | **Sensitivity** | **Specificity** |
| **Thresholds** | *KLF2*-based TB Scores | | | |
| 1.5 (often-used cutoff) | 16% (0%-32%) | 91% (88%-94%) | 10% (7%-14%) | 93% (88%-98%) |
| 1.99 (optimal cutoff suggested by Olbrich et al.) | 32% (11%-53%) | 76% (71%-80%) | 27% (22%-32%) | 80% (71%-88%) |
| 2.76 (cutoff which gave Olbrich et al. 90% sensitivity) | 63% (42%-84%) | 39% (35%-44%) | 64% (57%-69%) | 43% (33%-53%) |
| 3.10 (Youden's Index Suggested Cutoff) | 95% (84%-100%) | 22% (17%-26%) | 82% (77%-86%) | 28% (19%-38%) |
| **Thresholds** | *TBP*-based TB Scores | | | |
| -2.75 (Youden's Index Suggested Cutoff) | 16% (0%-32%) | 100% (99%-100%) | 1% (0%-3%) | 99% (97%-100%) |
| -0.65 (cutoff used by COMBO study [86]) | 68% (47%-90%) | 41% (36%-46%) | 62% (55%-67%) | 44% (34%-55%) |

*Subgroup-Specific Performance*

While my ability to assess subgroup-specific diagnostic performance at baseline was limited due to having only 19 MRS+ cases with valid baseline HRC results, I

nevertheless calculated AUCs for both TB scores in subsets of the cohort to explore whether performance appeared to be improved in specific cohorts of interest (**Table 3**). These exploratory evaluations were performed against both the MRS and CRS. No significantly improved performance was observed in the subgroups tested, which consisted of CLHIV, CHEU, or children who were female, older than two, demonstrating severe manifestations of disease, or presenting with SAM. However, performance of the HRC in children with SAM, when evaluated against the CRS (AUC = 0.65 (*KLF2*) and AUC = 0.63 (*TBP*)) appeared to be slightly improved when compared to overall performance evaluated against the same reference standard (AUC = 0.57 (*KLF2*) and AUC = 0.56 (*TBP*)). Similarly, performance of the HRC in older children (between the ages of two and five), when evaluated against the MRS (AUC = 0.63 (*KLF2*) and AUC = 0.61 (*TBP*)) appeared to be slightly improved when compared to overall performance evaluated against the same reference standard (AUC = 0.56 (*KLF2*) and AUC = 0.55 (*TBP*)). Of note, some participants were included in more than one subgroup. To illustrate this overlap between subgroups, an UpSet plot was constructed (**Figure 4**).

**Table 3.** AUCs for the diagnostic performance of HRC overall and in six subgroups of interest. Confidence intervals for AUCs were determined by 2000 bootstrap resamples. Reference standards are outlined in **Section 2.6**. Abbreviations: Children who are HIV-Exposed but Uninfected (CHEU), Children Living with HIV (CLHIV), Severe Acute Malnutrition (SAM).

| | **Microbiological Reference Standard (MRS)** | | **Composite Reference Standard (CRS)** | |
|---|---|---|---|---|
| | ***KLF2*-based TB Score** | ***TBP*-based TB Score** | ***KLF2*-based TB Score** | ***TBP*-based TB Score** |
| *Overall* | 0.56 (0.42-0.69) | 0.55 (0.39-0.69) | 0.57 (0.50-0.63) | 0.56 (0.49-0.62) |
| *Age ≥ two years n=12 MRS+, n=166 MRS- n=118 CRS+, n=42 CRS-* | 0.63 (0.47-0.78) | 0.61 (0.44-0.76) | 0.59 (0.49-0.69) | 0.59 (0.49-0.68) |
| *Female n=12 MRS+, n=176 MRS- n=124 CRS+, n=48 CRS-* | 0.57 (0.41-0.74) | 0.55 (0.37-0.74) | 0.62 (0.52-0.72) | 0.60 (0.50-0.69) |
| *CHEU n=2 MRS+, n=59 MRS- n=44 CRS+, n=14 CRS-* | 0.59 (0.15-0.98) | 0.52 (0.02-1.00) | 0.58 (0.41-0.74) | 0.54 (0.39-0.69) |
| *CLHIV n=4 MRS+, n=48 MRS- n=40 CRS+, n=5 CRS-* | 0.59 (0.25-0.87) | 0.58 (0.21-0.85) | 0.58 (0.40-0.75) | 0.50 (0.28-0.69) |
| *Children with SAM n=7 MRS+, n=115 MRS- n=83 CRS+, n=27 CRS-* | 0.48 (0.23-0.73) | 0.48 (0.20-0.75) | 0.65 (0.53-0.76) | 0.63 (0.50-0.74) |
| *Children with Severe Disease n=8 MRS+, n=134 MRS- n=116 CRS+, n=18 CRS-* | 0.58 (0.33-0.82) | 0.58 (0.31-0.85) | 0.56 (0.42-0.70) | 0.53 (0.40-0.65) |

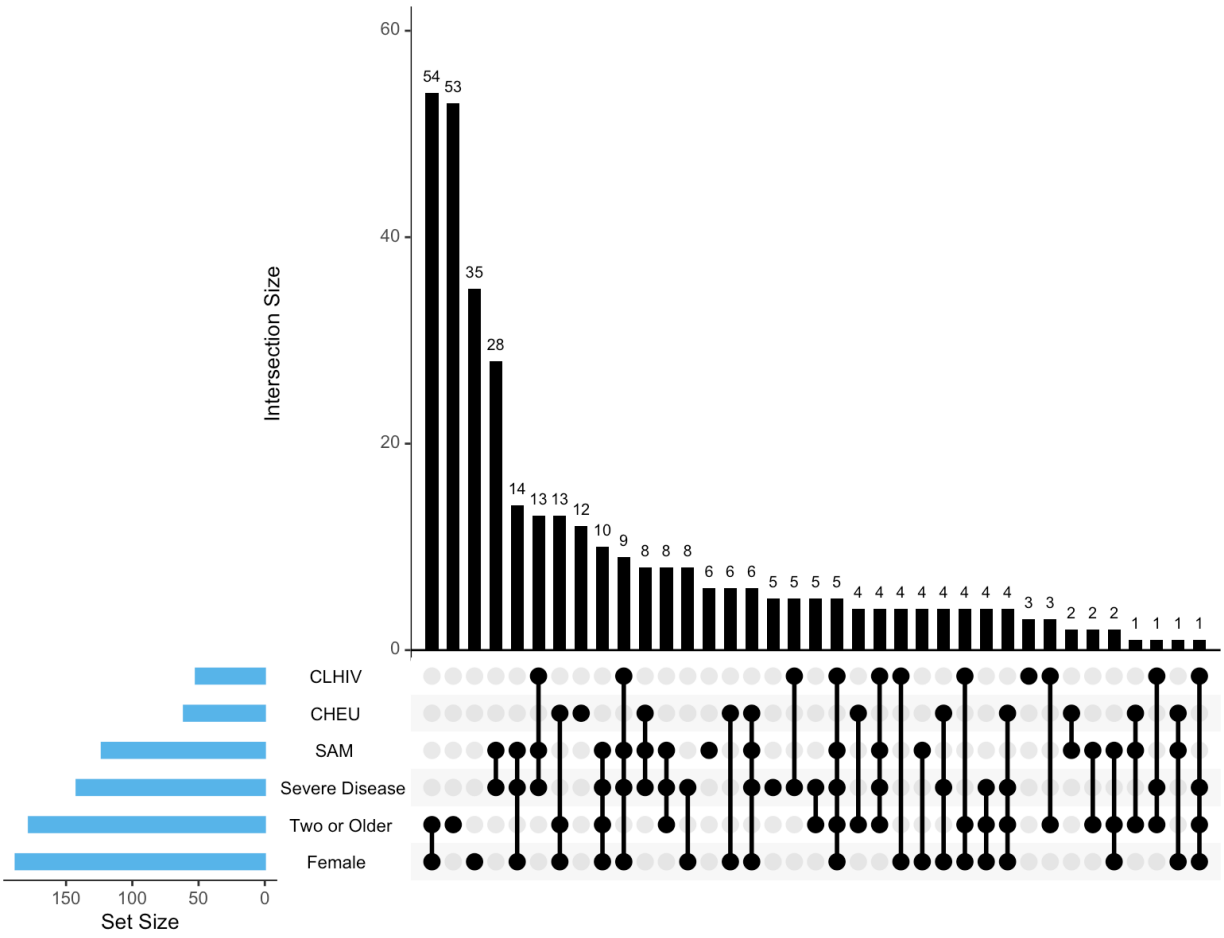



**Figure 4**. An UpSet plot demonstrating the overlap in subgroups for the host response cartridge analysis. Individuals unclassifiable by both the composite and microbiological reference standard are not included; however, individuals unclassifiable by only one reference standard are included. Reference standards were defined as detailed in **Section 2.6**. Abbreviations: Children Living with HIV (CLHIV), Children who are HIV-Exposed but Uninfected (CHEU), Severe Acute Malnutrition (SAM).

Additionally, the high number of participants with unconfirmed TB in our cohort allowed for intra-group evaluation of scoring changes in subgroups of interest, including across age, sex, HIV status, disease severity, and malnutrition status via linear regression models (**Figures 5** and **6).** The unconfirmed TB group was chosen for this analysis as it contained a majority of all participants and enabled an interrogation of covariates without

the potential confounding influence of participant TB status. After performing this analysis for the *KLF2*-based TB score, the only significant factor influencing the TB score, among those tested, was HIV status, which was 0.45 units higher among CHEU versus CLHIV ($p=0.01$), and 0.36 units higher among HIV-uninfected children versus CLHIV ($p=0.01$). However, the model only attributed HIV status to 3.2% of the variation in scores. For the *TBP*-based TB score, a similar influence of HIV status was observed, with the TB score being 0.42 units higher among CHEU ($p<0.01$) versus CLHIV, and 0.29 units higher among HIV-uninfected children versus CLHIV ($p=0.02$), which accounted for 3.6% of the variation in scores. Additionally, age had significant influence over the *TBP*-based TB score (but not the *KLF2*-based score), with the model finding the *TBP*-based score to decrease by 0.01 units for every additional month in the child's age ($p=0.02$), which accounted for 2.5% of the variation in scores. Otherwise, no significant influence of the evaluated factors on TB scores, amongst the 241 children with unconfirmed TB, was observed.

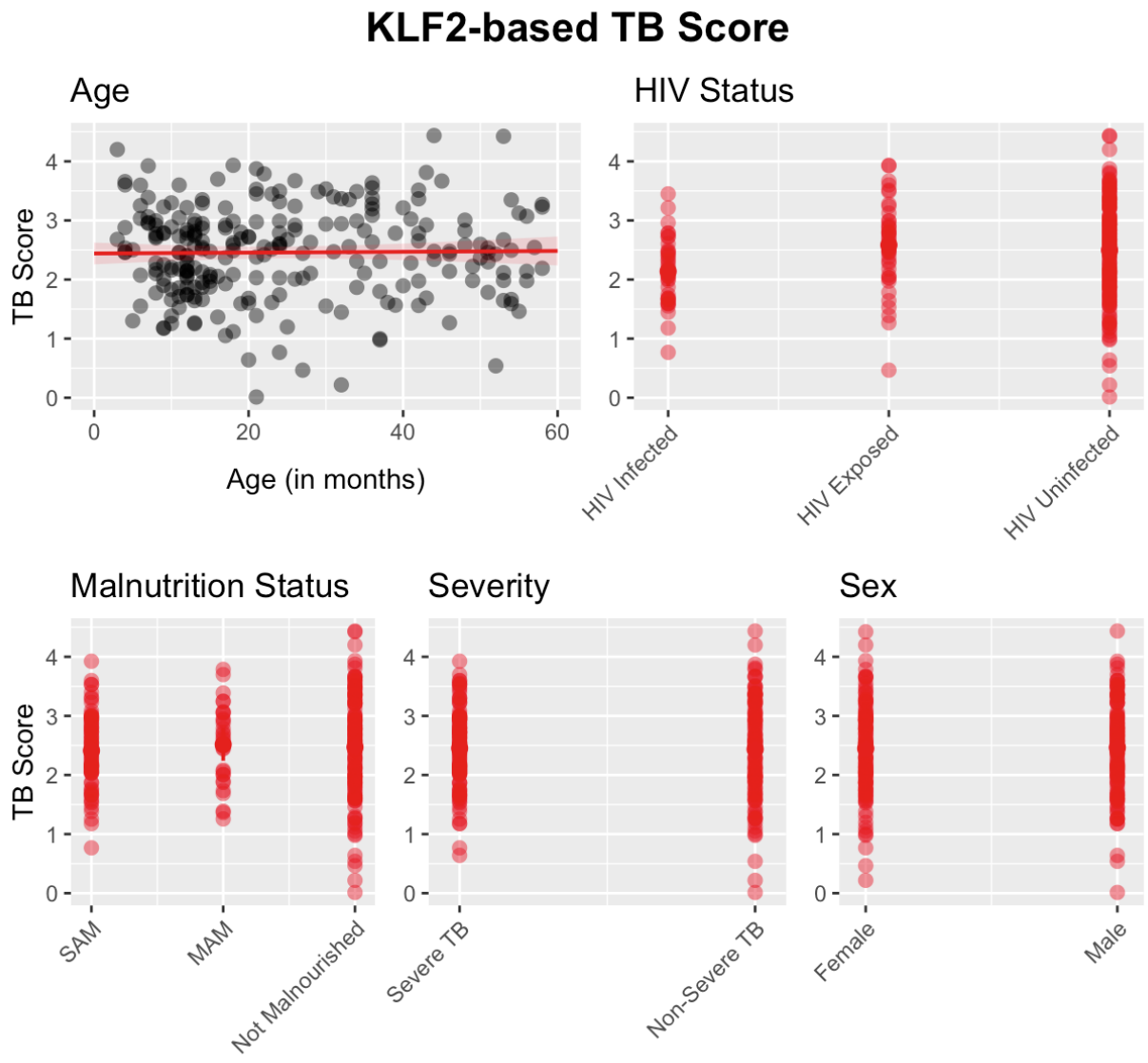



**Figure 5.** Linear regression models of the *KLF2*-based TB score amongst participants with unconfirmed TB. Case classifications were determined as detailed in **Section 2.5**. Abbreviations: Severe Acute Malnutrition (SAM), Moderate Acute Malnutrition (MAM).

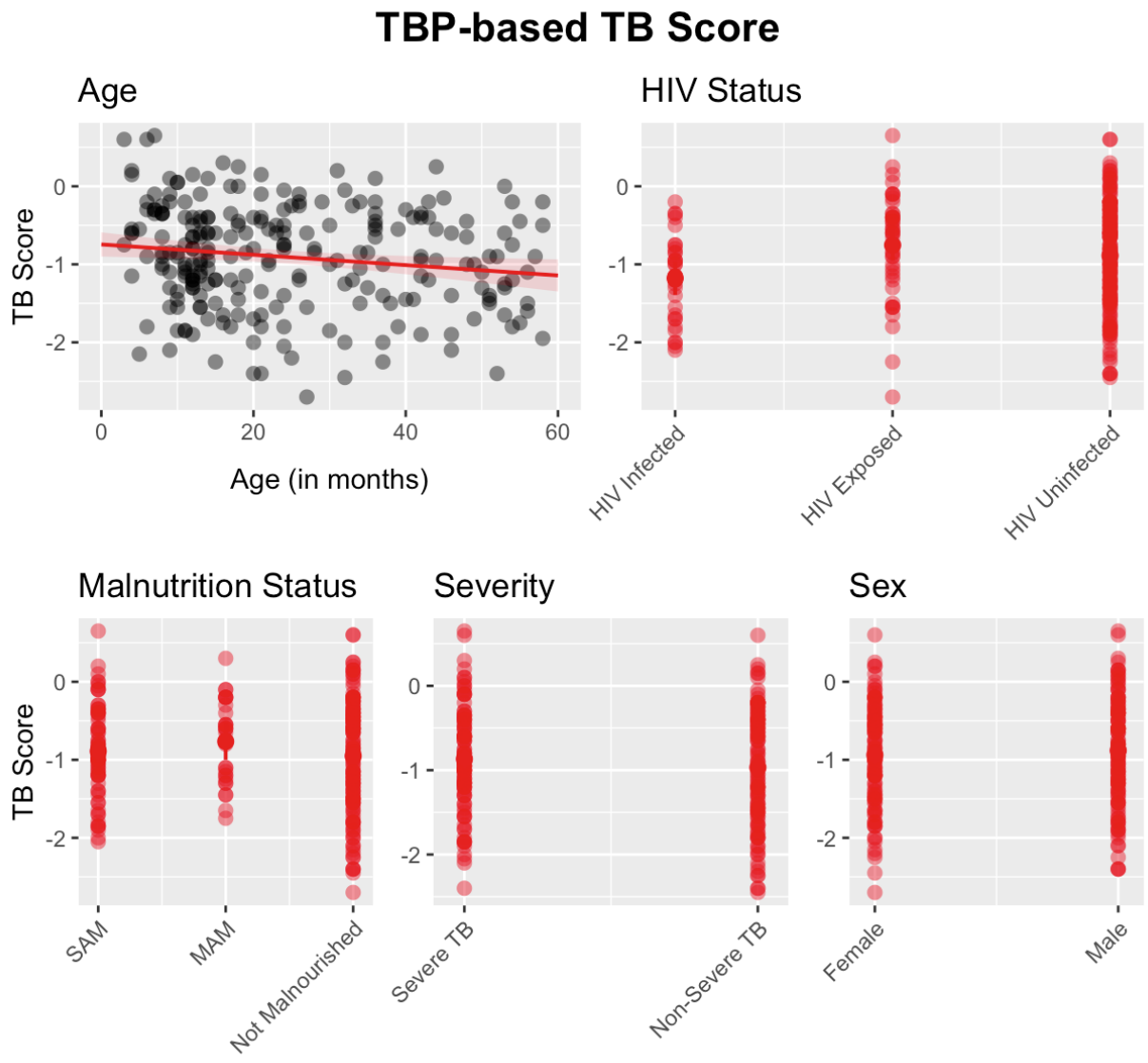



**Figure 6.** Linear regression models of the *TBP*-based TB score amongst participants with unconfirmed TB. Case classifications were determined as detailed in **Section 2.5**. Abbreviations: Severe Acute Malnutrition (SAM), Moderate Acute Malnutrition (MAM).

### 3.6 Longitudinal Changes to the HRC Score after TB Treatment

We also sought to evaluate how HRC scores changed after two weeks of TB treatment. For this analysis, I reviewed the TB scores of the 221 participants who were started on TB treatment at baseline, did not have known drug-resistant TB, and who had an

HRC test performed at both their baseline visit and a visit two weeks later. Across all participants, TB scores changed minimally between the baseline and two-week visits, with a median difference of -0.01 for the *KLF2*-based score and +0.05 for the *TBP*-based score; these differences were not statistically significant by a Wilcoxon Signed-Rank test at p=0.57 and p=0.35, respectively. The median individual TB score change and its significance, calculated within each clinical classification, is given in **Table 4.**

**Table 4.** The median change in TB scores between an HRC test taken at the baseline and two-week visits, for the 221 selected participants. Significance was calculated using a Wilcoxon Signed-Rank test. Case classifications were defined as detailed in **Section 2.5**.

| | **Overall (n=221)** | **Confirmed TB (n=16)** | **Unconfirmed TB (n=179)** | **Unlikely TB (n=20)** | **Unclassifiable (n=6)** |
|---|---|---|---|---|---|
| *Median Difference Across Timepoints, TBP-based Score* | 0.05 (p=0.35) | 0.10 (p=0.94) | 0.05 (p=0.26) | -0.03 (p=0.78) | -0.28 (p=1.00) |
| *Median Difference Across Timepoints, KLF2-based Score* | -0.01 (p=0.57) | -0.05 (p=0.86) | -0.02 (p=0.56) | 0.09 (p=0.45) | -0.42 (p=0.16) |

Given the uncertainty of case classifications in paediatric TB, particularly in the heterogenous “unconfirmed TB” group, I also evaluated individual-level changes in TB scores to look to see if a subset of participants exhibited more appreciable changes in their TB scores during treatment. The evolution of TB scores for each participant is shown in **Figures 7** and **8**, while **Figure 9** shows the distribution of participants’ change in TB scores

between their baseline and week two visits**.** Qualitatively, no clear subgroups of HRC-indicated responders/non-responders to treatment were observed upon visual inspection.

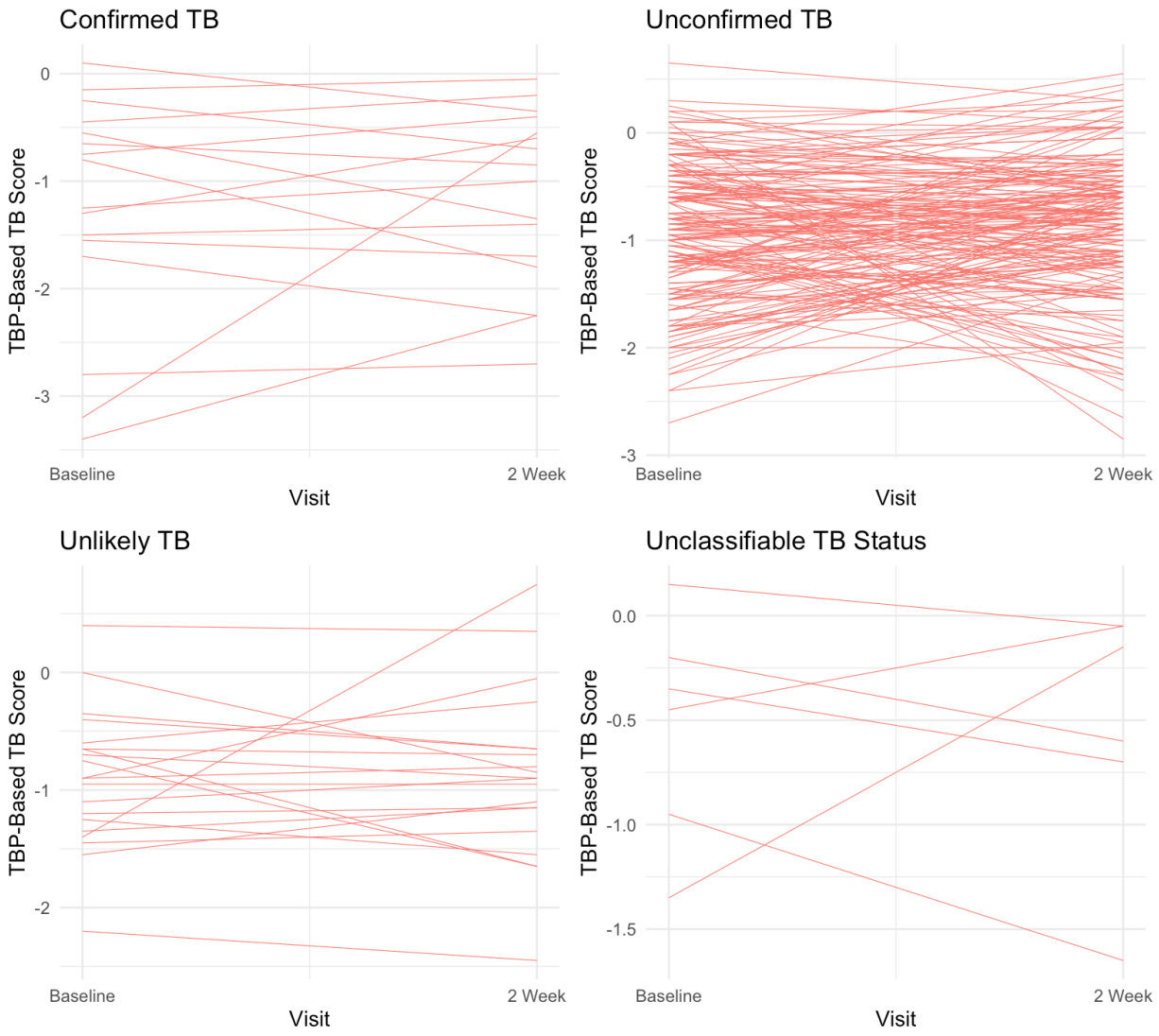



**Figure 7.** Line graphs showing the per-participant change in *TBP*-based TB scores from the baseline to two-week host response cartridge tests, split per NIH classification. Case classifications were determined as detailed in **Section 2.5**.

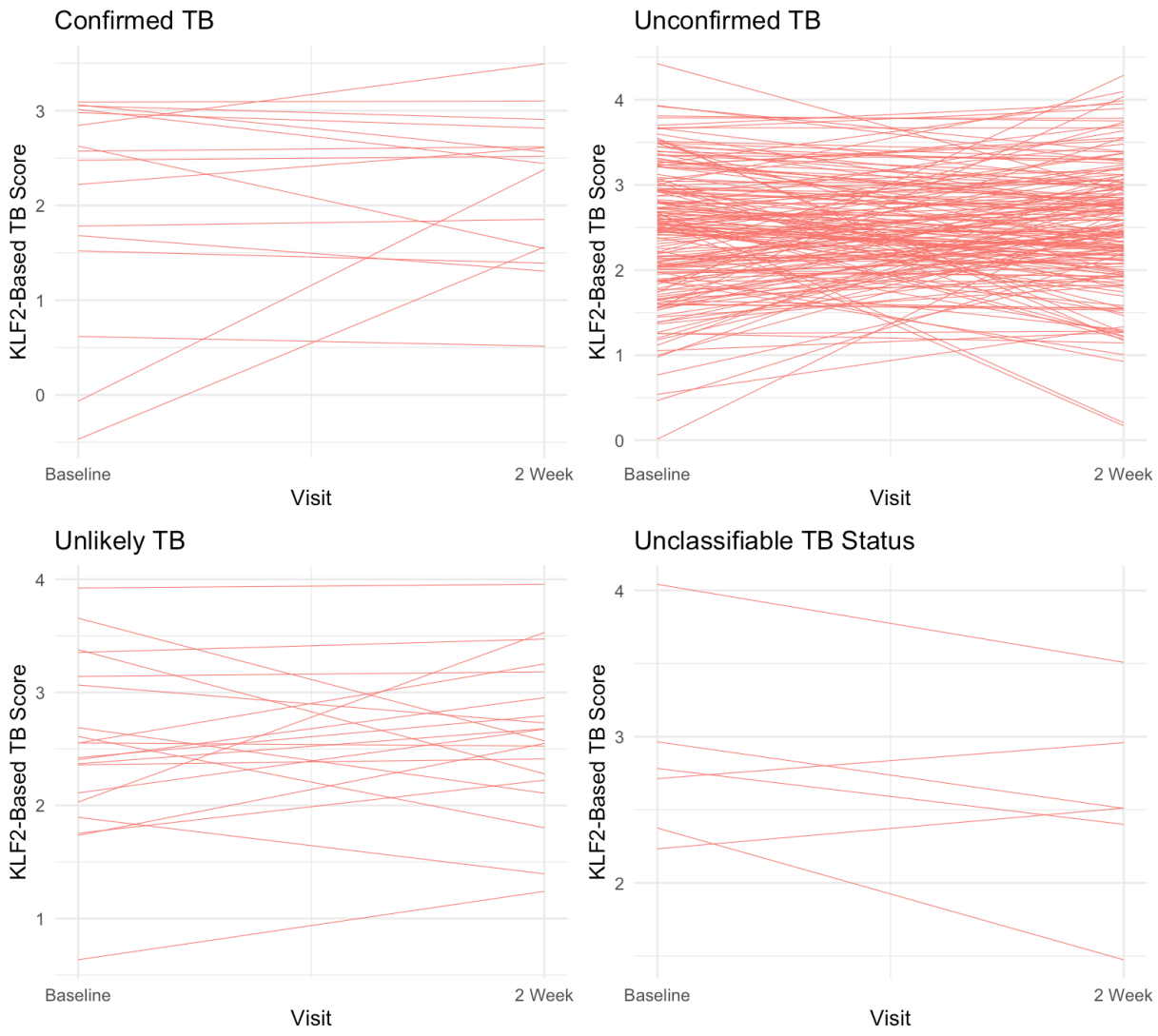



**Figure 8.** Line graphs showing the per-participant change in *KLF2*-based TB scores from the baseline to two-week host response cartridge tests, split per NIH classification. Case classifications were determined as detailed in **Section 2.5**.

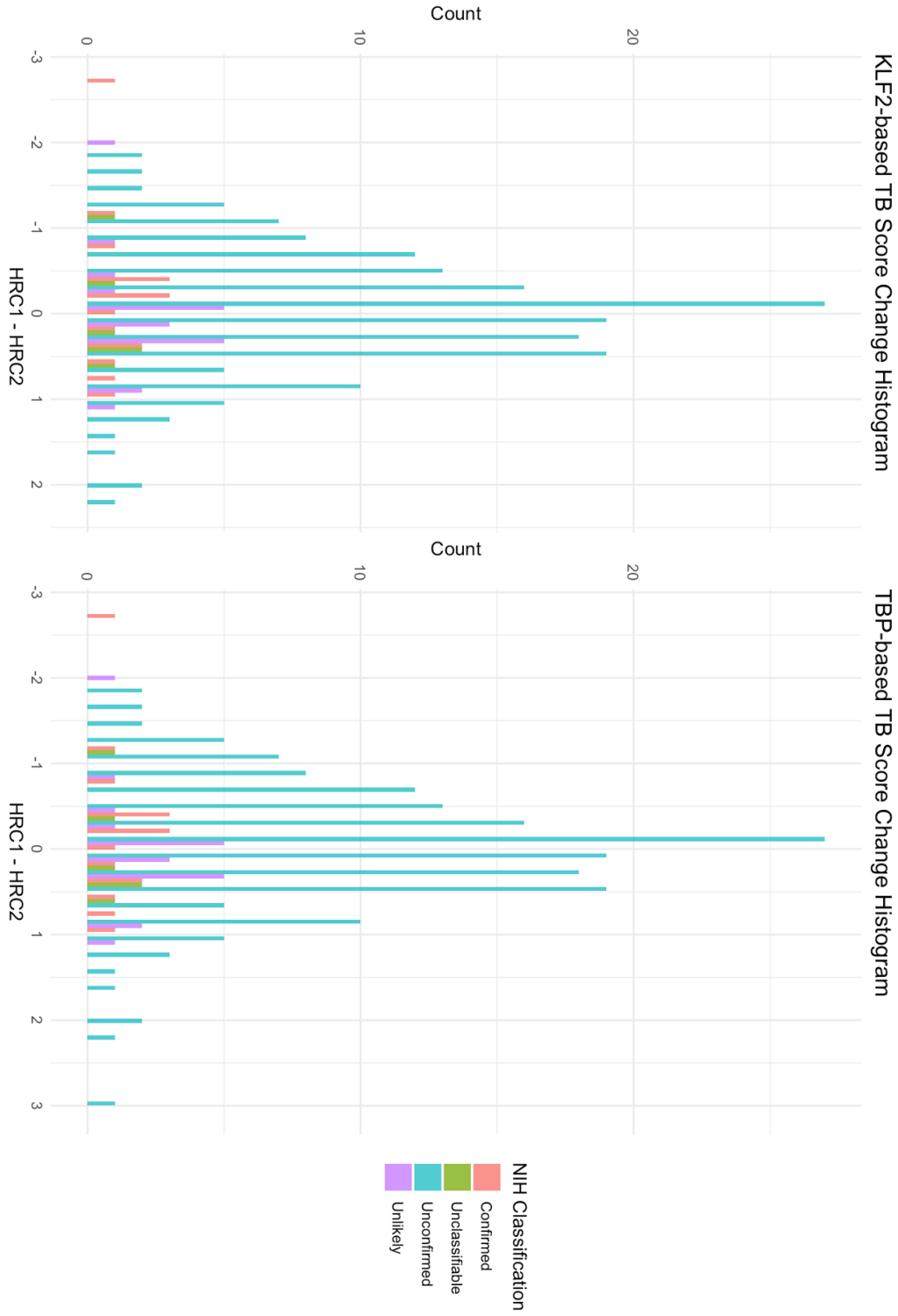



**Figure 9.** Histograms showing the distribution of the difference in baseline to two-week TB scores amongst the analysed participants, coloured by NIH case classification. Case classifications were determined as detailed in **Section 2.5**.

*Longitudinal Changes as a Retrospective Diagnostic*

To quantify the diagnostic performance of the calculated difference in TB scores after two weeks of treatment, I evaluated the ability of the change in *KLF2*- or *TBP*-based TB scores to retrospectively classify children on the basis of their CRS or MRS status **(Table 5)**. Retrospective changes in TB scores were not significantly better at predicting children's reference standard status than the TB scores determined at baseline.

**Table 5.** AUCs for the diagnostic performance of longitudinal changes in the TB scores in retrospectively differentiating reference standard positive individuals from those who are reference standard negative. 95% confidence intervals were determined using 2000 bootstrap resamples. Reference standards were defined as detailed in **Section 2.6**.

| | ***KLF2*-based TB Score** | ***TBP*-based TB Score** |
|---|---|---|
| *Microbiological Reference Standard* | 0.50 (0.36-0.63) | 0.49 (0.35-0.64) |
| *Composite Reference Standard* | 0.53 (0.40-0.65) | 0.54 (0.42-0.66) |

## 3.7 Correlation Between HRC Scoring Approaches

In addition to comparing the diagnostic performance of both the *KLF2*- and *TBP*-based signatures, I also sought to compare the two scoring methodologies themselves. To accomplish this, I plotted the paired *KLF2*- and *TBP*-based TB scores for each individual HRC test performed and determined the correlation between the two scoring approaches. This is particularly relevant for the comparison of our study's results with those reported in the prior literature, which primarily reported on the performance of the *KLF2*-based score.

Looking at the results of 773 valid tests, performed at both baseline and the two-week visit, the *KLF2*- and *TBP*-based scores were moderately correlated, with $R^2$=0.77 (**Figure 10)**. Of note, the *TBP*-based TB score is only reported in increments of 0.05 despite only spanning from -3.40 to 0.75 in our cohort, while the *KLF2*-based TB score spans an even larger range of -0.55 to 4.60 and is reported in increments of 0.001.

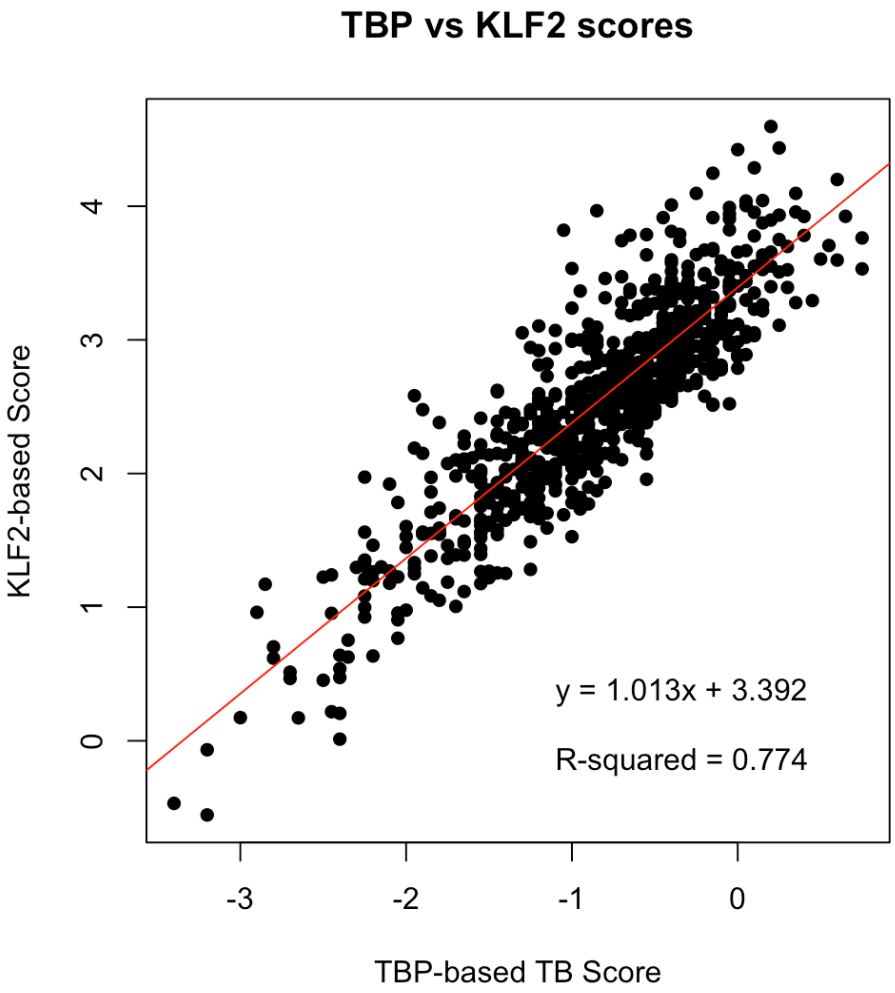



**Figure 10.** A comparison of the scoring output for both the *KLF2*- and *TBP*-based TB scores originating from the same tests.

## 3.8 Alternative Scoring Methodologies

The TB score is the primary output of the HRC and is the metric designed to be used for TB diagnosis per manufacturer guidance. However, the HRC also records the cycle threshold (Ct) values for each of the four genes tested (*KLF2* and *TBP* are both tested in each cartridge), along with within-instrument controls. As such, there is additional data generated from each HRC which could have enhanced diagnostic utility in paediatric populations. Recognizing this, I performed a few exploratory analyses to look for alternative scoring methodologies which might better diagnose paediatric TB.

*Diagnostic Performance—Individual Genes*

First, I measured the diagnostic performance of each gene individually, based on its Ct value. As Ct values can be influenced by RNA stability and concentration, I normalized them by subtracting the Ct value of the housekeeping gene *TBP*. The HRC test conducted at baseline was used for this analysis. AUCs (detailed in **Table 6** for each gene and reference standard) were not meaningfully improved for any individual normalized gene over the performance of the summary TB scores. *DUSP3* alone, however, was notable for having an AUC 0.10 units higher than the *KLF2*-based score and 0.11 units higher than the *TBP*-based score, when evaluated against the MRS. The diagnostic performances of each gene alone are graphically demonstrated in the ROC curves shown in **Figures 11** and **12**.

**Table 6.** AUCs for the performance of normalized individual gene Ct values in differentiating reference standard positive individuals from those who are reference standard negative. 95% confidence intervals were determined using 2000 bootstrap resamples. Reference standards were defined as detailed in **Section 2.6**.

| | **Normalized *KLF2* Ct** | **Normalized *DUSP3* Ct** | **Normalized *GBP5* Ct** | ***TBP* Ct** |
|---|---|---|---|---|
| *Microbiological Reference Standard* | 0.57 (0.46-0.68) | 0.66 (0.51-0.79) | 0.52 (0.36-0.67) | 0.55 (0.44-0.67) |
| *Composite Reference Standard* | 0.55 (0.47-0.62) | 0.55 (0.49-0.62) | 0.54 (0.47-0.60) | 0.52 (0.45-0.60) |

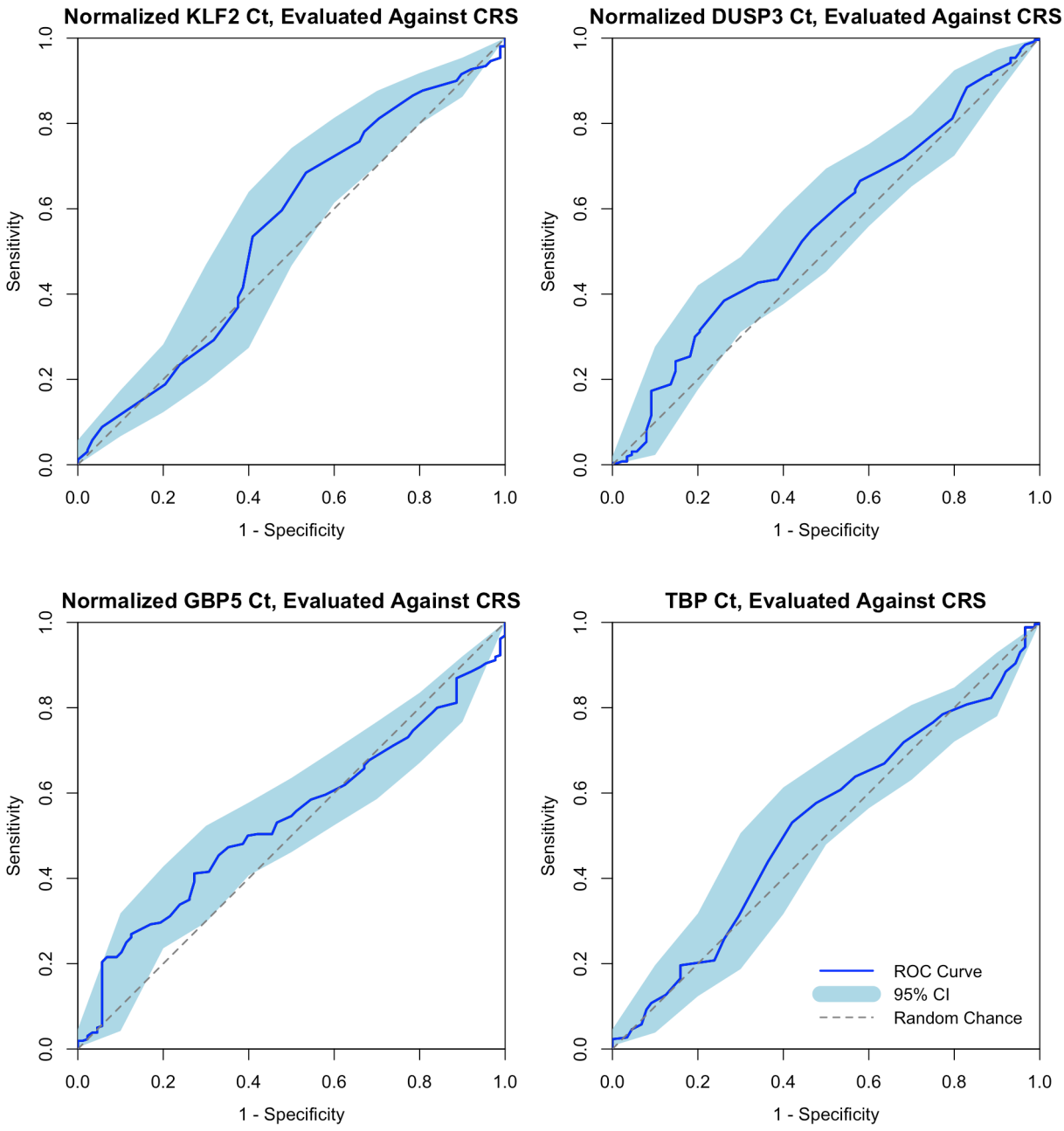



**Figure 11.** ROC curves showing the performance of individual gene cycle threshold (Ct) values in differentiating composite reference standard (CRS)+ individuals from those who are CRS-. 95% confidence intervals were determined using 2000 bootstrap resamples. Reference standards were defined as detailed in **Section 2.6**.

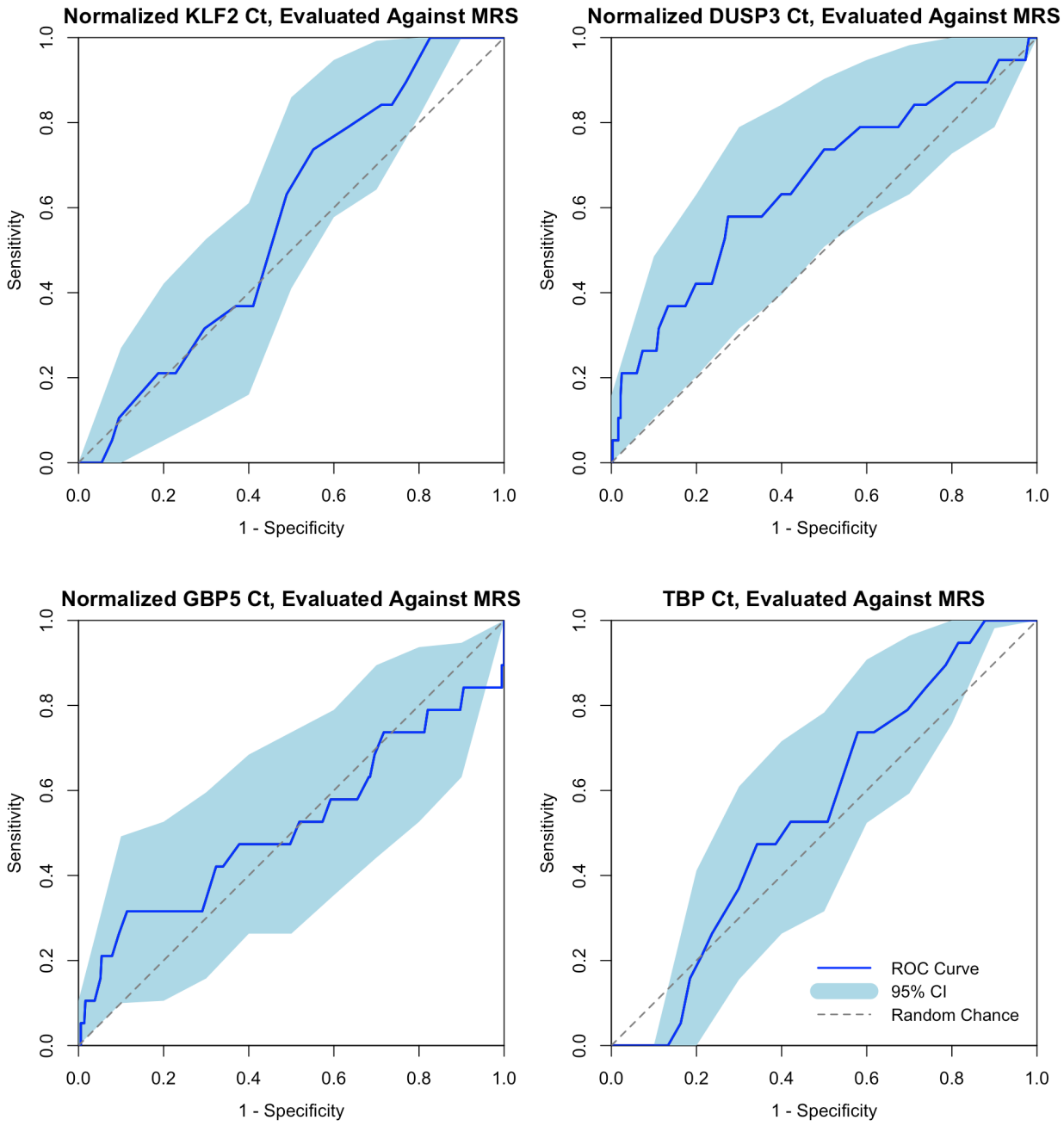



**Figure 12.** ROC curves showing the performance of individual gene cycle threshold (Ct) values in differentiating MRS+ individuals from those who are microbiological reference standard (MRS)-. 95% confidence intervals were determined using 2000 bootstrap resamples. Reference standards were defined as detailed in **Section 2.6**.

*Diagnostic Performance—Pairwise Scoring*

I also evaluated the performance of the differences in a pair of genes' normalized Ct values to diagnose TB, which approximates the relative expression between the two genes. I evaluated the differences between all non-housekeeping gene pairs, which were calculated as *GBP5-KLF2*, *DUSP3-KLF2*, and *DUSP3-GBP5*. Both *GBP5-KLF2* and *DUSP3-GBP5* did not perform better than the overall TB scores, as shown by the AUCs reported in **Table 7.** *DUSP3-KLF2*, similar to *DUSP3* alone, however, did perform significantly better than a random classifier, unlike either version of the TB score. Compared head-to-head, however, the confidence intervals for the AUC of the *DUSP3-KLF2* difference (0.67, 95% CI 0.55-0.79), like baseline *DUSP3* alone (0.66, 95% CI 0.51-0.79), substantially overlapped with those for the performance of the baseline *KLF2*-based (0.56, 95% CI 0.42-0.69) and *TBP*-based (0.55, 95% CI 0.39-0.69) TB scores.

**Table 7.** The AUCs of the difference in Ct values between two of the normalized HRC cartridge genes—a proxy for the ratio between the two genes' expression levels—for differentiating participants on the basis of reference standard positivity. 95% confidence intervals are given in parentheses and determined by bootstrapping with 2000 resamples. Reference standards were defined as outlined in **Section 2.6**.

| | ***GBP5 - KLF2*** | ***DUSP3 - KLF2*** | ***DUSP3 - GBP5*** |
|---|---|---|---|
| *Microbiological Reference Standard* | 0.49 (0.35-0.65) | 0.67 (0.55-0.79) | 0.57 (0.42-0.70) |
| *Composite Reference Standard* | 0.55 (0.49-0.62) | 0.56 (0.50-0.63) | 0.52 (0.45-0.58) |

*Diagnostic Performance—Longitudinal Changes in Alternative Scoring Methods*

As with the original TB scores, I also evaluated the ability for longitudinal changes in these novel scoring approaches—based on individual genes or gene pairs—to retrospectively diagnose children with TB. Longitudinal changes were evaluated as described in **Section 3.3**, using data from the same cohort of 221 children. No longitudinal change in any of the scoring methods developed meaningfully exceeded the diagnostic performance of TB scores at baseline alone (**Tables 8** and **9)**. For comparison, when calculated specifically in the cohort of 221 children, performance of the *KLF2*-based TB score at baseline was 0.56 (95% CI 0.42-0.71) against the MRS and 0.53 (95% CI 0.40-0.66) against the CRS; performance of the *TBP*-based TB score at baseline was 0.57 (95% CI 0.40-0.73) against the MRS and 0.50 (95% CI 0.37-0.62) against the CRS.

**Table 8.** The AUCs for the ability of longitudinal changes in the normalized individual HRC cartridge genes' Ct values to differentiate participants on the basis of their reference standard status. 95% confidence intervals were determined using 2000 bootstrap resamples. Reference standards were defined as outlined in **Section 2.6**.

| | **Normalized *KLF2* Ct, longitudinal change** | **Normalized *DUSP3* Ct, longitudinal change** | **Normalized *GBP5* Ct, longitudinal change** | ***TBP* Ct, longitudinal change** |
|---|---|---|---|---|
| *Microbiological Reference Standard* | 0.52 (0.41-0.64) | 0.52 (0.37-0.67) | 0.54 (0.40-0.68) | 0.56 (0.40-0.72) |
| *Composite Reference Standard* | 0.55 (0.42-0.68) | 0.59 (0.46-0.71) | 0.52 (0.40-0.64) | 0.50 (0.36-0.64) |

**Table 9.** The AUCs for the ability of longitudinal changes in the difference in Ct values for two non-housekeeping genes to differentiate participants on the basis of their reference standard status. 95% confidence intervals were determined using 2000 bootstrap resamples. Reference standards were defined as outlined in **Section 2.6**.

| | ***GBP5* - *KLF2*, longitudinal change** | ***DUSP3* - *KLF2*, longitudinal change** | ***DUSP3* - *GBP5*, longitudinal change** |
|---|---|---|---|
| *Microbiological Reference Standard* | 0.53 (0.38-0.67) | 0.46 (0.33-0.61) | 0.58 (0.44-0.71) |
| *Composite Reference Standard* | 0.56 (0.43-0.69) | 0.53 (0.41-0.65) | 0.56 (0.43-0.69) |

*Unsupervised clustering with t-SNE*

Finally, I sought to see if unsupervised clustering approaches could identify any subgroups with distinct gene expression profiles across the four genes measured by the HRC. To accomplish this, I input the normalized Ct values of both baseline and two-week testing for all 348 participants with valid test results at both timepoints available into a t-distributed Stochastic Neighbour Embedding (t-SNE) model. After qualitatively observing its output, shown in **Figure 13** and colour-coded by subgroups of interest, no obvious participant groupings emerged.

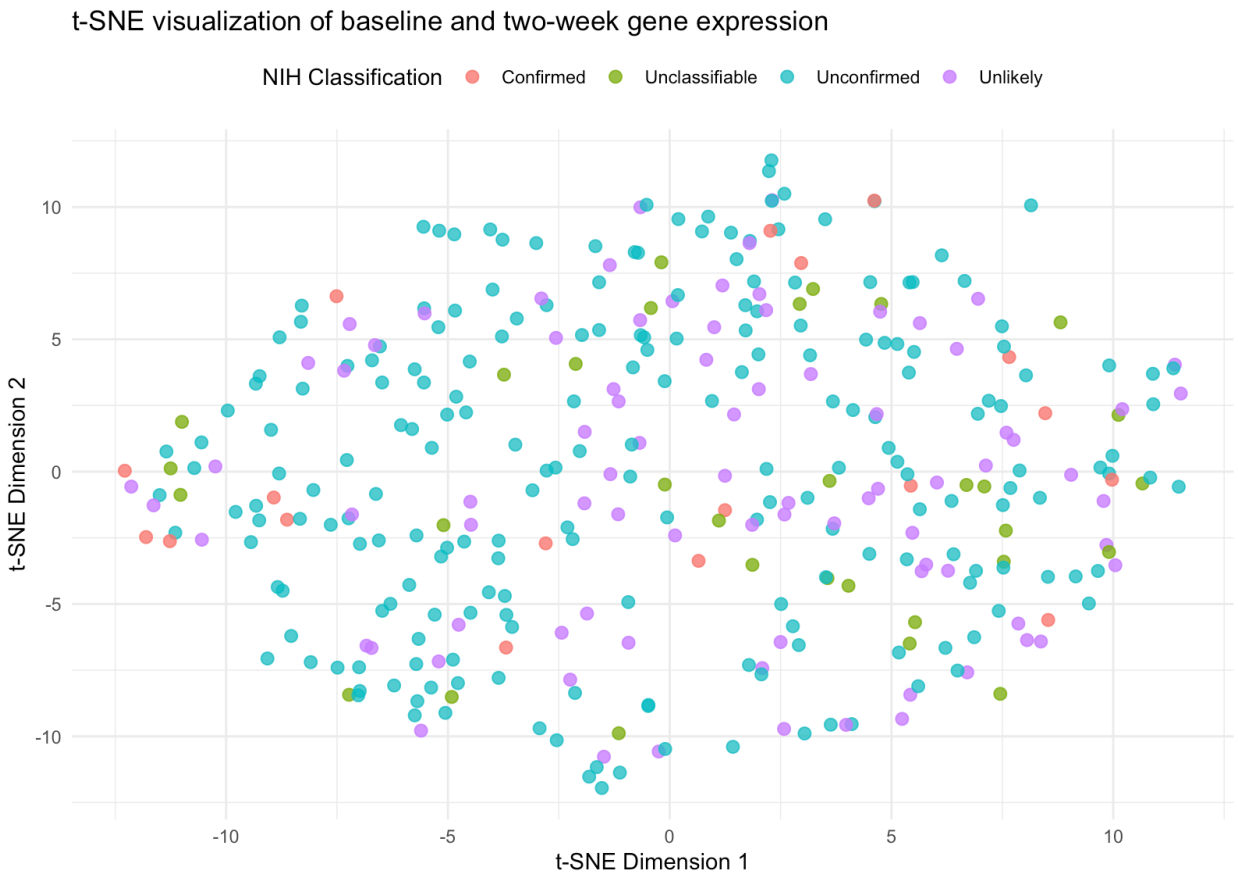



**Figure 13.** A t-SNE model, with perplexity of 30, using normalized Ct counts (excluding *TBP*, which was not normalized) from 348 participants with valid host response cartridge test results from both visits. Case classifications were determined as outlined in **Section 2.5**.

## 3.9 Discussion

*Discrepancies in HRC Performance across Paediatric Studies*

The HRC did not exhibit useful diagnostic accuracy in the NOD-pedFEND cohort, with performance not far exceeding that of a random classifier. This is in line with the results of the COMBO study [86], but contrary to those of RaPaed-TB [84] and the work of Sun et al. [87], who reported better performance in the under-five age group (**Table 10**). The underlying cause of this discrepancy is unclear and merits further investigation. It is

possible, for example, that the cohort composition in the RaPaed-TB or Sun et al. studies differs in some meaningful way which increases the proportion of children who are better detected by the HRC [84, 87]. To investigate this, exploratory analyses of the HRC's performance in clinical and demographic subgroups would be warranted. However, a key limitation of ours and others' data has been the limited ability to perform these subgroup-level analyses due to low sample sizes, particularly of individuals with confirmed TB disease. A meta-analysis of all available paediatric data would enable such an investigation and perhaps elucidate potential contributors to the disparate results observed.

**Table 10.** A comparison of the overall HRC performance in children under five, as reported by ourselves and other groups. Abbreviations: Strict Reference Standard (SRS), Microbiological Reference Standard (MRS).

| Study | Reference Standard | TB Score Cutoff Used | Sensitivity | Specificity | AUC |
|---|---|---|---|---|---|
| RaPaed-TB [84] | SRS (requiring culture confirmation) | KLF2-based score, 1.5 | 27/47 (57%) | 82/89 (92%) | Not provided |
| COMBO [86] | MRS | TBP-based score, -0.65 | 7/9 (78%) | 30/91 (33%) | Not provided |
| Sun et al. [87] | CRS, including healthy controls | KLF2-based score, 2.675 | 79% | 89% | 0.89 |
| NOD-pedFEND | MRS | KLF2-based score, 1.5 | 3/19 (16%) | 334/368 (91%) | 0.56 |
| | CRS | KLF2-based score, 1.5 | 26/260 (10%) | 82/88 (93%) | 0.57 |
| | MRS | TBP-based score, -0.65 | 13/19 (68%) | 150/368 (41%) | 0.55 |
| | CRS | TBP-based score, -0.65 | 160/260 (62%) | 39/88 (44%) | 0.56 |

Known differences between study inclusion/exclusion criteria, study setup, and baseline demographics/clinical presentation stand out as promising subgroups to interrogate in such a meta-analysis. For example, the studies which included children with

extrapulmonary TB (EPTB) in their cohorts (RaPaed-TB and the Sun et al., but not NOD-pedFEND or the COMBO study) reported more positive results for the HRC, suggesting that it might be worth investigating whether the HRC performs better in children with EPTB [84, 86, 87]. Another potential contributor might be the geographic basis of the study, as both COMBO and NOD-pedFEND recruited largely or entirely (respectively) from Uganda, while the Sun et al. cohort recruited from China and the RaPaed-TB cohort spanned South Africa, India, Malawi, Tanzania, and Mozambique [84, 86, 87]. While HIV statuses of participants in the Sun et al. cohort were not stated, the RaPaed-TB cohort also had a lower proportion of CLHIV (8% overall, with an unknown age distribution) than our cohort (13%) and COMBO (16%), which our results suggest might also have contributed to differences in HRC TB scores [84, 86, 87].

It should also be noted that the RaPaed-TB study did not utilize Xpert positivity on stool in their definition of microbiologically confirmed cases, while NOD-pedFEND did, and tests used for microbiological confirmation in the COMBO and Sun et al. cohorts were not immediately apparent [84, 86, 87]. Calculating the HRC's performance amongst only children with culture-confirmed disease, more than a single trace-positive result on Xpert Ultra, or positivity on a non-stool test would be an interesting analysis to begin to evaluate the importance of these differences.

Additionally, the RaPaed-TB cohort had substantially fewer children with unconfirmed TB, which may suggest slight variations in the studies' implementation of the definition of "unconfirmed TB", which could foreseeably influence diagnostic accuracy results [84]. However, such concerns are, in part, why different reference standards are

utilized—the MRS, being primarily influenced only by the tests performed, should be less susceptible to discrepancies in how the clinical and radiographic signs of TB are interpreted. With my results based upon the MRS also falling below those of the RaPaed-TB and Sun et al. cohorts, sufficiently powered analyses in subgroups of interest are truly needed to best understand the underlying drivers of the discrepancy in results observed [84, 87].

*Subgroup-Specific Performance*

I did, however, perform limited subgroup evaluations, which did not identify any groups for which the HRC performed dramatically differently. It should be noted, however, that the greater variation between results when switching between the MRS and CRS in subgroup evaluations may partially stem from the lower sample sizes in the MRS+ group, which—having only a maximum of 12 participants in the assessed subgroups—has a far lower sample size than the CRS+ group, which—having a maximum subgroup size of 124 participants—has a substantive sample size even after subgrouping. In fact, it is the CRS- sample size that is most limiting in our cohort for subgroup evaluations using the CRS; however, the maximum of 48 CRS- individuals in a subgroup is still significantly larger than the maximum of 12 reference standard positive individuals included in subgroup analyses using the MRS.

Given these sample size concerns, I also utilized data from the 241 baseline tests from children with unconfirmed TB to better understand the impact of participant characteristics on HRC test performance, independent of presumed TB diagnosis. By fitting linear regression models to the baseline test scores of all children with unconfirmed TB, I

identified a significant influence of HIV status on both the *KLF2*-based and *TBP*-based TB scores, with CLHIV having generally lower TB scores than CHEU or CHU. Additionally, I observed age to have a significant impact on the *TBP*-based score, with each additional month of life subtracting 0.01 from the expected TB score in our cohort, which spanned from years 0-5. While also a potential explanation for discrepancies in HRC performance between paediatric cohorts, this evidence for age's influence over HRC scoring may also help to explain why performance of the HRC appears inferior in children versus adults. Our data on the influence of age on the TB score somewhat aligns with the findings of Sun et al., who also reported a substantial drop in TB scores with increasing age; however, their work utilized the *KLF2*-based score, which did not have a significant drop in score over the 0-5 age range in our cohort. That being said, their results spanned from 0-18 years of age, with a small, statistically insignificant drop observed between 0-5. It should also be noted that this analysis does not consider any changes to TB incidence with increasing age or HIV infection, which could influence TB scores as a mediating variable. Regardless, however, it does appear clear that age and HIV status are associated with changes in TB scores, either directly or through other mediating factors.

*Validating HRC Technology*

Implementation of the HRC appeared to be quite feasible, with the test utilizing a relatively simple sample type and the GeneXpert platform already familiar and available to site lab technicians. Further, with only 40/813 (5%) of HRC tests yielding non-valid results, test implementation appeared to be successful. However, the rate of valid results does not provide any data to support the accuracy of the signature measurement itself. It is possible

that the poor performance of the HRC observed in children is a consequence of inaccurate signature scoring, rather than the signature itself not being diagnostic. This concern is partially attenuated by the knowledge that the HRC runs on the Xpert platform, which has been successfully used for PCR-based *Mtb* detection for years. Nevertheless, this is insufficient to confidently report that the chosen signature, and not the HRC cartridge itself, is the primary contributor to the assay's poor performance. Comparing the HRC's TB score to measurements of the same TB signature taken via alternative modalities, such as RNA-Seq, NanoString, or other RT-qPCR methods, would help to evaluate the accuracy of the HRC itself. If the HRC's TB score is shown to correlate well with the signature measurements of other established methods, it still may retain promise if a better signature of paediatric disease is discovered and the test is able to be repurposed to measure that signature instead.

*Longitudinal Changes in TB Scores*

I also investigated the longitudinal changes in HRC scoring observed after two weeks of TB treatment. Longitudinal changes have primarily been investigated for prospective use as a marker of treatment response [84, 87]. If clear changes in the TB score after treatment are indicative of the presence of TB disease, however, measuring such changes could also be useful for retrospective confirmation of TB disease status. While this use might not be particularly helpful in clinical practice, it would assist in research, allowing better clarity of who truly had TB amongst the heterogenous and unclear TB statuses of many children with presumptive TB, particularly those classified as "unconfirmed TB". Unfortunately, however, I did not observe a consistent change in TB

scores amongst even those children with microbiologically confirmed TB. While this limits the ability of the HRC to be used for treatment monitoring or retrospective classification—at least within two weeks of treatment initiation—it does suggest that, while suboptimal, results from delayed HRC testing performed after treatment initiation may be able to serve as a proxy for baseline TB scores, engendering some confidence in the short-term stability of the tested gene signature.

*Alternative Scoring Methodologies*

I also investigated the scoring methodology used by the HRC test and explored alternative methodologies which could be employed. Despite evaluating the performance of the component genes individually, as paired ratios, or through unsupervised clustering methods, no alternative scoring methodology was clearly superior to that of the calculated TB score, including when looking longitudinally. I did, however, see stronger-than-expected performance for the discrimination ability of *DUSP3* alone, along with the *DUSP3-KLF2* difference score; however, given the number of tests performed, concerns about this finding being a result of testing multiple hypotheses are warranted.

Along with the COMBO study, our data are some of the first to interrogate the new *TBP*-based HRC scoring methodology, as applied to children [86]. I observed moderate, but meaningful, differences in the *TBP*- and *KLF2*-based TB scores, suggesting that caution when comparing results from studies utilizing different TB score methodologies is warranted. Additionally, these results raise the question of how the *TBP*-based TB score methodology, which is not tested in most prior literature, would have performed in these prior cohorts. For the HRC's use in paediatric populations, the small number of HRC

evaluations performed thus far may make a meta-analysis with the calculation of missing *TBP*-based TB scores feasible and worthwhile, as doing so would both support cross-study comparisons and provide the power for more robust subgroup evaluations, as previously noted.

## 3.10 Conclusion

The Cepheid Host Response Cartridge performed poorly in the NOD-pedFEND cohort, with diagnostic accuracy not significantly better than that of a random classifier. The HRC also failed to demonstrate improvement in the TB score after two weeks of treatment in participants with confirmed TB. Alternative scoring methodologies also did not perform meaningfully better than the *KLF2*- or *TBP*-based TB scores, though *DUSP3* was revealed to have discriminatory ability in detecting microbiologically confirmed TB which was superior to that of a random classifier. Further, the rate of invalid tests was reasonable, but the performance of the HRC itself limits its utility. More research is needed to elucidate root causes of disparate HRC results across different studies, better ascertain HRC's diagnostic performance in subgroups of interest, evaluate the performance of the HRC technology itself, and identify better-performing signatures of paediatric TB disease.

# Chapter 4: Validation of Reported RNA Signatures of the Host Response to Tuberculosis in the NOD-pedFEND Cohort

## Introduction

### 4.1 Reported Transcriptomic Signatures of TB

While the signature developed by Sweeney et al. [61], as evaluated by the Cepheid Host Response Cartridge (HRC), failed to meet performance benchmarks [58] for its use as a triage or diagnostic test, numerous other transcriptomic signatures of TB have been reported. However, the majority of these signatures were developed in exclusively adult, or, at best, adult and adolescent cohorts, and their performance in paediatric cohorts has not yet been assessed. We sought to evaluate a selection of these signatures in the NOD-pedFEND cohort.

Transcriptomic signatures, including those for TB, are typically developed in a case-control design, comparing the gene expression between a "positive" group and one or more control groups. As such, reported signatures are a product of the datasets from which they are derived. The composition of the cohorts used in development of the signatures can, therefore, influence the ultimate signature developed and its applicability to external cohorts. Moreover, what disease states are defined as "cases" versus "controls" also can meaningfully influence reported signatures. Given that most existing signatures of TB were developed in cohorts composed primarily of adults, the signatures were not developed using the same TB reference standards used for paediatric research. As a result, while

some case definitions used in adult signature development map more cleanly onto a paediatric TB classification—such as individuals with TB disease broadly aligning with the confirmed TB definition in children, or individuals with "other diseases" with the unlikely TB classification—these equivalences are imperfect. A broad "TB disease" classification, for example, can include individuals diagnosed through a number of avenues, including smear microscopy, LAM testing, or even clinical diagnosis, not just the Ultra or culture testing that informed our confirmed TB classification.

Together, these factors motivate the need to explicitly validate and/or develop signatures in paediatric cohorts, especially since their distinct immune response to *Mtb,* further described in **Section 1.2**, may influence the performance of a signature derived from data obtained from adults. Additionally, factors such as the sample type used for analysis, algorithms used for signature development, and the technology/approach used to measure gene expression may also influence reported transcriptomes and the resultant signature of TB. Given this, what follows is a summary of the transcriptomic signatures evaluated in my analysis, shared to provide this critical context on how the signatures were originally derived.

Signatures have been developed for many stages of the TB clinical course, and those included in this evaluation are summarized in **Table 1.** These include diagnostic signatures, including those of TB disease or infection, but also prognostic signatures, including signatures for the risk of progression from infection to disease and for treatment failure. I address each category of TB signatures in turn.

**Table 1.** An overview of the TB signatures evaluated in our analysis, including their component genes, the algorithm used for signature development, the cohorts in which the signature was derived, and the technology used for measurement of gene expression. Underneath the authorship of each signature is the name by which it is referred to in this chapter's analysis. Abbreviations: TB Disease (TBD), TB Infection (TBI), Other Disease (OD), Household Contact (HHC), Healthy Control (HC), Interferon-Gamma Release Assay (IGRA), Peripheral Blood Mononuclear Cells (PBMCs), Reverse Transcription Polymerase Chain Reaction (RT-PCR), Support Vector Machine (SVM), Quantitative Real-Time Reverse Transcription Polymerase Chain Reaction (qRT-PCR).

| Signature | Upregulated Genes | Downregulated Genes | Original model | Population (for development) | Originally designed for | Assessed by |
|---|---|---|---|---|---|---|
| **Darboe et al. [89]**<br><br>(Darboe_RISK_11) | *BATF2, ETV7, FCGR1B, GBP1, GBP2, GBP5, SCARF1, SERPING1, STAT1, TAP1, TRAFD1* | N/A | Support Vector Machine (SVM)-based gene pair model; developed by removing redundant or less reproducible genes from Zak_RISK_16 signature [90]. | ACS Cohort (12-18yr, South Africa, 46 who progressed to TBD and 107 with TBI who did not) | Risk of developing TBD in TBI | qRT-PCR of splice junctions from whole blood or PBMC samples |
| **Francisco et al. [91]**<br><br>(Francisco_OD_2) | *GBP5* | *KLF2* | Random forest | >18yr; China; 419 TBD, 499 OD, 499 HCs | Active TB vs OD or HC | RT-PCR of whole blood or PBMCs |
| **Jacobsen et al. [92]**<br><br>(Jacobsen_3) | *LTF, CD64* | *Rab33a* | Linear Discriminant Analysis | Adults; Germany; 37 TBD, 7 TBI, and 15 HCs | TBD vs TBI or HC | Microarray performed on cDNA from PBMCs |
| **Kaipilyawar et al. [93]**<br><br>(Zhao_NANO_6) | *ANKRD22* | *BLK, CCR6, ATP1A1, DARS2, and EXOC2* | Lasso regression | >6yr, South India, 80 people with TBD and 102 HHCs with TBI | TB vs TBI | NanoString on whole blood, measuring 107 genes |
| **Kaul et al. [94]**<br><br>(Kaul_3) | *FCGR1B, GBP1, and GBP5* | N/A | Decision tree (OneR) and binary logistic regression | HHCs ≥ 14 years old (threshold may have been lower), India, 43 TBI and 37 HCs | TBI vs HC amongst HHCs | qRT-PCR of six genes from whole blood |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Kulkarni et al. [95]**<br><br>(Kulkarni_HIV_2) | *RAB20, INSL3* | N/A | Decision tree | Adults with advanced HIV (CD4 count < 100 cells/mm$^3$); India and Uganda; 34 with TBD and 29 without (TBI status not assessed) | Concurrent TBD and HIV vs HIV alone | RNA-Seq of whole blood; five samples from India removed due to failing QC checks |
| **Leong et al. [96]**<br><br>(Leong_RISK_29) | *APOL6, CTSA, DERA, FUT4, HM13, IL31RA, MOB3C, NAGA, SLC3A1, SPSB1, SRBD1, TCN2, WARS, ZNRF1* | *ZNF419, SH2D1B, GSTA4, AGAP9, LUC7L, CIRBP, PRSS53, MDN1, SNRNP70, SPDYE5, CCDC14, NEIL1, ENO3, CCDC78, ZNF202* | Discovery: Ensemble feature selection model<br><br>Testing: Any of ridge logistic regression, SVMs, random forests, or gradient boosting | ACS Cohort (12-18yr, South Africa, 46 who progressed to TBD and 107 with TBI who did not) | Early prediction of TBD development in individuals with TBI | RNA-Seq on whole blood |
| **Lopez et al. [97]**<br><br>(Trace_4) | *ZNRF1, IL31RA, SRBD1, WARS* | N/A | Gene Set Variation Analysis (GSVA) | Adults; Kenya, South Africa, Uganda; 90 with presumptive TBD and “very low” or “trace” Ultra results, 17 of whom were culture positive. | Identifying culture positive individuals with “very low” or “trace” Ultra results | NanoString nCounter performed on whole blood |
| **Maertzdorf et al. [98]**<br><br>(Maertzdorf_4) | *GBP1, P2RY14, IFITM3* | *ID3* | Conditional inference decision tree | Adults; India; 113 TBD, 56 TBI, and 20 HC | Active TB vs TBI or HC | RT-PCR array of 372 genes from whole blood |
| **Penn-Nicholson et al. [99]**<br><br>(PennNich_RISK_6) | *GBP2, FCGR1B, and SERPING1* | *TUBGCP6, TRMT2A, and SDR39U1* | SVMs were used to identify the pair-wise gene ratios, which were used as an ensemble to determine signature score | ACS Cohort (12-18yr, South Africa, 46 who progressed to TBD and a cohort of those with TBI who did not) | Three uses: Active TB vs OD, TBI or HC, marker of treatment response, and risk of developing TBD in TBI | RNA-Seq on whole blood for discovery, qRT-PCR for validation (whole and capillary blood) |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Roe et al. [100]**<br><br>(Roe_3) | *BATF2, GBP5, SCARF1* | N/A | Discovery: Stability selection, followed by SVMs<br><br>Testing: SVM only | Stability selection cohort: Adults, United Kingdom, TBD. Second cohort (ACS): 12-18yr, South Africa, 46 who progressed to TBD and 107 with TBI who did not | Risk of developing disease in TBI | RNA-Seq and whole-genome microarrays on whole blood |
| **Sambarey et al. [101]**<br><br>(Sambarey_HIV_10) | *FCGR1A, HK3, RAB13, RBBP8, IFI44L, TIMM10, BCL6, SMARCD3, CYP4F3 and SLPI* | *None* | Linear Discriminant Analysis | Adults; India; 19 TBD, 13 TBI, 15 HCs, 7 IGRA- individuals living with HIV | TBD vs TBI vs HIV vs HC | RNA-Seq (not for all participants) and qRT-PCR of 21 candidate genes (both on whole blood) |
| **Sloot et al. [102]**<br><br>(Sloot_HIV_2) | *IL-13* | *AIRE* | Lasso regression | Adults >25 years old using drugs and living with HIV; Amsterdam; 14 people who developed TB, 15 who did not develop TB | Detect TB up to eight months before diagnosis | Dual-colour reverse transcriptase multiplex ligation-dependent probe amplification (dcRT-MLPA) for 141 genes, measured from PBMCs |
| **Suliman et al. [103]**<br><br>(Suliman_RISK_4) | *GAS6, SEPT4* | *CD1C, BLK* | Support Vector Machine (SVM)-based gene pair model | GC6-74 cohort (10-60 years old; Uganda, Ethiopia, South Africa, and The Gambia; HHCs, 79 who progressed to TBD and 328 matched individuals who did not) | Risk of developing disease in TBI | RNA-Seq from whole blood |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Sweeney et al. [61]**<br><br>(Sweeney_OD_3) | *GBP5, DUSP3* | *KLF2* | (*GBP5* + *DUSP3*) / 2 - *KLF2* | Meta-analysis of 14 publicly available datasets; 3 used for discovery (all from adults) and 11 for validation | Active TB vs TBI vs OD vs HC | Discovery: Meta-analysis of three publicly available microarray datasets on whole blood. Validation: remaining 11 datasets, all on whole blood |
| **Zak et al. [90]**<br><br>(Zak_RISK_16) | *ANKRD22, APOL1, BATF2, ETV7, FCGR1A, FCGR1B, GBP1, GBP2, GBP4, GBP5, SCARF1, SEPT4, SERPING1, STAT1, TAP1, TRAFD1* | N/A | Support Vector Machine (SVM)-based gene pair model | ACS Cohort (12-18yr, South Africa, 46 who progressed to TBD and 107 with TBI who did not) | Risk of developing disease in TBI | RNA-Seq for discovery, qRT-PCR of splice junctions for validation (both on whole blood) |

*Signatures of TB Disease*

Signatures of TB disease typically differentiate those with TB disease from those with other diseases, TB infection alone, or from individuals considered to be "healthy controls". A three-gene signature developed by Jacobsen et al., for example, compared RNA in peripheral blood mononuclear cells (PBMCs) in 37 adults with pulmonary TB to that from 7 healthy adults with TB infection and 15 without infection. In so doing, they reported that the combination of *Rab33a, CD64,* and *LTF* differentiated between those with and without disease with a predictive accuracy of 85% [92]. Other signatures of disease include a six-gene signature proposed by Kaipilyawar and colleagues. This signature, designed to distinguish TB disease from infection in a cohort of TB cases and their household contacts ≥ 6 years old in South India, also performed well, with an AUC of between 0.70 and 1.00 across 10 independent datasets. Maertzdorf et al. have also proposed a four-gene

signature to discriminate TB disease from infection and healthy controls based on data collected from 200 adults in Bangalore, India, with this signature demonstrating an AUC of 0.82 and 0.89 in two independent datasets [98]. For identifying TB infection, Kaul et al. proposed a three-gene signature, developed amongst household contacts in India, which predicted IGRA results with an AUC of 0.68 [94]. Signatures of TB disease also include the three-gene signature reported by Sweeney et al., as previously described [61], along with a two-gene derivative of the signature proposed by Francisco et al. from a large adult cohort in China [91]. Additionally, a four-gene signature has been proposed to identify culture-positive TB from amongst symptomatic individuals with "very low" or "trace" Ultra semi-quantitative results, with its performance yielding an AUC of 0.82 in the authors' test set [97].

*Signatures of TB in Individuals Living With HIV*

Three signatures designed for use solely for people living with HIV have also been proposed. The first, a two-gene signature by Sloot et al., purports to identify TB up to eight months prior to TB diagnosis amongst adults living with HIV, with an AUC of 0.80 in their cohort [102]. The second, a two-gene signature proposed by Kulkarni et al., was designed to distinguish concurrent TB and HIV from HIV infection alone in those with CD4+ cell counts less than 100 cells/$mm^3$. Though the signature demonstrated an AUC of 0.95 for detecting culture-positive TB disease in the discovery cohort, its performance diminished in two independent validation sets, with AUCs of between 0.62 and 0.75 [95]. A third signature, while not solely for individuals with HIV, was designed to differentiate between TB disease, TB infection, healthy controls, and IGRA-negative individuals living with HIV;

this signature consists of 10 genes and was developed by Sambarey et al. [101]. When evaluated in an external cohort, the signature performed well, with AUCs between 0.70 and 0.93, depending on the study site and disease states being compared.

*Prognostic Signatures for TB*

Prognostic signatures have also been proposed, though many, including most evaluated here, originate from data collected in the same study: the Adolescent Cohort Study (ACS). The study, which enrolled 3,595 participants between 12 and 18 years old with TB infection in South Africa, collected RNA-stabilized whole-blood samples from participants either every six months for two years or at baseline and two years later, with follow-up identifying which participants progressed to TB disease [90]. From this design, the study identified 46 children who progressed to disease and matched controls 2:1 from those in the cohort who did not develop disease. Using samples from these children, several signatures of disease progression have emerged.

First, a 16-gene signature developed by Zak et al. reported an AUC for predicting TB disease development of 0.69 in a test set, which was replicated in an independent validation cohort. This signature was later refined down to 11 genes by Darboe et al., who removed genes using three criteria: genes with redundant signals, genes which were utilized in the fewest number of gene pairs in the SVM-based model, and genes that had poorer reproducibility when evaluated by qRT-PCR [89]. This condensed signature performed equivalently to the original signature in their evaluation, with a Spearman's correlation coefficient of 0.99. Later, a new six-gene signature was developed from the same cohort, which performed better (AUC=0.88) than the other signatures (for both, AUC

= 0.82) when evaluated using only samples taken 12 months or less before TB diagnosis. This signature was also designed for use as a diagnostic as well as a marker of treatment response. For diagnosis, it performed well, with AUCs ≥ 0.85 in all independent cohorts tested; for treatment response, performance was also strong in an independent cohort, with an AUC of 0.77 for predicting treatment response before treatment was given, and an AUC of 0.95 for identifying non-responders after the treatment course [99]. For its ability to identify individuals at risk of progressing from TB infection to disease, however, the signature's performance was diminished in an external cohort, with an AUC 0.71. Of note, both the 16-gene and 6-gene risk signatures were validated in the same external cohort: that of the Grand Challenges 6—74 project, which included household contacts of adults with culture-positive TB disease across four nations in Sub-Saharan Africa [103].

That these signatures were evaluated in the same cohort speaks to a larger concern about reported signatures of TB progression. Studies of TB's progression from infection to disease are highly resource-intensive, requiring large numbers of enrolees and extended follow-up to identify the relatively small percentage of individuals who progress to TB disease. As such, available cohorts for signature development are limited. As a result, many signatures of progression are developed or validated in shared datasets, which limits their diversity and the understanding of their performance in heterogenous populations.

Recently, new prognostic signatures of TB have been developed which utilized another dataset, though in conjunction with the ACS, to inform their gene selections, which begins to address this concern. The first, a three-gene signature of *BATF2*, *GBP5*, and *SCARF1* reported by Roe et al., was developed in a two-cohort approach. First, using a

cohort of adults with TB disease previously recruited from the United Kingdom [104], top-performing genes were identified by stability selection, in which 10,000 iterations of subsetting and SVM-model fitting were performed to identify genes that consistently discriminated between those with active and treated TB disease. Once this shortlist of genes was identified, SVMs were then fit to the ACS cohort, with an AUC of 0.96. When the signature was evaluated in an independent third cohort of close contacts of TB cases recruited from TB clinics in London, UK, 4/6 progressor cases had signature scores above the 99.9$^{th}$ percentile of IGRA- contacts' scores, with the remaining two cases well below this threshold [100]. A second signature was developed using 29 genes, determined using an ensemble feature selection method, also based off of the ACS cohort; however, signature validation was performed in a novel cohort: that of 14 Brazilian TB patients and 37 household contacts, 16 who progressed to TB disease and 21 who did not. In this cohort, the gene signature's AUC was an average of 0.67 across 4 different machine learning models [96]. A final signature, developed by Suliman et al., consisted of four genes derived entirely from the aforementioned Grand Challenges 6—74 cohort; upon validation in the ACS cohort, its AUC for predicting TB progression was 0.69 [103].

Together, these signatures represent a range of approaches to signature development, yielding signatures which have largely gone untested in young paediatric populations. Evaluating the performance of these signatures in infants and children is valuable, then, both for assessing their use in this patient population as well as for providing broader insights on the universality of TB signatures and the distinctiveness of the

paediatric host response to *Mtb*. With this in mind, we sought to evaluate the performance of these signatures in diagnosing TB in the NOD-pedFEND cohort.

## 4.2 Use of NanoString nCounter Technology for the Validation of RNA Signatures

For this work, signatures were evaluated using the NanoString nCounter system [105], which has been previously used to validate TB signatures in a cohort which was composed of >80% adults [93]. The technology's distinct advantage is its ability to evaluate gene expression without requiring reverse transcription of RNA to DNA, avoiding the subsequent irregularities in amplification that process can cause. This has been reported to both decrease variations in measurement [106] and enable testing on samples with lower quality RNA [107]. A limitation of NanoString, however, is its targeted approach, which requires genes of interest to be pre-determined. This, unfortunately, prevents assessment of signatures containing genes which were not pre-selected for analysis, limiting novel signature discovery and hindering one's ability to easily change the evaluated genes. The technology's specifications do, however, uniquely position it for validation work, where accuracy—and the ability to work with at-times lower quality samples—is of greater importance than the ability to evaluate the entire transcriptome, given the focus on pre-established signatures.

*Principles of the NanoString nCounter System*

The NanoString nCounter system utilizes hybridization-based barcoding to enable the direct counting of RNA sequences of interest. This works via three overarching steps, summarized in **Figure 1**. First, a proprietary set of two probes, which comprise the "CodeSet", bind to an RNA sequence of interest measuring 100 nucleotides in length. The first probe, dubbed the "reporter probe", carries a barcode on its 5' end composed of six fluorophores arranged in an order unique to each target RNA sequence. The second probe, dubbed the "capture probe", carries biotin on its 3' end. The binding properties of biotin are exploited in the nCounter system's second step to bind the target-probe complex to a streptavidin-coated surface as unbound probes are washed away. After washing away unbound probes, the barcodes on the bound target-probe complexes are immobilized and made to lay flat against an imaging surface. A high-resolution fluorescence microscope then scans the imaging surface for these barcodes, which are then counted and linked back to the RNA sequence to which they're known to be associated. As such, RNA sequences are directly measured, without the need for any amplification steps. Further, counts of RNA sequences can be normalized against counts of known housekeeping genes and in-built positive controls, enabling easy comparison across samples and batches run by the NanoString instrument.

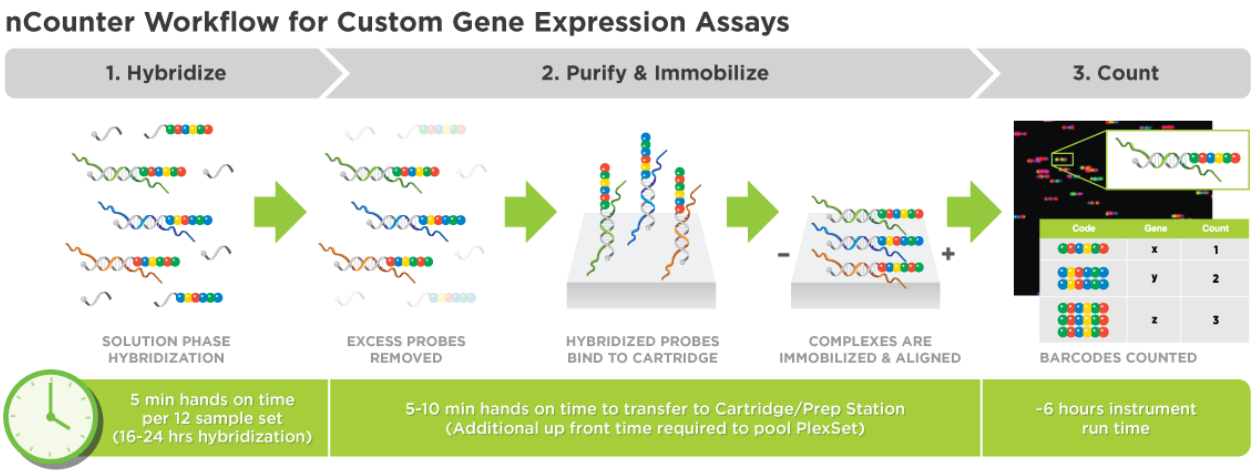



**Figure 1.** A manufacturer-provided diagram detailing the three primary steps which enable quantification of gene expression using the NanoString nCounter system. [108]

As mentioned, NanoString nCounter has been previously used for the evaluation of TB signatures in a primarily adult cohort, in which its performance aligned with results previously obtained from RNA-Seq performed on samples from the same cohort [93]. Given NanoString's use of probes specific to RNA targets of interest, to undertake this work, the authors designed a custom set of capture and reporter probes specific to genes implicated in established TB signatures like those described above. A modified version of this set of probes, along with the established and validated methodology of Kaipilyawar et al., was made available for use in this analysis, as further described in **Methods** [93].

## 4.3 Approaches to Signature Development and Evaluation

A crucial aspect of the development of signatures of disease are the algorithms and approaches used for signature creation and validation. Signature development, in particular, requires a delicate balance between fitting models to the underlying biological

truth revealed by the available data without "overfitting" models to be influenced by random noise which is unlikely to be present when the model is applied to other datasets. It is unfortunately common to have signatures perform well in the cohort used for signature development, only to disappoint when the signature is applied to an independent, but theoretically similar, cohort.

Many distinct models have been employed to structure, develop, and evaluate TB signatures. Each present a unique set of advantages and trade-offs, including on the prevention of overfitting. Different models can also have dramatically different performances in different datasets [109]. As such, multiple models are oftentimes compared to identify the best performing combination of genes and the best method to converting the expression of those genes into the desired classifications.

*Machine Learning Models for Signature Development and Scoring*

Models can be used for the identification and development of TB signatures as well as for the "scoring" of signatures themselves. Most models utilize some aspect of machine learning; linear discriminant analysis (LDA), for example, finds a linear combination of genes which best discriminates between disease categories using labelled training data. Similarly, support vector machines, often considered a gold-standard approach for gene expression data, focus on the "boundary" between labelled groups (those data points that are similar but come from different groups), identifying features that maximize the difference between these boundary points. SVMs attempt to correct for overfitting using a regularization parameter (tuneable in accordance with the desired risk of over/underfitting) and a structure that intrinsically lowers the risk of overfitting by building classifiers which

maximally separate the boundaries between groups. LDA and logistic regressions, in contrast, are less robust to overfitting concerns, as they do not contain these or other mitigation strategies. Additional machine learning approaches include random forest classifiers and lasso regressions, both of which utilize labelled training data and, through distinct approaches, also attempt to attenuate overfitting risk. All these machine learning approaches, however, are designed primarily to, once made specific to the training data and labelling task, predict the classification/label which would best be applied to new data. An advantage of these approaches is that they may be able to identify more complicated patterns in gene expression data which are best predictive of disease; for example, interactions between gene expression values or small contributions from a large number of genes.

*Statistical Methods for TB Signature Discovery*

In contrast, statistical methods can also be used to identify gene signatures of interest. However, these methods generally perform best when relationships between gene expression and disease can be represented by relatively simple, prespecified models (for example, a few genes universally elevated in those with a disease). Statistical methods often rely on these simpler model structures and pre-specified analyses, depending to a greater extent on a human interpreter to predict and identify patterns in gene expression between groups; as this becomes more difficult with larger numbers of genes, or more complicated relationships between those genes, machine learning methods become better suited to the task.

*Prevention of Overfitting*

In both cases, however, another larger structural approach is often utilized to reduce the risk of signature overfitting. By withholding a subset of the overall dataset, creating a so-called “training/test split”, a portion of a dataset—not yet seen by the model—can be used to evaluate models developed on the remaining data. This allows for the model to be evaluated on data it wasn’t specifically developed to align with, yielding a less biased assessment of model performance. The percent of data withheld for testing is partially a reflection on the volume of data available, as well as the concern for overfitting, and can regularly range from 30% to a leave-one-out approach, where a model is developed multiple times using all but a single, randomly selected data point which is used for evaluation (with performance averaged over all iterations). Additional subsets of data are at-times taken too to help refine model performance before formal testing. However, these approaches do not fully account for systemic noise/population-level variation which might influence signature performance in other populations; as such, validation in external cohorts, like NOD-pedFEND, remains the gold-standard for the evaluation of signature robustness.

*Gene Set Scoring Methods*

For external validation, gene set scoring methods are often used in lieu of explicit recreation of the original signature models. This is a reflection of the fact that recreating models for each individual signature is often impossible using only the data which is available in the public domain. Even if recreation of a model is possible, it is typically a fairly resource-intensive process. Gene set scoring methods, however, have been shown to

provide comparative evaluative performance to the recreation of signatures' original models [110]. While the models can be necessary to identify the component genes in a signature, it appears that the models are less critical for the process of converting gene expression levels into a group categorization once the genes in the signature are determined.

One such approach to gene set scoring is Pathway Level Analysis of Gene Expression (PLAGE), which utilizes singular value decomposition to calculate activity levels for a gene set/signature, which can then be used to evaluate signature performance [111]. An advantage of PLAGE over other gene set scoring methods is its ability to include both up- and down-regulated genes in a single evaluated signature. Additionally, unlike other methods, PLAGE does not use rank-based gene set scoring, which would be easily manipulated in cases where a large proportion of genes not in the evaluated signature, but in the evaluated dataset, are also associated with the disease of interest. This is particularly true when working with datasets focused on the expression of disease-related genes, like the datasets produced by NanoString.

*Experimental Aims*

With these considerations in mind, we sought to utilize an established NanoString assay on PAXgene samples to evaluate the performance, in children, of sixteen previously reported signatures of TB, as analysed using PLAGE. Using these data, I also investigated novel approaches to the use of the transcriptome as a paediatric TB diagnostic, including by deriving a novel signature of TB disease and by evaluating longitudinal changes in signature scores, given their potential to change following TB treatment. Additionally, I

interrogated the performance of the above analyses in pre-defined subgroups of interest. Finally, as an exploratory analysis, I also utilized the transcriptomic data produced by NanoString to technically validate the output scores produced by the HRC.

## 4.4 Methods

*Sample Selection*

Samples were selected for this analysis from the pool of all 1,165 PAXgene samples collected during NOD-pedFEND: 619 at baseline and 546 at the two-week visit. Participants were first selected for testing of baseline PAXgene samples, of which a portion were later selected for additional testing of their two-week samples.

Samples were selected in two rounds from among those available in the testing laboratory at the time of selection (slightly biasing towards earlier-collected samples that had had time to ship). The primary reference standard used for this analysis was that of the strict reference standard, which compares those with confirmed TB to those with unlikely TB. As such, all 40 available baseline samples from participants with confirmed TB (MRS and SRS positive) were used for this analysis, with a similar number of children with unlikely TB (n=47) selected for comparison testing. Given the limited number of participants with confirmed TB, additional baseline samples from MRS- CLHIV with a positive LAM test (n=6) were also selected and tested, though samples from these children were not used for comparisons based on the strict reference standard. To enable comparisons against a more clearly defined asymptomatic baseline, a small number of

healthy control cases (n=33) were also tested, each having been age-, sex- and HIV-status matched to a participant with confirmed TB. When possible, symptomatic participants (those who were not healthy controls) were selected to enrich for malnutrition, CLHIV, and participants who had been selected for other NOD-pedFEND projects; otherwise, selection occurred at random.

Of note, one MRS- LAM+ CLHIV had a case classification of “unlikely” TB, while all other MRS- LAM+ CLHIV children were classified as “unconfirmed TB”. While explicit application of the strict reference standard would include the one child with unlikely TB in the SRS- group used for comparison, to align with the treatment of other samples from MRS- LAM+ CLHIV—and given the uncertainty a LAM+ test brings to the “unlikely TB” classification—I opted to exclude the child from the SRS as implemented for the NanoString project, creating a “modified” SRS (demonstrated in **Figure 2**). I did, however, also perform comparisons using the extended microbiological reference standard, which only differed from the SRS by including MRS- LAM+ CLHIV in the eMRS+ group and rendering two participants with indeterminate LAM results eMRS unclassifiable.

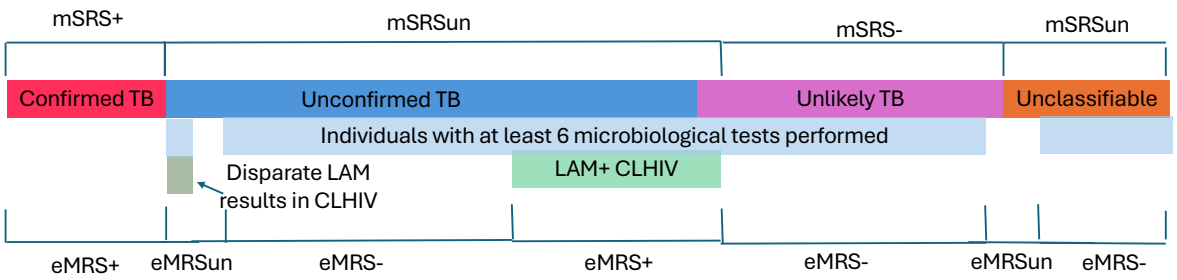



**Figure 2.** A diagram detailing the reference standards used for this analysis. The modified SRS is structured, as described in **Section 2.6**, in accordance with the SRS, apart from excluding one MRS- lipoarabinomannan (LAM)+ CLHIV with an "unlikely TB" classification. The eMRS is also structured precisely as detailed in **Section 2.6,** with children with indeterminate LAM test results (shown with an arrow) being unable to be classified by that reference standard. Abbreviations: Microbiological Reference Standard (MRS), Extended Microbiological Reference Standard (eMRS), Unclassifiable (un), Lipoarabinomannan (LAM), Children Living with HIV (CLHIV).

Two-week samples were selected from all symptomatic participants selected in the first round of baseline sample selection. This cohort constituted 28 participants with confirmed TB, 4 participants who were LAM+ CLHIV, and 32 participants classified as unlikely TB, totalling 64 two-week samples selected for analysis, each from a participant with a baseline sample which had also been selected for analysis.

*RNA Extraction*

Whole blood samples were collected in PAXgene tubes (PreAnalytiX/QIAGEN, Zürich, Switzerland) at the baseline and two-week visits from all NOD-pedFEND participants, incubated at room temperature for 2 to 72 hours, aliquoted, frozen for 24 hours at -20° C, and then finally put into long-term storage at -80° C. To extract RNA, PAXgene tubes were thawed at room temperature for 20 hours, after which RNA was extracted using PAXgene Blood RNA Kits (PreAnalytiX/QIAGEN, Zürich, Switzerland) using

the established manufacturer protocol for manual processing. Once RNA was extracted, its concentration and quality were measured on a NanoDrop spectrophotometer (Thermo Scientific, Massachusetts, USA), and subsequently frozen at -80° C or used immediately for gene expression measurement by NanoString (Bruker Spatial Biology, Washington, USA).

*NanoString nCounter Measurements and Data Processing*

For gene expression analysis using NanoString, we based our approach off of an established methodology previously used for the validation of TB signatures in a primarily adult cohort [93]. In particular, I utilized an updated version of the NS-TB107 CodeSet used by Kaipilyawar et al., modified to evaluate the expression of 101 genes of interest, for our analysis. The 101 genes included in this CodeSet, detailed in **Appendix 2**, dictated the signatures which could be measured. Measurements were performed on a NanoString nCounter Sprint instrument; upon completion, a background threshold was set to two standard deviations above the mean of measured negative controls, and geometric means were used for normalization against included positive controls using nSolver 4.0, the manufacturer-developed software utilized for initial data processing. Normalization was also performed using NanoString Advanced Analysis v2.0.134 against the geometric mean of a subset of housekeeping genes with low coefficients of variation between themselves, when measured across samples, as advised by the manufacturer.

Compiled results were imported into and analysed in R v4.5.2 using tidyverse v2.0 for data wrangling and ggplot2 v4.0.3 for data visualization. pROC v1.19.0.1 was used for construction of the ROC curves, MixOmics v6.34 for the Sparce Partial Least Squares

Discriminant Analysis, UpSetR v1.4 for the UpSet plot, and DESeq2 v1.50.2 for calculation of the principal component analysis.

Statistics were performed as described in **Section 2.8**. When indicated, corrections for multiple hypothesis testing followed the Benjamini-Hochberg (BH) procedure.

*Signature Scoring and Analysis*

The performance of gene signatures was determined using the TBSignatureProfiler package v.1.22 [112], and dependencies, in R 4.5.2. To calculate gene signature scores for each tested sample, the PLAGE algorithm—as implemented in the TBSignatureProfiler package—was utilized. As PLAGE scores for one sample are influenced by all other samples included in the scored dataset, scores were independently determined for each analysis, using only the samples to be compared in that analysis. There were two exceptions to this approach. The first was my analysis of longitudinal changes in gene signatures, which utilized signature scores calculated using PLAGE from all available samples. This approach was taken to best account for the unbalanced selection of samples to compare in this analysis (which was heavily skewed towards children with confirmed TB), as well as the need to derive scores for baseline and two-week samples that were directly comparable. The comparison of HRC scores with those derived from NanoString also utilized PLAGE scores calculated from all samples. This approach was similarly chosen to account for the mix of baseline and two-week samples tested, as well as to reflect how HRC scoring was not similarly dynamic to the specific dataset being scored.

Additionally, it should be noted that the PLAGE algorithm is sign indeterminate, with scores indicative of relative but not absolute polarity. This means that positive signature values do not necessarily indicate upregulation of certain genes, just that their genes are regulated in, overall, the opposite direction of samples with negative PLAGE values for a given signature. For our dataset, the PLAGE algorithm consistently assigned signature scores (for all but Sloot_HIV_2) for TB participants to the negative pole. To align with the assumption that diagnostic scores for TB would be elevated in individuals with TB, all scores were transformed using a global sign flip (multiplying by -1) prior to visualization.

*Derivation of a Novel Signature*

Signature development was performed in R 4.5.2 using the 3.46.0.10 version of the H2O package [113]. Baseline gene expression data, normalized as described above and converted into a $\log_2$ counts per million format, were batch corrected using the ComBat function of the sva package v3.58.0 [114] and subsequently run through the H2O package's AutoML function, which used the data to train an assortment of models to predict a participant's mSRS status. The top three signatures were selected for further analysis.

# Results

## 4.5 Overall Performance of RNA Extraction and the NanoString Cartridge

*Sample Characteristics*

For this work, RNA was extracted from 190 samples from 126 NOD-pedFEND participants, chosen as described above. From these extractions, median RNA concentration was measured as 101.0 ng/ul (IQR 55.9-170.8) by a NanoDrop spectrophotometer. RNA quality, also measured by Nanodrop, was within acceptable limits, with a median A260/A280 ratio of 2.08 (IQR 2.06-2.10) and a median A260/A230 ratio of 1.91 (IQR 1.72-2.01).

187 of these samples were chosen for analysis by NanoString (three had too little RNA content obtained from extraction to be run on the instrument). These samples consisted of 123 collected at baseline, with the remainder (64), having been collected at the two-week visit. Of the 187 total samples, 14 were rerun on the NanoString instrument due to missed QC thresholds (n=12) or the premature release of reagents from the NanoString cartridge upon sample loading (n=2). QC thresholds included a binding density of less than 0.07 features per $\mu m^2$, a limit of detection greater than or equal to a count of 36, and if, upon manual inspection of the data, there was a clear concern for data integrity (for example, if genes typically found in high abundance in all other samples were entirely undetected in a sample). In all cases, the repeat result was used for downstream analysis. As such, 187 samples, from 123 NOD-pedFEND participants, were included in this analysis.

NanoString QC metrics were acceptable, with median binding densities of 0.14 features/$\mu m^2$ (IQR 0.11-0.16). Normalization against housekeeping genes was performed excluding *OAZ*, as determined by NanoString Advanced Analysis, due to its relatively high coefficient of variation with other housekeeping genes. After normalization, four samples were flagged for an mRNA positive control normalization factor of above three (ranging from 3.02-4.31). No other QC thresholds were surpassed; further, it was determined that the mRNA positive control normalization factor threshold, when exceeded, was exceeded by a sufficiently small margin to permit the relevant samples' continued inclusion in the analysed dataset, in line with manufacturer guidance [115]. QC values and participant demographics, per case classification, for the tested samples are given in **Tables 2** and **3.**

**Table 2.** A summary of participant demographics and QC values for the baseline samples chosen for gene expression analysis, presented per participant case classification. Of note, one included LAM+ CLHIV was classified as unlikely TB; the remainder were classified as unconfirmed TB. Where medians are given, IQRs are also provided in parentheses. For continuous variables, p values were calculated using the Kruskal-Wallis Rank Sum Test. For categorical variables, chi-squared tests were used. Case classifications were determined as detailed in **Section 2.5**. Abbreviations: Lipoarabinomannan (LAM), Severe Acute Malnutrition (SAM), Moderate Acute Malnutrition (MAM), Children Living with HIV (CLHIV), Absorbance Ratio at 260 nm and 280 nm (A260/A280), Absorbance Ratio at 260 nm and 230 nm (A260/A230).

| | **Confirmed TB (n=40)** | **LAM+ CLHIV (n=6)** | **LAM- Unlikely TB (n=46)** | **Healthy Controls (n=31)** | **P Value** |
|---|---|---|---|---|---|
| *Median age (in months)* | 19.5 (13.5-36.0) | 15.0 (13.0-20.0) | 15.0 (12.2-27.0) | 24.0 (12.5-42.5) | 0.32 |
| ***Sex Distribution*** | | | | | |
| *% Female* | 23/40 (58%) | 4/6 (67%) | 22/46 (48%) | 17/31 (55%) | 0.74 |
| *% Male* | 17/40 (43%) | 2/6 (33%) | 24/46 (52%) | 14/31 (45%) | |
| ***HIV Status*** | | | | | |
| *% HIV Positive* | 6/40 (15%) | 6/6 (100%) | 3/46 (7%) | 4/31 (12.9%) | <0.001 |
| *% HIV Exposed-Uninfected* | 6/40 (15%) | 0/6 (0%) | 7/46 (15%) | 6/31 (19.4%) | |
| *% HIV Negative* | 28/40 (70%) | 0/6 (0%) | 36/46 (78%) | 21/31 (68%) | |
| *% HIV Status Unknown* | 0/40 (0%) | 0/6 (0%) | 0/46 (0%) | 0/31 (0%) | N/A |
| ***Malnutrition Status*** | | | | | |
| *% with SAM* | 13/40 (33%) | 5/6 (83%) | 22/46 (48%) | 0/31 (0%) | <0.001 (with and without unknown group) |
| *% with MAM* | 7/40 (18%) | 1/6 (17%) | 8/46 (17%) | 0/31 (0%) | |
| *% Without Malnutrition* | 18/40 (45%) | 0/6 (0%) | 14/46 (30%) | 29/31 (94%) | |
| *% With Unknown Malnutrition Status* | 2/40 (5%) | 0/6 (0%) | 2/46 (4%) | 2/31 (6%) | |
| ***Disease Severity*** | | | | | |
| *% with Severe Disease* | 20/40 (50%) | 6/6 (100%) | 9/46 (20%) | N/A | <0.001 |

| | | | | | |
|---|---|---|---|---|---|
| *% with Non-Severe Disease* | 15/40 (38%) | 0/6 (0%) | 1/46 (2%) | N/A | (with unclassifiable group), 0.025 (without unclassifiable group) |
| *% with Unclassifiable Disease Severity* | 5/40 (13%) | 0/6 (0%) | 36/46 (78%) | N/A | |
| ***RNA Quality Data*** | | | | | |
| *Median concentration (ng/ul)* | 112.7 (67.8-195.4) | 102.6 (53.3 - 158.0) | 144.30 (81.1-224.8) | 70.8 (51.0-114.0) | 0.004 |
| *Median A260/A280* | 2.08 (2.06-2.09) | 2.07 (2.05-2.09) | 2.09 (2.07-2.10) | 2.08 (2.06-2.09) | 0.11 |
| *Median A260/A230* | 1.84 (1.35-1.94) | 1.86 (1.69-1.88) | 1.86 (1.54-2.00) | 1.97 (1.88-2.03) | 0.005 |
| ***NanoString Quality Data*** | | | | | |
| *Median % fields of view counted* | 100 (100-100) | 99.5 (99-100) | 100 (99-100) | 100 (100-100) | 0.01 |
| *Median binding density (features/$\mu m^{2)}$* | 0.15 (0.12-0.16) | 0.17 (0.15-0.19) | 0.14 (0.11-0.18) | 0.14 (0.11-0.15) | 0.06 |
| *Median positive control linearity ($R^2$)* | 0.99 (0.99-1.00) | 1.00 (0.99-1.00) | 0.99 (0.99-1.00) | 1.00 (1.00-1.00) | <0.001 |
| *Median limit of detection (count)* | 17.0 (11.3-23.2) | 17.5 (16.5-19.0) | 15.7 (12.6-20.5) | 21.1 (15.9-24.4) | 0.14 |
| *% Probes above Threshold* | 87.6 (82.7-90.1) | 88.6 (84.4-92.1) | 87.1 (81.2-90.1) | 80.2 (76.7-82.2) | <0.001 |
| *Median mRNA Positive Normalization Factor* | 1.04 (0.83-1.39) | 0.87 (0.79-0.88) | 1.12 (0.85-1.42) | 1.01 (0.77-1.22) | 0.24 |
| *Median mRNA Content Normalization Factor* | 1.00 (0.82-1.22) | 0.83 (0.57-1.00) | 0.90 (0.71-1.19) | 1.15 (0.92-1.39) | 0.03 |

**Table 3.** A summary of participant demographics and QC values for the two-week samples chosen for gene expression analysis, presented per participant case classification. Where medians are given, IQRs are also provided in parentheses. For continuous variables, p values were calculated using the Kruskal-Wallis Rank Sum Test. For categorical variables, chi-squared tests were used. Case classifications were determined as detailed in **Section 2.5**. Abbreviations: Lipoarabinomannan (LAM), Severe Acute Malnutrition (SAM), Moderate Acute Malnutrition (MAM), Children Living with HIV (CLHIV).

| | **Confirmed TB (n=28)** | **LAM+ CLHIV (n=4)** | **LAM- Unlikely TB (n=32)** | **P Value** |
|---|---|---|---|---|
| *Median age (in months)* | 19.5 (14.0-36.5) | 15.0 (11.2-18.0) | 14.5 (11.8-36.0) | 0.33 |
| ***Sex Distribution*** | | | | |
| *% Female* | 15/28 (54%) | 3/4 (75%) | 15/32 (47%) | 0.67 |
| *% Male* | 13/28 (46%) | 1/4 (25%) | 17/32 (53%) | |
| ***HIV Status*** | | | | |
| *% HIV Positive* | 5/28 (18%) | 4/4 (100%) | 0/32 (0%) | ≤0.001 |
| *% HIV Exposed-Uninfected* | 3/28 (11%) | 0/4 (0%) | 6/32 (19%) | |
| *% HIV Negative* | 20/28 (71%) | 0/4 (0%) | 26/32 (81%) | |
| *% HIV Status Unknown* | 0/28 (0%) | 0/4 (0%) | 0/32 (0%) | N/A |
| ***Malnutrition Status*** | | | | |
| *% with SAM* | 9/28 (32%) | 3/4 (75%) | 14/32 (44%) | 0.48 (including unknown malnutrition status), 0.69 (excluding unknown malnutrition status) |
| *% with MAM* | 6/28 (21%) | 1/4 (25%) | 7/32 (22%) | |
| *% Without Malnutrition* | 12/28 (43%) | 0/4 (0%) | 10/32 (31%) | |
| *% With Unknown Malnutrition Status* | 1/28 (4%) | 0/4 (0%) | 1/32 (3%) | |
| ***Disease Severity*** | | | | |
| *% with Severe Disease* | 15/28 (54%) | 4/4 (100%) | 3/32 (9%) | ≤0.001 (including unclassifiable severity), 0.16 (excluding unclassifiable severity) |
| *% with Non-Severe Disease* | 10/28 (36%) | 0/4 (0%) | 0/32 (0%) | |
| *% with Unclassifiable Disease Severity* | 3/28 (11%) | 0/4 (0%) | 29/32 (91%) | |
| ***RNA Quality Data*** | | | | |

| | | | | |
|---|---|---|---|---|
| *Median concentration (ng/ul)* | 114.8 (65.5 - 188.1) | 164.5 (80.9-229.7) | 113.9 (77.0-202.5) | 0.87 |
| *Median A260/A280* | 2.08 (2.07-2.09) | 2.06 (2.05-2.09) | 2.08 (2.05-2.09) | 0.43 |
| *Median A260/A230* | 1.99 (1.84-2.08) | 1.86 (1.78-2.02) | 2.00 (1.80-2.05) | 0.82 |
| ***NanoString Quality Data*** | | | | |
| *Median % fields of view counted* | 100% (100%-100%) | 100% (100%-100%) | 100% (100%-100%) | 0.58 |
| *Median binding density (features/$\mu m^{2)}$* | 0.13 (0.12-0.16) | 0.15 (0.13-0.17) | 0.13 (0.11-0.15) | 0.68 |
| *Median positive control linearity ($R^2$)* | 1.00 (1.00-1.00) | 1.00 (1.00-1.00) | 1.00 (1.00-1.00) | N/A |
| *Median limit of detection (count)* | 19.1 (14.8-24.4) | 21.4 (17.0-22.8) | 18.1 (13.9-23.3) | 0.77 |
| *% Probes above Threshold* | 80.7 (75.7-84.4) | 84.7 (81.7-85.9) | 81.7 (76.2-83.2) | 0.47 |
| *Median mRNA Positive Normalization Factor* | 1.08 (0.88-1.25) | 0.85 (0.72-1.09) | 0.98 (0.74-1.16) | 0.51 |
| *Median mRNA Content Normalization Factor* | 1.23 (1.02-1.62) | 1.10 (0.97-1.25) | 1.21 (1.10-1.45) | 0.57 |

While significant differences in participant characteristics were noted between classification groups (including HIV status, disease severity, malnutrition status, etc.), these differences are primarily representative of intrinsic or definitional differences between groups that could not be entirely eliminated in sample selection. For example, high semi-quantitative values on Xpert Ultra would classify a child as having both confirmed TB and severe disease, making the two categorizations correlated. As a result,

significant differences in clinical characteristics between NIH case classifications should be interpreted cautiously.

Additionally, some significant differences in QC metrics were noted between classifications for baseline samples. However, as the QC values for all samples had already passed the predetermined QC criteria, which designate the range in which sample measurements are reliable, no additional adjustments to the analysis were needed to account for these differences. Moreover, as the NanoString technology does not rely on enzyme-catalysed reactions, it is resilient to changes in the A260/A230 ratio, and RNA concentration differences were accounted for in the amount of sample loaded onto the NanoString cartridge. Similarly, then, no adjustments to the analysis were made to account for differences in the RNA QC values, as, while significant, the differences were not meaningful, being unlikely to significantly influence the normalized gene expression levels measured due to intrinsic characteristics of the NanoString technology.

## 4.6 Clustering Analysis of Baseline Gene Expression

Baseline samples from 123 NOD-pedFEND participants were measured by NanoString, which yielded expression data on 101 genes. Prior to evaluating signature performance, I first assessed the ability of the full gene panel to discriminate between participants' classifications. In the unsupervised hierarchical clustering performed in construction of the heatmap shown in **Figure 3**, no clear differentiation between classifications was observed. Further interrogating this, a principal component analysis (PCA) of gene expression data (shown in **Figure 4**), while able to differentiate a portion of

participants with confirmed TB, also did not demonstrate clear separation between groups. It was also noted that 12 genes were at or below the background threshold value for >80% of all tested samples. As this was not the case for all participants for any one gene, however, no genes were excluded from the analysis. Nevertheless, given the genes' expression frequently falling below the limit of detection, the data they provided for classification was limited.

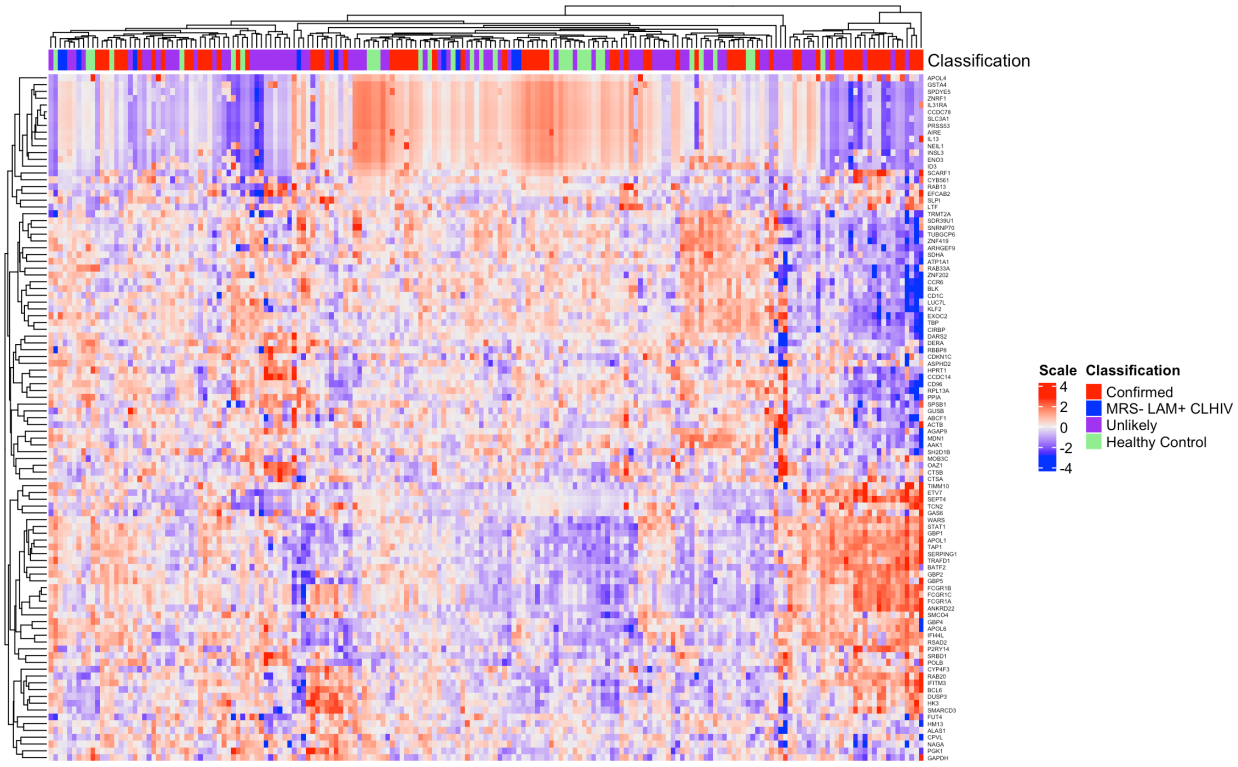



**Figure 3.** A heatmap of all baseline gene expression data measured by NanoString, scaled and annotated by the participant's classification. Classifications were determined as outlined in **Section 2.5** and **Section 2.6**. Abbreviations: Lipoarabinomannan (LAM), Microbiological Reference Standard (MRS), Children Living with HIV (CLHIV).

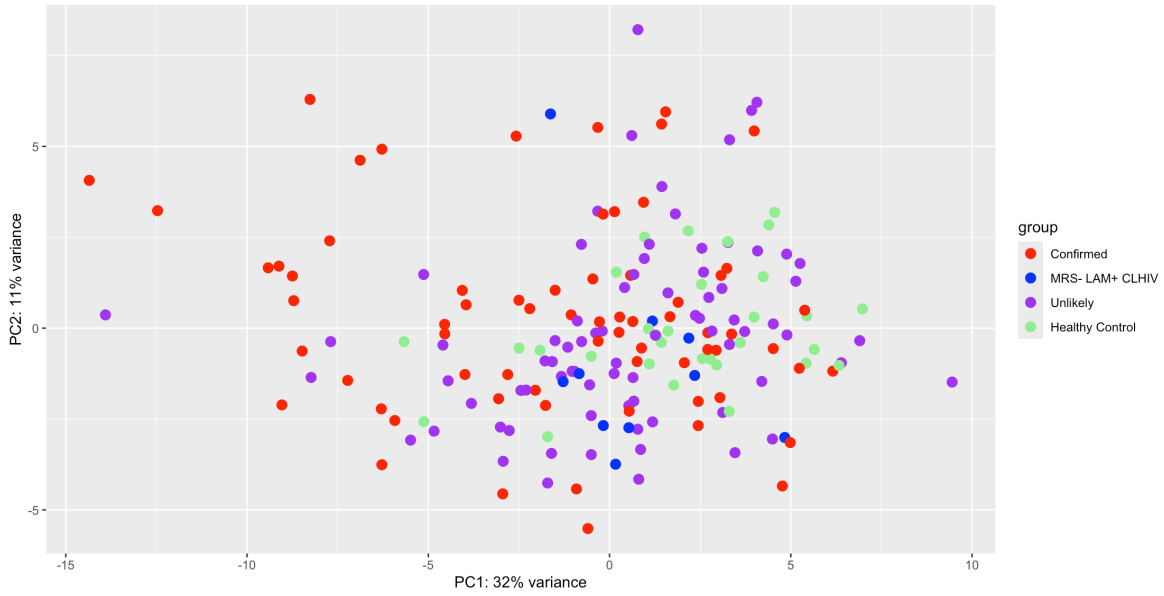



**Figure 4.** A principal component analysis (PCA) plot of baseline gene expression, as measured by NanoString, coloured by participant classification. Classifications were determined as outlined in **Section 2.5** and **Section 2.6**. Abbreviations: Lipoarabinomannan (LAM), Microbiological Reference Standard (MRS), Children Living with HIV (CLHIV).

To take advantage of group labels to help inform clustering, I also performed a Sparce Partial Least Squares Discriminant Analysis (sPLS-DA) to perform supervised clustering on the full baseline gene expression dataset (**Figure 5)**. The number of genes included in each component of the sPLS-DA were tuned iteratively, using three-fold cross-validation to evaluate the balanced error rate (BER) across the range of gene counts which could be chosen. From these results, the top 50 genes were chosen for inclusion in component one and the top 70 genes for inclusion in component two. This optimally balanced the BER with gene count, as including more genes in the analysis increased the risk of overfitting. Nevertheless, supervised clustering also did not clearly differentiate between groups, although it also identified a subgroup of participants with confirmed TB with a distinct pattern of gene expression. Of note, within the 17 individuals in the lower left

quadrant of the graph, in which all but 2/17 (12%) were children with confirmed TB, 13/17 (76%) had some form of malnutrition.

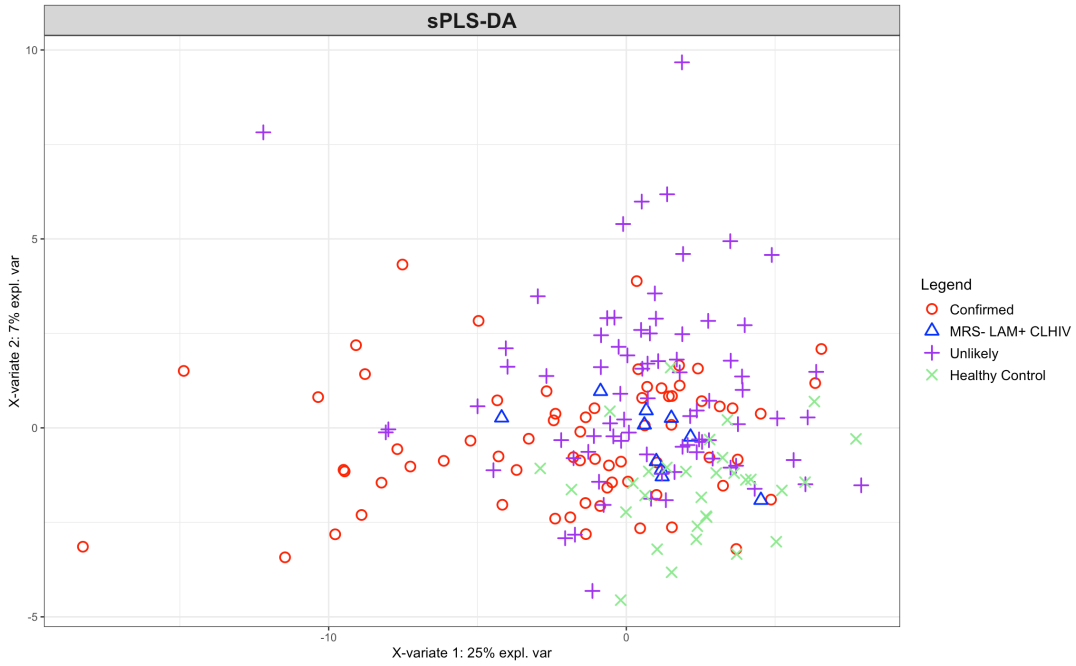



**Figure 5.** Plotted results of a sparce Partial Least Squares Discriminant Analysis (sPLS-DA), which attempted to group participants on the basis of their baseline gene expression data in accordance with their colour-coded classification. Classifications were determined as outlined in **Section 2.5** and **Section 2.6**. Abbreviations: Lipoarabinomannan (LAM), Microbiological Reference Standard (MRS), Children Living with HIV (CLHIV).

## 4.7 Performance of Previously Reported RNA Signatures at Baseline

*Distribution of Signature Scores*

Using the same dataset, covering 123 baseline samples, I next compared the distribution in an individual signature's scores between all groups. Some signatures exhibited significant differences between groups, as demonstrated in **Figure 6** and evaluated using a Wilcoxon rank-sum test corrected for multiple tests. PennNich_RISK_6

[99] and Roe_3 significantly differentiated between participants with confirmed and unlikely TB, with p=0.02 and p=0.01, respectively. Moreover, the three-gene signature proposed by Jacobsen et al. [92] had the most significant difference between any two classification groups, with healthy controls and children with confirmed TB differing with a significance of <0.001. Unfortunately, however, no signature displayed clean differentiation between any two of the tested groups.

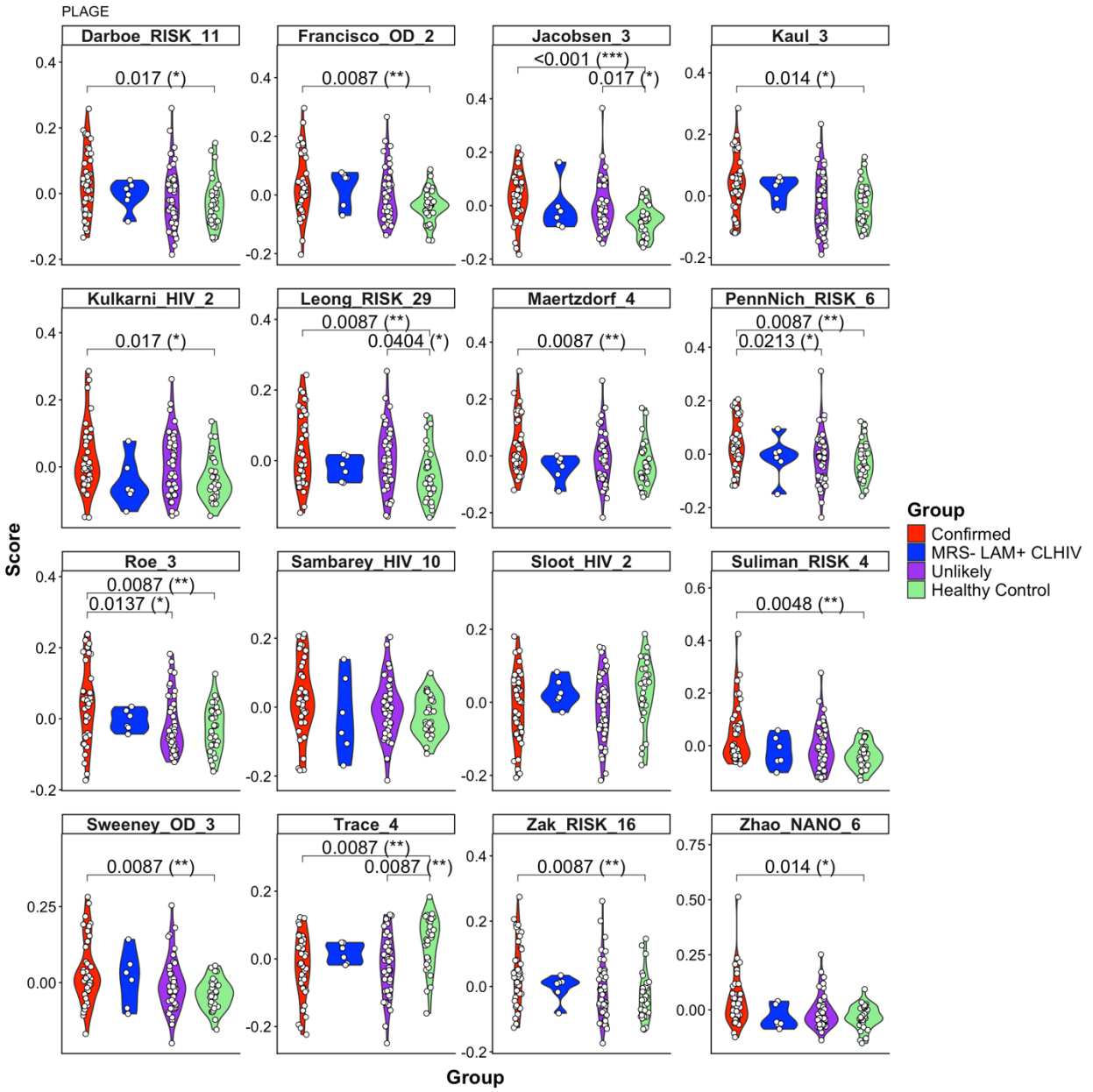



**Figure 6.** A violin plot showing the distribution of signature scores, determined by the PLAGE algorithm, of all included baseline samples, grouped by classification. Significance was determined using a Wilcoxon rank-sum test with the Benjamini-Hochberg (BH) procedure to correct for multiple statistical tests being performed. Classifications were determined as outlined in **Section 2.5** and **Section 2.6**. Abbreviations: Lipoarabinomannan (LAM), Microbiological Reference Standard (MRS), Children Living with HIV (CLHIV).

*Diagnostic Performance of TB Signatures*

I also evaluated the performance of individual TB signatures against the established reference standards. Given the lack of clear segregation between classifications used in the clustering analysis, I chose to first analyse signatures against the mSRS, focusing on comparing the groups—confirmed TB and LAM- unlikely TB—which were thought to have their TB status known with the highest level of certainty. Additionally, I contextualized our results through comparisons with the extended microbiological reference standard, as well as the gene expression of the evaluated healthy controls.

I plotted ROC curves and calculated AUCs for each individual TB signature using the TB Signature Profiler package to demonstrate the diagnostic performance of each. ROC curves and AUC values were calculated against both the mSRS (**Figure 7** and **Table 4**) and eMRS (**Figure 8** and **Table 5)**. For the eMRS calculation, as previously mentioned, two participants were excluded due to having inconclusive LAM results. From this analysis, the three-gene signature originally reported by Roe et al. [100] performed best, with an AUC of 0.69 when evaluated against the mSRS and an AUC of 0.70 when evaluated against the eMRS.

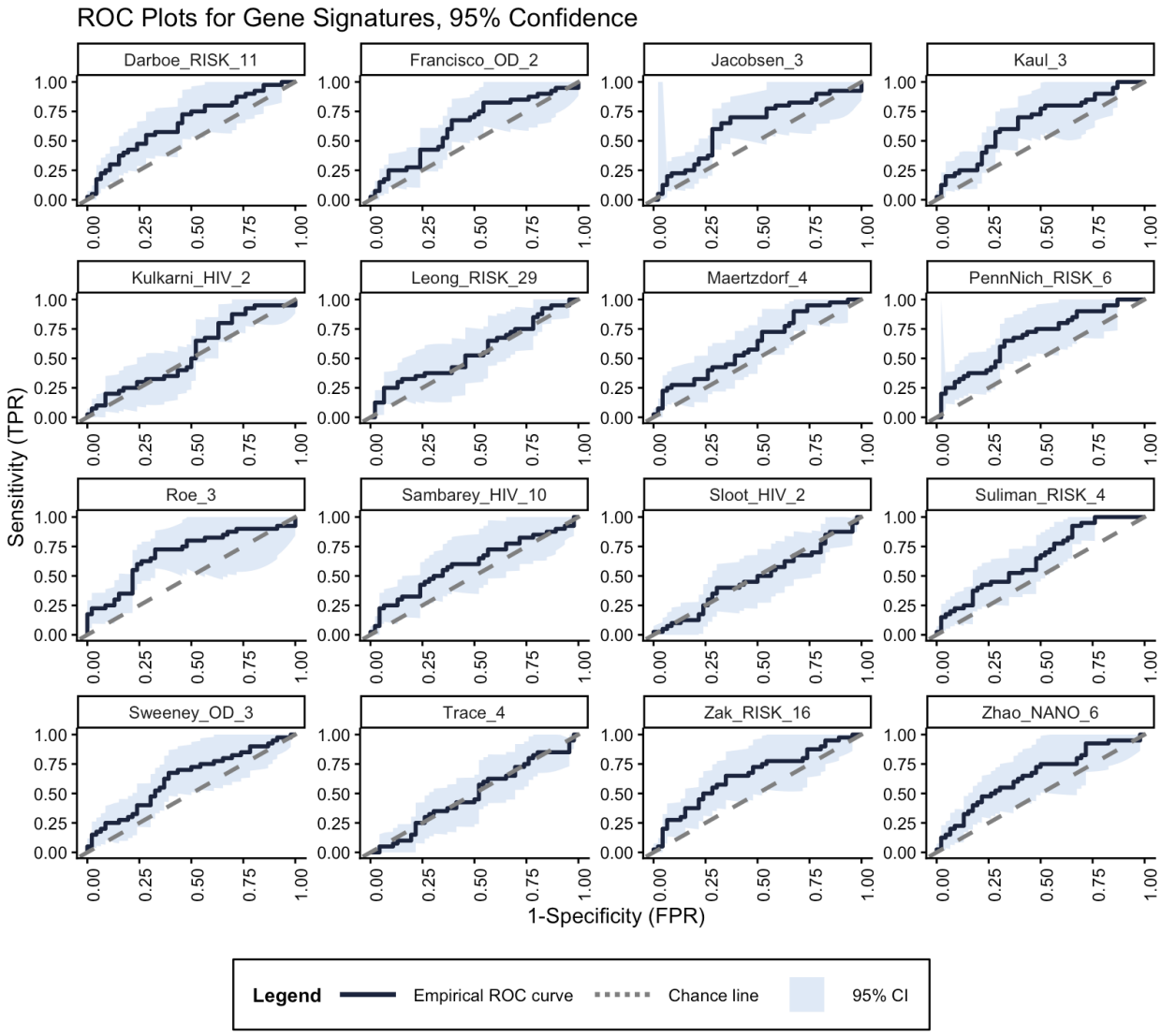



**Figure 7.** ROC plots demonstrating the ability of individual signatures to distinguish between modified strict reference standard (mSRS)+ and mSRS- participants. Reference standards were defined as detailed in **Section 2.6**.

**Table 4.** AUC values for the ROC curves shown in **Figure 7**, evaluated against the modified strict reference standard. 95% confidence intervals were determined by bootstrapping with 100 resamples. P values were calculated using a two-sided t-test. Reference standards were defined as detailed in **Section 2.6**.

| Signature | P Value | Lower 95% CI, AUC | AUC | Upper 95% CI, AUC |
|---|---|---|---|---|
| Roe_3 | 0.003 | 0.60 | 0.69 | 0.82 |
| PennNich_RISK_6 | 0.004 | 0.55 | 0.68 | 0.78 |
| Darboe_RISK_11 | 0.01 | 0.53 | 0.66 | 0.76 |
| Kaul_3 | 0.01 | 0.55 | 0.66 | 0.75 |
| Zak_RISK_16 | 0.01 | 0.56 | 0.65 | 0.78 |
| Zhao_NANO_6 | 0.02 | 0.56 | 0.65 | 0.77 |
| Suliman_RISK_4 | 0.01 | 0.52 | 0.65 | 0.74 |
| Jacobsen_3 | 0.13 | 0.52 | 0.63 | 0.74 |
| Sweeney_OD_3 | 0.05 | 0.51 | 0.63 | 0.74 |
| Francisco_OD_2 | 0.08 | 0.54 | 0.63 | 0.73 |
| Maertzdorf_4 | 0.04 | 0.52 | 0.62 | 0.72 |
| Sambarey_HIV_10 | 0.12 | 0.51 | 0.60 | 0.72 |
| Leong_RISK_29 | 0.21 | 0.50 | 0.56 | 0.66 |
| Kulkarni_HIV_2 | 0.39 | 0.50 | 0.55 | 0.68 |
| Sloot_HIV_2 | 0.82 | 0.50 | 0.52 | 0.66 |
| Trace_4 | 0.72 | 0.50 | 0.51 | 0.63 |

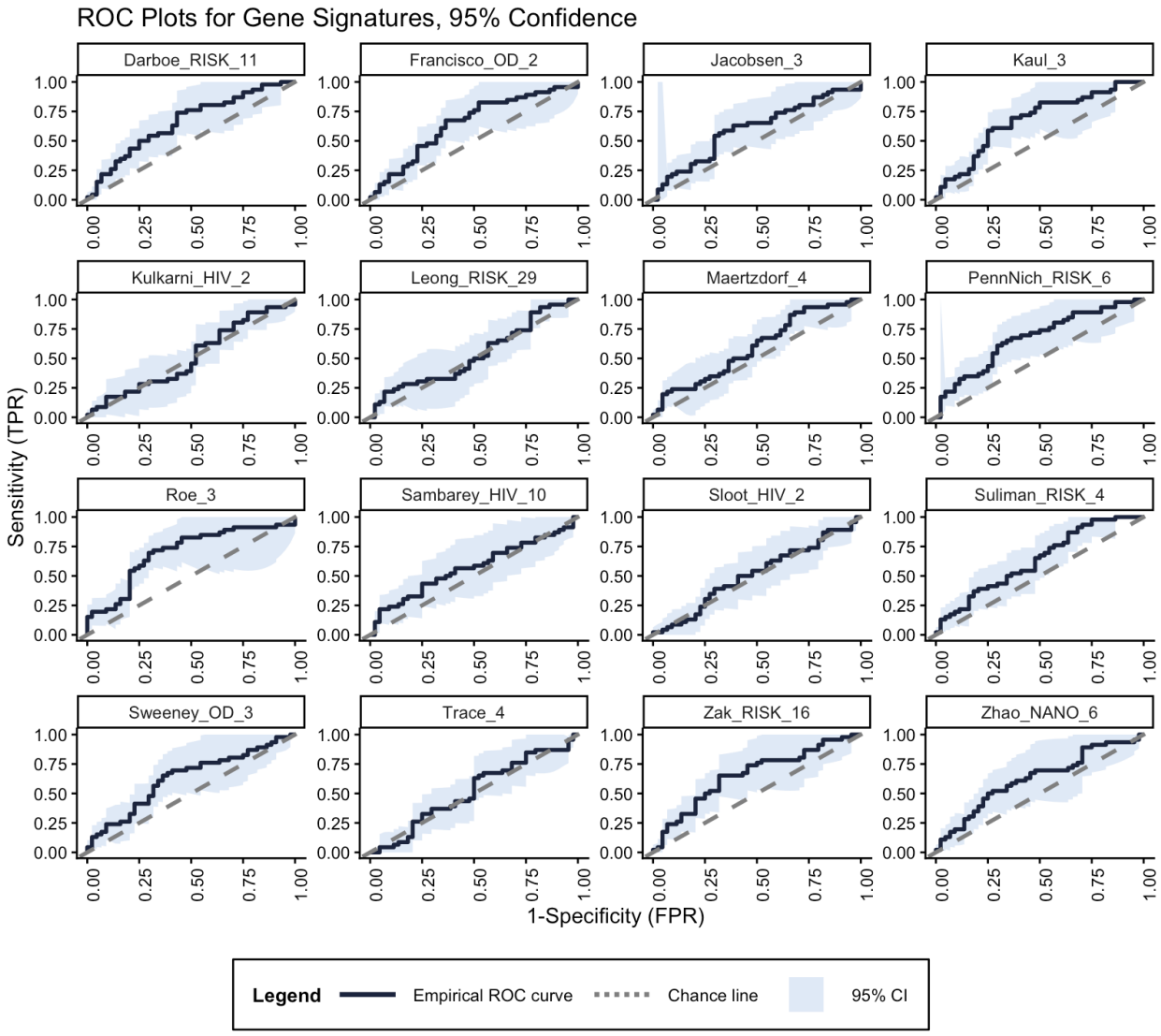



**Figure 8.** ROC plots demonstrating the ability of individual signatures to distinguish between extended microbiological reference standard (eMRS)+ and eMRS- participants. Reference standards were defined as detailed in **Section 2.6**.

**Table 5.** AUC values for the ROC curves shown in **Figure 8**, evaluated against the extended microbiological reference standard. 95% CIs were determined by bootstrapping with 100 resamples. P values were calculated using a two-sided t-test. Reference standards were defined as detailed in **Section 2.6**.

| Signature | P Value | Lower 95% CI, AUC | AUC | Upper 95% CI, AUC |
|---|---|---|---|---|
| Roe_3 | 0.002 | 0.59 | 0.70 | 0.82 |
| Kaul_3 | 0.003 | 0.58 | 0.68 | 0.76 |
| PennNich_RISK_6 | 0.01 | 0.58 | 0.67 | 0.79 |
| Darboe_RISK_11 | 0.01 | 0.57 | 0.66 | 0.77 |
| Zak_RISK_16 | 0.01 | 0.55 | 0.66 | 0.75 |
| Francisco_OD_2 | 0.05 | 0.55 | 0.65 | 0.79 |
| Suliman_RISK_4 | 0.02 | 0.52 | 0.63 | 0.74 |
| Sweeney_OD_3 | 0.05 | 0.52 | 0.63 | 0.76 |
| Zhao_NANO_6 | 0.05 | 0.52 | 0.63 | 0.72 |
| Jacobsen_3 | 0.24 | 0.51 | 0.60 | 0.69 |
| Maertzdorf_4 | 0.10 | 0.50 | 0.60 | 0.71 |
| Sambarey_HIV_10 | 0.25 | 0.51 | 0.57 | 0.68 |
| Leong_RISK_29 | 0.36 | 0.50 | 0.53 | 0.65 |
| Trace_4 | 0.99 | 0.50 | 0.52 | 0.63 |
| Kulkarni_HIV_2 | 0.76 | 0.50 | 0.51 | 0.62 |
| Sloot_HIV_2 | 0.85 | 0.50 | 0.51 | 0.66 |

Although not required for the PLAGE algorithm, for each signature composed of both up- and down-regulated genes, I also calculated the AUCs for each directional component separately. For all bi-directional signatures, this analysis was conducted for each of its directional components (up or down) composed of at least two genes. Diagnostic performance was not improved in any of the uni-directional components analysed alone, besides PennNich_RISK_6 which was minimally (AUC +0.04) improved (**Table 6**).

**Table 6.** AUC values for unidirectional components composed of two or more genes from bidirectional signatures. AUCs were determined using the extended microbiological reference standard. 95% CIs were determined by bootstrapping with 100 resamples. P values were calculated using a two-sided t-test. Reference standards were defined as detailed in **Section 2.6**.

| Signature | P Value | Lower 95% CI, AUC | AUC | Upper 95% CI, AUC |
|---|---|---|---|---|
| PennNich_RISK_6_UP | 0.001 | 0.60 | 0.71 | 0.82 |
| Sweeney_OD_3_UP | 0.002 | 0.57 | 0.67 | 0.77 |
| Suliman_RISK_4_DOWN | 0.06 | 0.52 | 0.63 | 0.73 |
| Zhao_NANO_6_DOWN | 0.08 | 0.52 | 0.63 | 0.74 |
| Maertzdorf_4_UP | 0.04 | 0.51 | 0.62 | 0.71 |
| Suliman_RISK_4_UP | 0.03 | 0.51 | 0.61 | 0.71 |
| PennNich_RISK_6_DOWN | 0.13 | 0.50 | 0.59 | 0.70 |
| Jacobsen_3_UP | 0.37 | 0.50 | 0.56 | 0.70 |
| Leong_RISK_29_DOWN | 0.11 | 0.50 | 0.56 | 0.67 |
| Leong_RISK_29_UP | 0.78 | 0.50 | 0.52 | 0.65 |

Confidence intervals (CIs) between all signatures' AUCs considerably overlapped; as such, it is difficult to definitively rank the performance of each signature. Interestingly, however, the top 5 signatures (Roe_3, in addition to PennNich_RISK_6, Kaul_3, Darboe_RISK_11, and Zak_RISK_16) all were derived from cohorts which included adolescents.

*Subgroup Analysis*

Given this higher performance of signatures derived from adolescent-inclusive cohorts, I sought to determine the influence of age on signature performance, along with the influence of other covariates of interest: malnutrition, disease severity, and sex. For this analysis, the eMRS was utilized to increase the size of the reference standard positive

group. Across the top six signatures, three—Darboe_RISK_11, Zak_RISK_16, and Francisco_OD_2—exhibited enhanced performance in all subgroups. Of the remaining three, subgroup performance was higher than overall performance for all but one evaluated subgroup: for Roe_3 and Kaul_3, children with severe disease; and for PennNich_RISK_6, female children (**Figure 9**). However, wide confidence intervals prevented any observed differences from meaningfully exceeding the overall performance of the signature. It should also be noted that some participants were members of multiple subgroups, as shown in the UpSet plot in **Figure 10**.

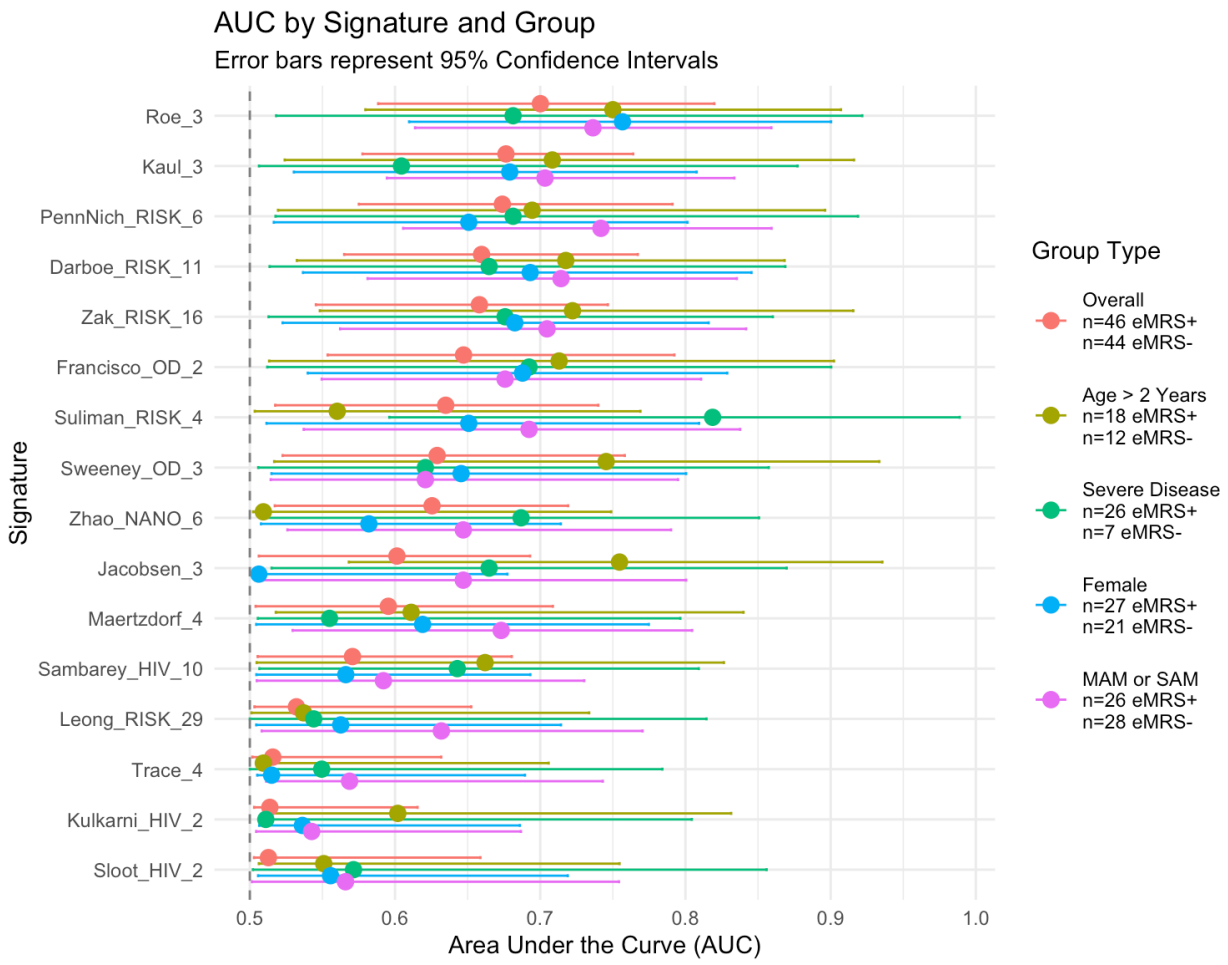



**Figure 9.** AUCs for each signature, evaluated against the eMRS, split by subgroup. 95% CIs were calculated as before, using bootstrapping with 100 resamples. Reference standards were defined as detailed in **Section 2.6**. Abbreviations: Extended Microbiological Reference Standard (eMRS), Moderate Acute Malnutrition (MAM), Severe Acute Malnutrition (SAM).

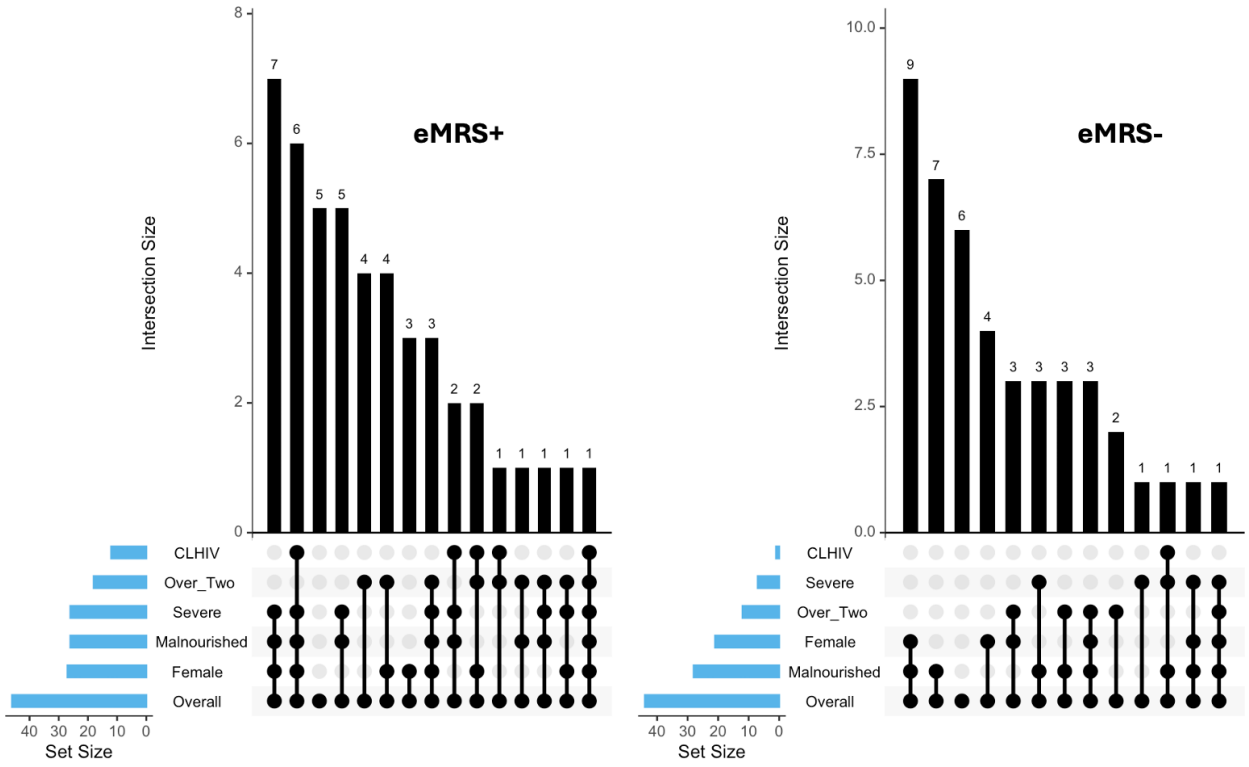



**Figure 10.** An UpSet plot demonstrating the participant overlap between subgroups studied in **Figure 9.** While HIV status was not studied as a subgroup due to the low number of extended microbiological standard (eMRS)- children living with HIV (CLHIV) tested, it is included in the UpSet plot for reference. Reference standards were defined as detailed in **Section 2.6**. The "malnourished" label includes children with both severe acute malnutrition (SAM) and moderate acute malnutrition (MAM).

## 4.8 Ability of Longitudinal Changes to Signature Scores to Retrospectively Identify TB Disease

I also looked to determine how TB signatures changed after two weeks of TB treatment by comparing signature scores from samples collected at both the baseline and two-week visits from participants without drug-resistant TB who were started on treatment at the baseline visit. In total, 35 of the participants with samples tested by NanoString met these criteria to be included in the longitudinal analysis. 27 of these participants had

confirmed TB, 4 were MRS- LAM+ CLHIV, and 4 were classified as unlikely TB. The MRS- LAM+ CLHIV who had been otherwise classified as unlikely TB did not meet inclusion criteria for this analysis. However, to account for how signature scores calculated using the PLAGE algorithm are influenced by the expression levels of other samples in the dataset, I scored all samples (baseline and two-week, from all participants) together for this analysis. Only after computing signature scores for all participants' samples were the samples from the 35 participants selected for longitudinal analysis subset from the overarching dataset.

From the obtained signature scores, I compared the per-participant differences between the two-week and baseline scores across classifications. In so doing, it was observed that no inter-group comparisons reached statistical significance, as shown in **Figure 11**, though this analysis was limited by the small size of the unlikely TB and MRS- LAM+ CLHIV groups. Additionally, I evaluated the discriminatory ability of the difference in two-week and baseline scores to differentiate participants on the basis of the modified strict reference standard. Analysis against the extended microbiological reference standard was performed but is not shown as, given that included reference standard negative cases were the same for both standards, the results were near-indistinguishable between standards. Using the modified strict reference standard, however, I plotted ROC curves for each signature (shown in **Figure 12**) using the difference between two-week and baseline scoring as the evaluated "signature score". Additionally, I calculated the AUC for each of these plots (**Table 7**). While using two-week changes in scores yielded notably higher AUC values for some signatures, the confidence intervals for this analysis precluded

any claims from being made about the approach's performance relative to that of baseline samples alone.

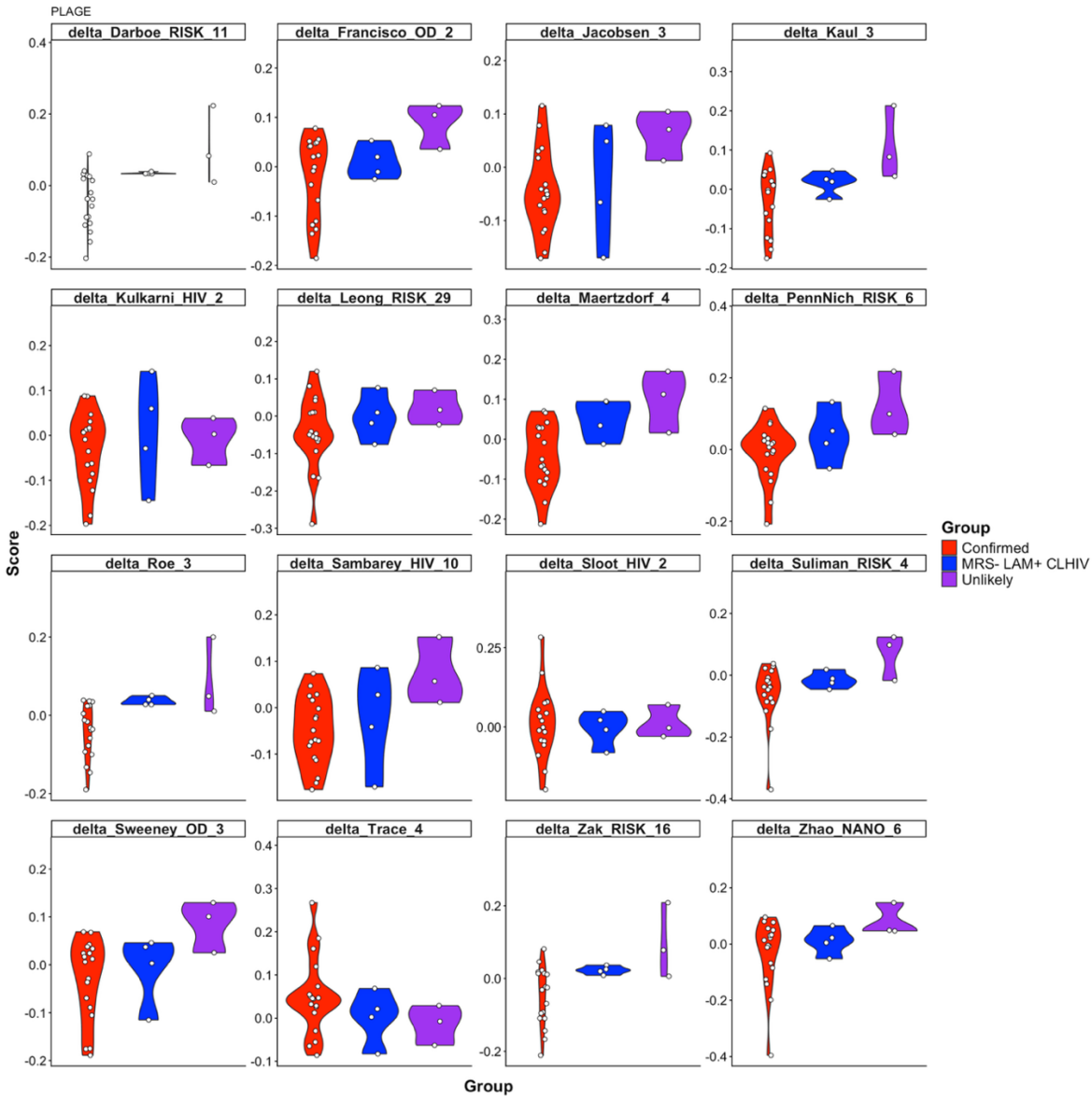



**Figure 11.** A violin plot showing the distribution of differences in PLAGE-determined signature scores between the baseline and two-week visits, of all selected participants, grouped by classification. Significance was determined using a Wilcoxon rank-sum test with the Benjamini-Hochberg (BH) procedure to correct for multiple statistical tests being performed; however, after performing the BH procedure, no inter-group comparisons reached statistical significance. Classifications were determined as outlined in **Section 2.5** and **Section 2.6**. Abbreviations: Lipoarabinomannan (LAM), Children Living with HIV (CLHIV), Microbiological Reference Standard (MRS).

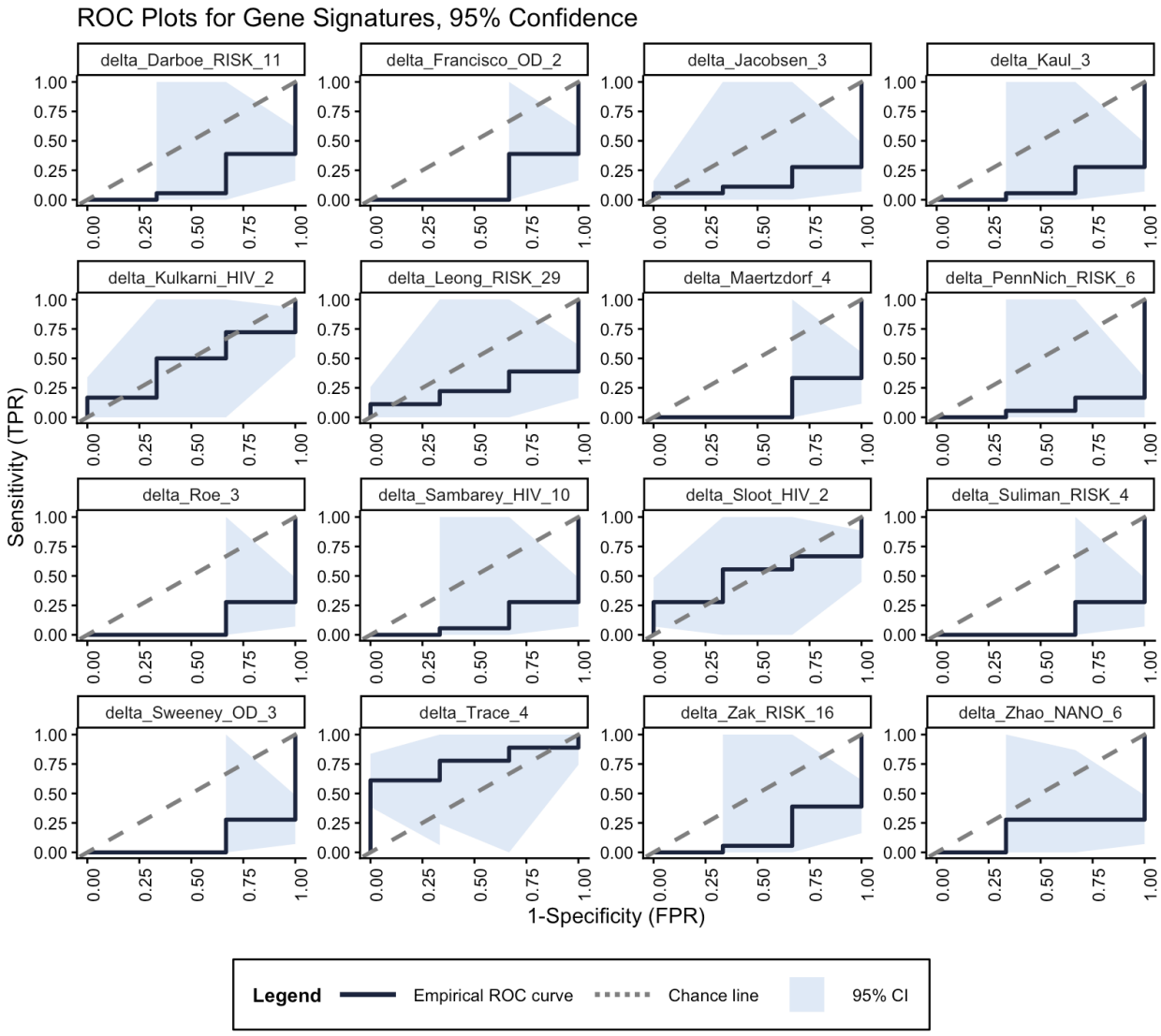



**Figure 12.** ROC plots demonstrating the ability of the two-week change in individual signatures to distinguish between modified strict reference standard (mSRS)+ and mSRS- participants. Reference standards were determined as outlined in **Section 2.6**.

**Table 7.** AUC values for the ROC curves shown in **Figure 12**, evaluated against the modified strict reference standard. 95% CIs were determined by bootstrapping with 100 resamples. P values were calculated using a two-sided t-test. Reference standards were defined as detailed in **Section 2.6**.

| Signature | P Value | Lower 95% CI, AUC | AUC | Upper 95% CI, AUC |
|---|---|---|---|---|
| delta_PennNich_RISK_6 | 0.12 | 0.77 | 0.93 | 1.00 |
| delta_Roe_3 | 0.15 | 0.62 | 0.91 | 1.00 |
| delta_Suliman_RISK_4 | 0.07 | 0.71 | 0.91 | 1.00 |
| delta_Sweeney_OD_3 | 0.04 | 0.58 | 0.91 | 1.00 |
| delta_Kaul_3 | 0.10 | 0.70 | 0.89 | 1.00 |
| delta_Maertzdorf_4 | 0.07 | 0.69 | 0.89 | 1.00 |
| delta_Sambarey_HIV_10 | 0.08 | 0.60 | 0.89 | 1.00 |
| delta_Francisco_OD_2 | 0.03 | 0.60 | 0.87 | 1.00 |
| delta_Darboe_RISK_11 | 0.13 | 0.64 | 0.85 | 1.00 |
| delta_Jacobsen_3 | 0.03 | 0.65 | 0.85 | 1.00 |
| delta_Zak_RISK_16 | 0.12 | 0.57 | 0.85 | 1.00 |
| delta_Zhao_NANO_6 | 0.04 | 0.61 | 0.81 | 1.00 |
| delta_Leong_RISK_29 | 0.12 | 0.59 | 0.76 | 0.93 |
| delta_Trace_4 | 0.12 | 0.53 | 0.76 | 0.95 |
| delta_Kulkarni_HIV_2 | 0.58 | 0.50 | 0.54 | 0.87 |
| delta_Sloot_HIV_2 | 0.98 | 0.50 | 0.50 | 0.80 |

## 4.9 Development of a Novel Transcriptomic Signature from Assessed Genes

Given that the evaluated signatures each did not exhibit strong performance, I explored whether deriving a novel signature from our NanoString data might yield a signature with improved performance in paediatric cohorts like our own. To evaluate this, I first batch-corrected all baseline gene expression data obtained from NanoString testing using the sva package [114]. I then used the $\log_2$ counts per million for all genes evaluated on the NanoString CodeSet, for all participants tested at baseline, to train 200 models to discriminate between mSRS+ and mSRS- cases, using an assortment of machine learning

methods. These methods included generalized linear models with regularization, random forests, and gradient boosting machines. To maximize the training data available for model development, no training/test split was utilized; instead, all models were evaluated with five-fold cross-validation, with the expectation that this approach would slightly overestimate discriminative performance due to overfitting. However, even with this lack of a true test dataset, the best-performing model, a generalized linear model with regularization, only demonstrated an AUC of 0.66 when evaluated against the mSRS. The two next best performing models, both gradient boosting machines, performed similarly, with AUCs of 0.64 and 0.63. Given that these performances were in fact worse than that of other established signatures, even with the lack of a true test set, the performance of these signatures was not investigated further.

However, despite the derived signatures performing sub-optimally, the component genes which had highest importance in the models may still be informative to future signatures. As such, I interrogated each model to ascertain which genes contributed most to the model's ultimate classifications. I determined this by using each gene's relative importance score, which is calculated based off of how much a given gene reduces the squared error across decision trees, with the most important gene being normalized to a score of 1. Several genes appeared in the list of top 10 most important genes for more than one model (shown in **Figure 13**), with three—*DERA*, *SCARF1*, and *TUBGCP6*—appearing in this list for all three models.

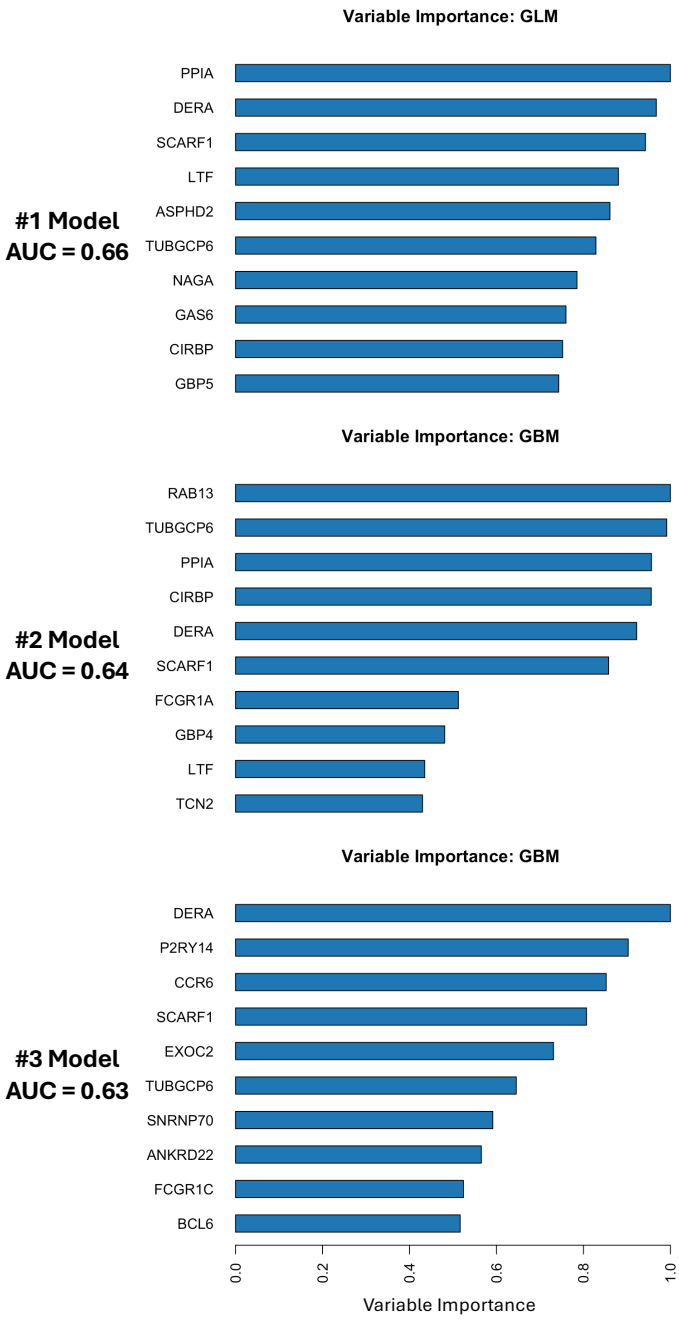



**Figure 13**. The variable importances of the top 10 most important genes to each of the three top models derived from baseline gene expression data from mSRS+ and mSRS- NOD-pedFEND participants, as measured by NanoString. Reference standards were determined as outlined in **Section 2.6**. Variable importance is normalized to one for the gene of greatest importance in each model. Abbreviations: Gradient Boosting Machines (GBM), Generalized Linear Model (GLM).

## 4.10 Correlation Between NanoString- and HRC-Derived TB Signature Scores

Given that the three-gene signature proposed by Sweeney et al. was evaluated by both NanoString and the HRC analysed in the prior chapter, I sought to compare the signature measurements of the two technologies. 31 participants had a baseline blood signature evaluated by both NanoString and the Cepheid Host Response Cartridge (HRC): 17 participants with confirmed TB, 13 with unlikely TB (including one LAM+ CLHIV), and 1 other MRS- LAM+ CLHIV. Nine of these participants—six with confirmed TB and three with unlikely TB—also had a two-week sample measured by both instruments. Of note, while HRC utilized a capillary blood sample measured within an hour of collection, NanoString utilized RNA-stabilized whole blood (collected in PAXgene tubes), which was frozen and thawed before measurement.

To begin, I directly compared the Sweeney_OD_3 score calculated by NanoString to the *KLF2*-based TB score calculated by HRC for each participant, at each measured timepoint, by directly plotting the obtained scores against one another. I then calculated scores for a modified Sweeney_OD_3 signature using PLAGE, which, like with the HRC, replaced *KLF2* with *TBP*. I then plotted these scores against the *TBP*-based TB scores determined by the HRC for the same participants and timepoints, constructing the same graph as was constructed for the *KLF2*-based score. Together, these plots form **Figure 14**. From this analysis, I observed a strong correlation between the signature scores given by both technologies, but particularly for the *TBP*-based signature, which had an $R^2$ value of 0.89.

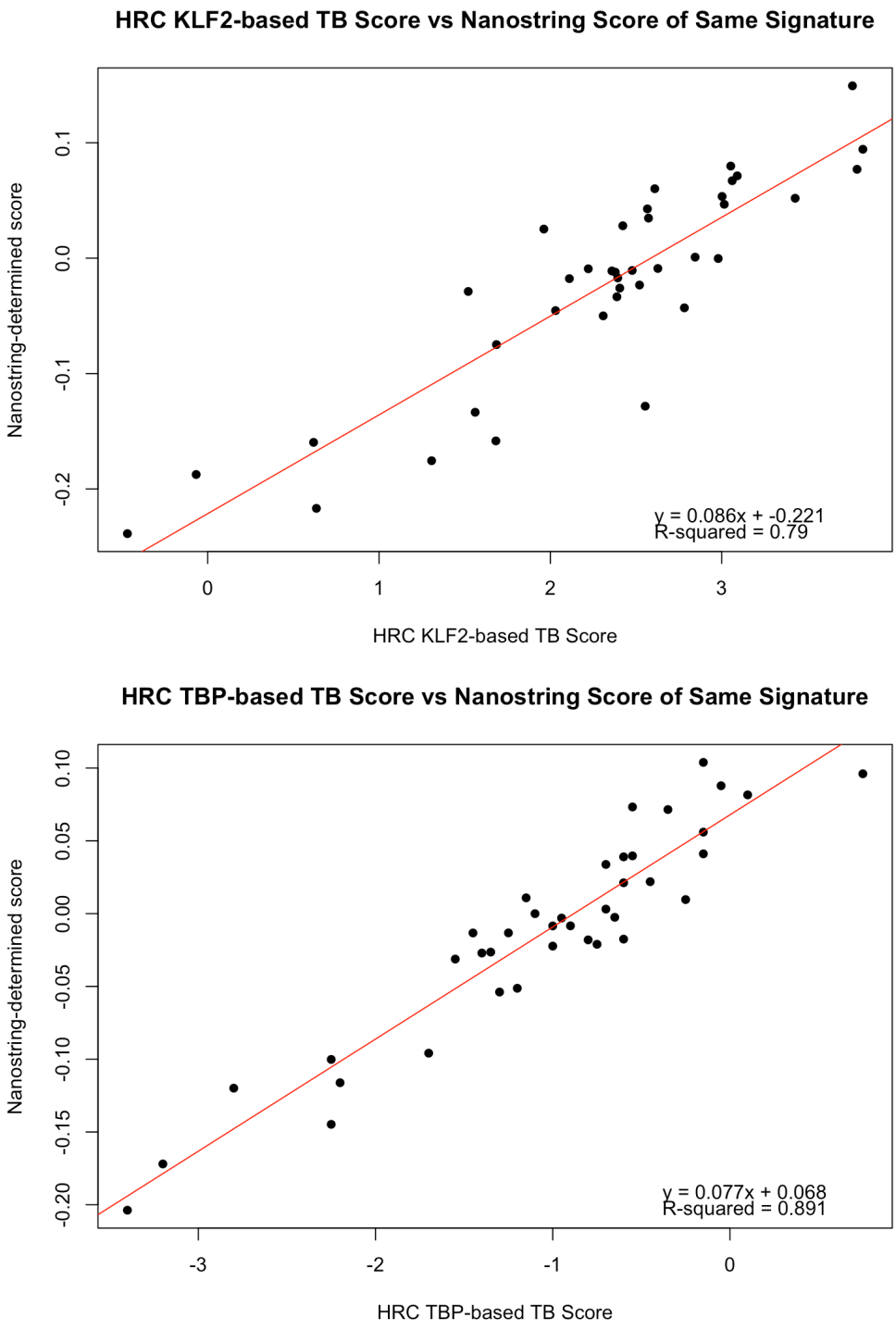



**Figure 14.** The relationship between the PLAGE-determined Sweeney_OD_3 signature score (for both the *KLF2*- and *TBP*-based versions) given by NanoString and the TB score reported by the host response cartridge (for the matched signature), on distinct samples taken, per protocol, on the same day.

Looking at the baseline scores only, I also calculated ROC curves (**Figures 15** and **16)** and their AUCs (**Table 8**) for the original and modified Sweeney signature, as measured by each technology and evaluated against the mSRS and eMRS. I observed similarly poor diagnostic performance for both versions of the signature across both technologies, with a maximum AUC of 0.59.

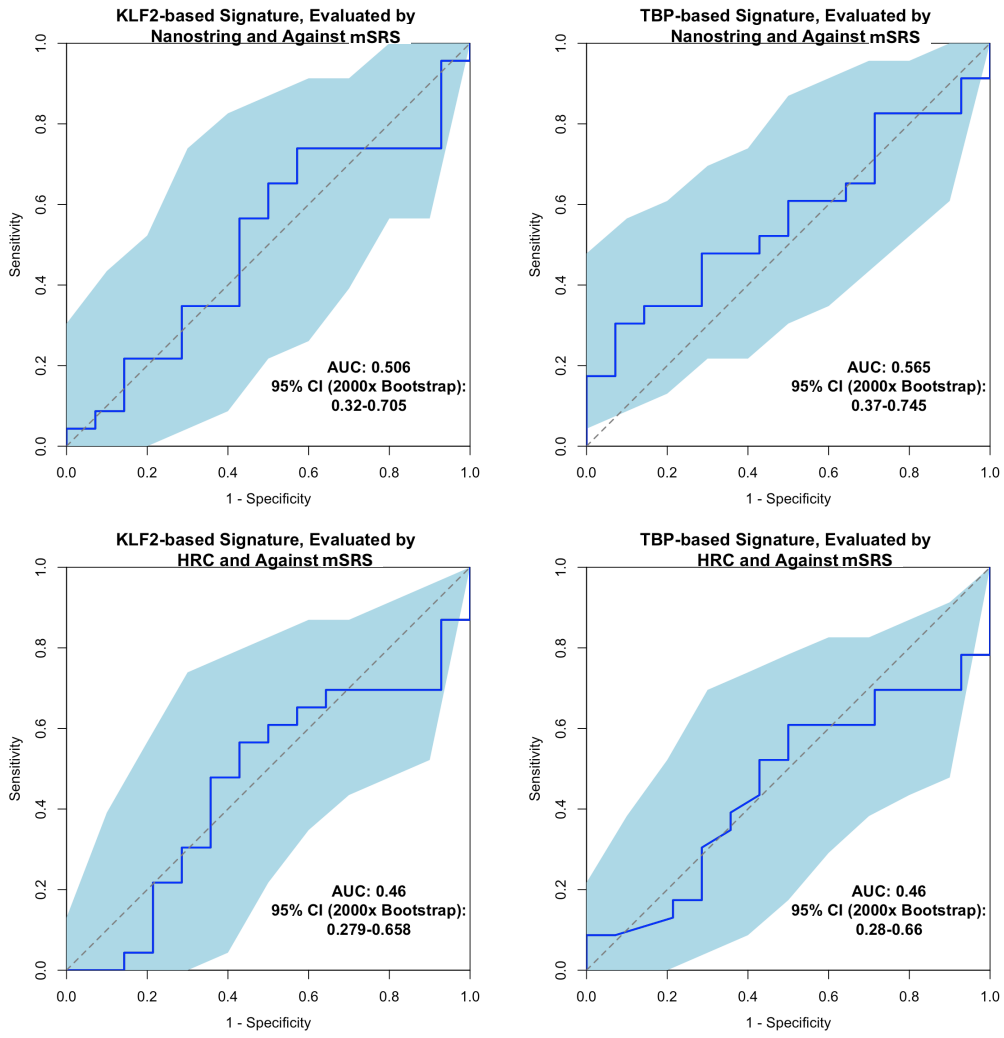



**Figure 15.** ROC curves for both the *TBP*- and *KLF2*-based TB signatures, measured using both technologies and compared against the mSRS. Reference standards were determined as outlined in **Section 2.6**. Abbreviations: Host Response Cartridge (HRC), Modified Strict Reference Standard (mSRS).

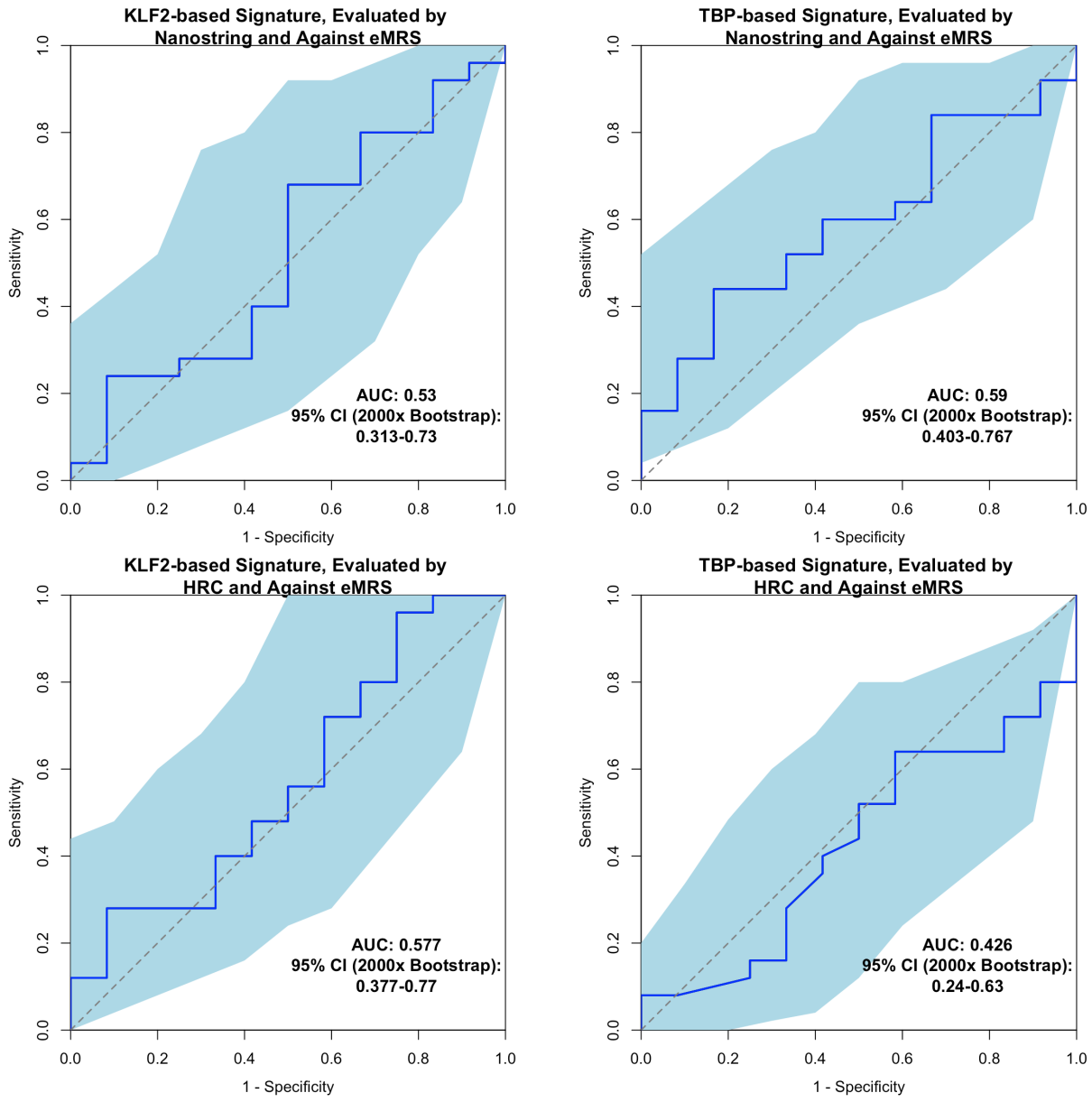



**Figure 16.** ROC curves for both the *TBP*- and *KLF2*-based TB signatures, measured using both technologies and compared against the extended microbiological reference standard. Reference standards were determined as outlined in **Section 2.6**. Abbreviations: Host Response Cartridge (HRC), Extended Microbiological Reference Standard (eMRS).

**Table 8.** AUCs for the ability of NanoString or the HRC to distinguish participants on the basis of the mSRS or eMRS, for each version of the TB signature and reference standard used. 95% CIs, determined by bootstrapping with 2000 resamples, are given in parentheses. Reference standards were defined as detailed in **Section 2.6**.

| | **NanoString** | **Host Response Cartridge (HRC)** |
|---|---|---|
| *KLF2-Based TB Signature* | | |
| ***Modified Strict Reference Standard*** | 0.51 (0.32-0.71) | 0.46 (0.28-0.66) |
| ***Extended Microbiological Reference Standard*** | 0.53 (0.31-0.73) | 0.58 (0.38-0.77) |
| *TBP-Based TB Signature* | | |
| ***Modified Strict Reference Standard*** | 0.57 (0.37-0.75) | 0.46 (0.28-0.66) |
| ***Extended Microbiological Reference Standard*** | 0.59 (0.40-0.77) | 0.43 (0.24-0.63) |

## 4.11 Discussion

I evaluated the performance of 16 individual signatures of TB disease in children under five, finding that no signatures, when evaluated at baseline, reached an AUC value reflective of good diagnostic performance overall (here taken to be 0.80). Longitudinal changes in signature scores did demonstrate some promise; however, sample size limitations reduced interpretability of the observed improvement in AUC values, particularly when accounting for multiple hypothesis testing. Diagnostic performance was also limited even when utilizing the entire gene expression profile measured or a novel signature derived from it. That being said, our data do not rule out the potential value of

transcriptomic biomarkers for paediatric TB and, nevertheless, highlight important considerations for continued work in the field.

*Subgroup-Specific Performance of Signatures*

In reviewing the performance of individual signatures, a key pattern emerged: the seven signatures with the best diagnostic performance at baseline each came from signatures derived from cohorts which included adolescents, in contrast to the 8/9 (89%) of remaining signatures which were developed exclusively in adult cohorts. Multiple high-performing signatures, in fact, originated from the same South African adolescent cohort study, with signatures differing just from the specific samples and computational methods used in signature construction [90]. However, this could also mean that another shared factor, besides age, contributed to the signatures' high performance. That host-based signatures exhibited age-specificity, however, should not come at complete surprise given the unique immune responses generated by young children and adults with TB, as further detailed in **Section 1.2**. Additionally, when I compared the top six signatures' general performance to their performance in multiple subgroups of interest, I observed diagnostic performance to be elevated in, at minimum, all but one of the subgroups evaluated. Performance was elevated in 4 of the top 6 signatures for children with more severe disease, 5/6 for female children, and 6/6 for malnourished children or those who were older (two to five years old, in our cohort). These findings align with the work of others. For example, the RISK6 signature was observed to perform better in identifying individuals with smear-positive TB, compared to those with culture-confirmed smear-negative disease, which supplements my finding on its improved performance in children with severe

disease [99]. The signature also exhibited improved predictive performance in males, which was corroborated by our own results [116]. Additionally, Berry et al. also found their transcriptomic signature of TB disease to perform best amongst individuals with culture-confirmed (due to, likely higher bacillary burden) TB [117]. Conversely, work in a symptomatic South African cohort did not observe any impact of age, HIV status, sex, or sputum-smear status on the performance of Sweeney_OD_3 or Roe_3 [118].

While these results demonstrate some signals of improved performance worthy of future study, they also emphasize the need to better study and characterize the large cohort of children with unconfirmed TB, a consideration further detailed in **Section 6.3**. Given that children with true TB who lack laboratory confirmation of disease are likely to have lower bacillary burdens, host-based tests may also perform worse in this cohort, necessitating new approaches. Moreover, if such children could be robustly included in analyses, it would greatly increase the power by which subgroup-specific conclusions could be drawn.

Additionally, integrating aspects of clinical presentation/demographics, like those I used for subgrouping, into multimodal signatures of TB has also been hypothesized to improve predictive performance compared to transcriptomic signatures alone. In a signature derived from an Indian cohort with pulmonary TB, for example, it was observed that integrating these clinical observations improved the AUC of transcriptomic signatures of TB by between 0.11 to 0.30, depending on the outcome being predicted [119]. A similar observation has been reported for Darboe_RISK_11 [120].

It should be noted, though, that the interpretability of the subgroup-specific findings is limited due to small sample sizes and participant overlap between subgroups. Future work developing transcriptomic biomarkers should keep these findings in mind, however, by focusing on paediatric-specific cohorts in signature development and evaluation, as well as continually evaluating indications for use in particular clinical subgroups. A further limitation of our study comes from our decision to utilize a similar NanoString CodeSet to that used for a study of a primarily adult cohort [93]. While this better aligned our methodology and results with those which had been already validated in adults, because of this alignment in gene selection, the CodeSet did not include genes specific to many paediatric signatures. As a result, I was unable to evaluate several signatures specifically developed in paediatric cohorts, including those proposed by Gjøen et al. and Chen et al. [121, 122]. Given the age-specificity I observed in our results, evaluating these signatures in the NOD-pedFEND cohort too is a promising avenue for future research.

*Influence of Low-Expression Genes*

An additional caveat which should be considered in the interpretation of our data is the fact that 12 genes had measured expression levels at or below the background threshold value for >80% of all tested samples. As a result, signatures which rely on those genes may exhibit improved performance on ultra-sensitive, amplification-based assays, as compared to the NanoString-based analysis. This is likely seen in the performance of Sloot_HIV_2, which is solely determined by the expression levels of AIRE and IL-13; both genes were often below the limit of detection of the assay. However, for all 187 samples analysed for this project, the limit of detection values obtained indicated that, given the

RNA content of an average cell, RNA was being detected at a level lower than two transcripts per cell [115]. As such, the 12 genes often below the limit of detection likely had expression levels lower than this, on average, in the non-erythrocytic cells captured in whole-blood samples. Given this, signatures reliant on these 12 genes would likely have difficulty with robustness and, potentially, interpretability, as random noise/fluctuations in measured expression levels could dramatically alter the reported signature score. Additionally, assays evaluating these signatures would have to be incredibly sensitive, limiting available technological approaches for clinical rollout.

It is also possible, however, that the probes for these "low-expression" genes did not adequately hybridize with their target RNA. While validation work is needed to conclusively exclude this possibility, the CodeSet's prior successful use for detection of these low-expression genes (both AIRE and IL-13, for example) [93] helps to alleviate technical concerns. As these genes have been measured at higher expression levels in other studies run on the same NanoString CodeSet [93], it also appears likely that these genes simply aren't highly expressed in our cohort. Nevertheless, a dedicated positive control sample for genes with low measured expression may be of use in future experiments to confirm the accuracy of these measurements of gene expression.

*Comparison with HRC*

I also compared the performance of the original and modified three-gene signature proposed by Sweeney et al., as measured by both NanoString and the HRC [61]. I observed strong correlation between the two approaches for measurements of the signature, building confidence in the use of either technological approach for gene expression

analysis. This result is particularly impressive given that both technologies used related, but distinct samples, with the HRC relying on capillary blood tested within one hour and NanoString on PAXgene tubes that were frozen and thawed before testing. Additionally, while all blood samples, including capillary blood and the utilized PAXgene sample, were typically collected at around the same time of day, this was not always possible given resource constraints. As a result, other environmental factors such as time of day or recent food or drink—which have been shown to influence gene expression profiles in prior work [123]—could also have influenced some HRC and NanoString results. That the Sweeney signature appears minimally affected by these influences would be exciting if the signature itself performed better. However, the similarity in signature scores between those obtained from capillary and whole blood still supports the potential of capillary-blood-based tests. Given the implementation advantages of the sample type, attempting to measure and develop future signatures on capillary blood would be a worthwhile approach.

It should also be noted that the scoring approach used by NanoString is fundamentally different to that used by assays developed for commercialization. Assessment of signatures by NanoString, for our analysis, relied on a score determined by the PLAGE algorithm. As the PLAGE algorithm's scoring is influenced by other samples in the dataset, it is possible that cohort selection could influence our results. Nevertheless, for this analysis comparing HRC to NanoString, I attenuated this risk by calculating signature scores for all samples prior to subsetting to only those from participants with matching HRC data. Despite this limitation of the PLAGE algorithm, however, it otherwise

remained best suited for our data, given the low number of non-TB-related genes in the dataset and the use of signatures with both up- and down-regulated components.

*Novel Signature Development*

Using the full set of gene expression data, I also developed a novel signature for TB specific to our cohort, which demonstrated performance worse than that of the highest performing established signatures I evaluated, even despite being evaluated in cross-validation from the training set. That performance was poor despite the possibility of overfitting suggests that, for our samples, the input data was simply insufficient to adequately differentiate within our reference standards. This is further supported by the lack of clear clustering on the dendrogram, PCA, and sPLS-DA and raises concerns about the genes tested in our targeted approach. This could, however, be due to a number of factors beyond the genes chosen for measurement, including flaws in the reference standards utilized or biological heterogeneity; additionally, relatively wide confidence intervals could have led to meaningful underestimation of diagnostic performance simply by random chance.

Lessons may still be found in the most important genes in the derived signatures, however; it may be that some genes important to a TB signature were correctly identified but signature performance was diminished by lacking data from other critical genes for which I didn't test. The top genes identified in the three chosen signatures were Peptidylprolyl Isomerase A (PPIA), a gene implicated in HIV replication along with protein folding and inflammation [124]; RAB13, which is involved in macrophage polarization [125]; and deoxyribose-phosphate aldolase (DERA), which contributes to the stress response

[126]. These genes may represent promising targets for future signatures of TB disease, developed from broader gene expression datasets.

*Future Work and Next Steps*

While our results do not refute the potential of transcriptomic signatures of paediatric TB disease, they do indicate that signatures of disease derived from adult cohorts—at least when assessed on whole-blood samples taken at baseline—may not be able to be translated to the paediatric population. As a result, future work may benefit from less targeted approaches to gene expression, such as bulk RNA-Seq, to enable inclusion of additional genes which might assist in case detection. However, such work has been previously described in other paediatric cohorts [127], so understanding how those derived signatures perform, and how their performance might be able to be improved, would allow for meaningful iteration without repetition. Ultimately, if signatures derived from dedicated paediatric cohorts also do not generalize sufficiently to other paediatric cohorts, it may become time to consider alternative approaches.

Given the more subdued immune response of young children to *Mtb*, alternative sample types than the unstimulated whole blood used for this analysis may be necessary to induce a sufficient immune response to meaningfully distinguish between TB cases and controls. One approach might be to use RNA-stabilized blood samples stimulated with TB antigens, as has been recently described in children [128], and which might also increase the expression levels of low-expression genes in our cohort, like IL-13. Such samples were collected in NOD-pedFEND, and their analysis is planned. Additionally, a multi-modal signature of disease—using demographics, clinical characteristics, or other biomarkers,

such as cytokines or metabolites, in conjunction with transcriptomic markers—may also perform better than any one approach alone. Ultimately, further research is needed to identify more effective approaches for developing host-based biomarkers of paediatric TB.

*A Note on Priorities and Experimental Design*

In considering future approaches, a crucial differentiator between the priorities of adult and paediatric TB research emerges. Biomarker research in adult TB focuses in large part on identifying a signature which can be translated to a commercial diagnostic test, ideally one which could be deployable at the point of care. However, research in paediatric TB is distinguished by a challenge that overshadows a range of research activities: knowing, for near certain, who does and does not have TB. As a result, even complicated or retrospectively determined biomarkers, if able to be implemented in research studies, could meaningfully advance paediatric tuberculosis research by supporting improved cohort selection and retrospective classification of disease status.

This thinking also guided our sample selection strategy. Given the uncertainty that undergirds paediatric TB research, one approach to biomarker discovery has been to compare children with confirmed TB to healthy controls (with or without TB infection), as the two groups are thought to be known with the highest degree of certainty. However, while this approach can assist in a preliminary, less resource-intensive assessment of a biomarker, solely structuring comparisons in this manner presents two challenges which can limit the applicability of any signature which emerges. First, as healthy controls are by definition not symptomatic, resultant signatures can simply be signatures of inflammation or illness that are not specific to TB. To combat this, I primarily structured our comparison

against the modified strict and extended microbiological reference standards, believing that the comparison of two symptomatic groups was worth the slightly increased potential for misclassification. For both standards, misclassification could occur if a participant in the unlikely TB group in fact had TB disease, or if a participant's positive microbiological testing result was incorrect. While a false positive culture or Xpert result would equally affect a comparison with healthy controls, false positive results are of greater concern for the eMRS (specifically, the group of LAM+ CLHIV), as LAM tests have a specificity of only around 76% in CLHIV [67].

The second concern, which influences all these approaches, is the lack of study of children with unconfirmed TB who are not LAM+ CLHIV, as children with unconfirmed TB often make up a majority of presumptive TB cases (including in NOD-pedFEND) and are the population in greatest need of a novel diagnostic [129]. However, given that current signatures fail to discriminate between confirmed and unlikely TB well (as measured by the mSRS), it would be difficult to determine whether elevated signature scores in children with unconfirmed TB were likely due to true disease or misidentification. For this reason, I did not include children with unconfirmed TB (unless they were a LAM+ CLHIV) in my analysis. If a signature is discovered and validated to perform well against the mSRS/eMRS, however, an immediate next step would be to evaluate how that signature classifies children with unconfirmed TB, including in NOD-pedFEND or other paediatric cohorts. In sum, I believe our sample selection approach best balances diagnostic certainty with clinical relevance amidst the ambiguity which complicates paediatric TB research.

## 4.12 Conclusion

By performing targeted gene expression analysis on blood samples collected from a subset of children from the NOD-pedFEND study, I evaluated the diagnostic performance, in children < 5 years old, of 16 previously reported transcriptomic signatures of TB. The best signature—that developed by Roe et al.—demonstrated an AUC of 0.70 against the eMRS in the cohort overall. Diagnostic performance appeared to improve in subgroups of interest [100]; limited sample sizes, however, warrant caution in interpreting these findings. Similarly, while longitudinal changes in signature scores post-treatment appeared to exhibit relatively strong diagnostic performance, small sample sizes require further study. Finally, I compared the signature scores produced by NanoString and the HRC and found strong correlations between the two technologies, supporting each technology's precision as well as the use of capillary blood samples for future transcriptomic signatures of TB [61].

Taken together, I found limited utility in evaluating transcriptomic signatures of adult-type TB in symptomatic infants and young children at baseline. My findings emphasize the importance of the tested population in signature evaluation. They also, however, highlight promising approaches for future study, including the use of longitudinal changes in signature scores to retrospectively diagnose children with TB. Subgroup-specific performance improvements and alternative sample types also merit further investigation, which—our results would suggest—would likely be best accomplished using a non-targeted approach to measuring the transcriptome, given the observed difficulty in developing signatures from genes already implicated in adult-type disease. Ultimately, these results demonstrate the difficulty of translating signatures of adult-type TB disease to

TB in children, suggesting that the development of novel signatures of TB for this population is likely necessary.

# Chapter 5: Proteomic Signatures of Paediatric TB in Blood Plasma

## Introduction

### 5.1 Reported Proteomic Signatures of Tuberculosis

*Approaches to Proteomic Signature Development*

Given the lack of a high-performing, generalizable transcriptomic biomarker of paediatric TB, proteomic signatures for TB infection and disease have also been explored [130]. As with transcriptomic biomarkers, this work has been undertaken via both targeted and untargeted approaches. Targeted approaches, in which a limited number (typically on the order of 10s) of proteins are pre-selected for measurement, include enzyme-linked immunosorbent assays (ELISAs) and newer technologies which expand upon the core principles of ELISAs, including Meso Scale Discovery (Meso Scale Diagnostics; Rockville, Maryland, USA) [131] and Luminex (Luminex; Austin, Texas, USA) [132] assays. Using these technologies, researchers can measure the concentrations of multiple, specific proteins in biological samples at once, typically at a substantially lower cost than broader approaches. In contrast, a broad, proteome-wide search can be performed, which measures the levels of thousands of proteins not necessarily thought to be implicated in TB. While mass spectrometry is the classic methodology employed for this untargeted approach, newer affinity-based technologies, such as O-Link (Thermo Fisher Scientific; Uppsala Science Park, Sweden) and SomaScan (Illumina; Boulder, Colorado, USA), have

also been developed [133, 134]. While these technologies allow for less-biased evaluations of the full proteome, they are also expensive and, by virtue of measuring a far greater number of proteins, also have drastically higher false discovery rates which, if corrected for by raising significance thresholds, can in turn increase the risk of masking true biological signals.

*Existing Proteomic Signatures of Paediatric TB*

As a result, most work developing proteomic signatures of paediatric tuberculosis has utilized a targeted approach. Using Luminex assays and similar technologies, levels of specific protein biomarkers of interest in, typically, blood plasma samples, have been compared between children with and without TB. While less common, plasma from children with TB has also been evaluated by an untargeted approach, utilizing mass spectrometry [135]. To my knowledge, however, only one recent evaluation, performed by Fossati et al., utilized mass spectrometry to develop a unique proteomic signature of paediatric TB [136]. Together, these targeted and untargeted evaluations have yielded several proteomic signatures of paediatric TB, which are detailed in **Table 1**. While many signatures have been reported to perform well, with AUCs ranging between 0.75 and 1.00, most have not been evaluated in an independent test set or external validation cohort, making their generalized performance uncertain. Moreover, as most signatures were developed using Luminex, with the concentrations of many of the same proteins measured between studies, by the same/similar antibodies, it is possible that promising signatures of paediatric TB are being missed just because the field has yet to broaden its methodological approach. Given this gap in the field, we sought to use blood plasma samples collected

during NOD-pedFEND to validate and develop proteomic signatures of TB disease using a yet-unused technology for biomarker discovery: Meso Scale Discovery (MSD) V-Plex assays.

**Table 1.** Examples of reported blood-based proteomic signatures of paediatric TB. The directionality of changes in protein concentrations is shared if clearly reported. Abbreviations: Body Mass Index (BMI), Composite Reference Standard (CRS), Healthy Controls (HCs), Strict Reference Standard (SRS), TB Disease (TBD), TB Infection (TBI).

| Citation | Technology Used; Signature Purpose | Sample Type | Proteins | Reported Performance |
|---|---|---|---|---|
| *Kumar, 2021 [137]* | Luminex; CRS+ TB | Blood Plasma | CCL1, CXCL1, CXCL10 (all higher in TB) | AUC of 1.00 in the training set |
| *Fossati, 2025 [136]* | Mass spectrometry; SRS+ TB | Blood Plasma | A three-, four-, five-, and six-protein signature, each with some of the following proteins: APOM, MMP2, and AMBP (lower in TB); TNC, CD44, HEG1, IGHV3-33, IGKV1D-33, FCGR3A, and WARS1 (higher in TB) | AUCs ranging from 0.79-0.89 in the test set |
| *Kumar, 2021 [138]* | Luminex; CRS+ TB | Blood Plasma | TNF-α, IL-17A, IL-2 (all higher in TB) | AUC of 1.00 in a validation cohort |
| *Chiappini, 2012 [139]* | ELISPOT against AlaDH antigen; TBD vs TBI | PBMCs | IL-2 (higher in TB) | AUC of 0.90 in the full cohort |
| *Tebruegge, 2015 [140]* | Luminex; TBD vs TBI | Stimulated Whole Blood | TNF-α, combined with either IL1-RA or IL-10 (all higher in TB) | Classification accuracy of 96% and 100% in the full cohort, respectively |
| *Lighter-Fisher, 2010 [141]* | Luminex; TBD vs TBI | Stimulated Whole-Blood Supernatants | TGF-β1 (higher in TB), IL-2 (lower in TB) | Not evaluated |

| | | | | |
|---|---|---|---|---|
| *Gourgouillon, 2012 [142]* | Luminex; TBD vs TBI | Stimulated Whole-Blood Supernatants | TNF-α (higher in TB), IL-2 (lower in TB) | 88% sensitivity and 83% specificity in the full cohort |
| *López-Suárez, 2024 [143]* | Luminex and sandwich immunoassays; TBD or TBI vs controls without TBI | Blood Plasma | Multiple signatures, which included: IFN-γ, and IL-2 (higher in TB); MCP-1 (lower in TB) | AUC of 0.82 for both TBD or TBI vs controls, as evaluated in the full cohort |
| *Jenum, 2016 [144]* | Luminex; TBD vs HCs | Stimulated Whole Blood | No differences in the proteomic biomarkers assayed were observed | N/A |
| *Togun, 2020 [145]* | Luminex; TBD vs OD | Blood Plasma | IL-1RA, IL-7, IP-10 (all higher in TB), along with age group and BMI-for-age z-score | AUC of 0.75 in the test set |
| *Sudbury, 2019 [146]* | Luminex; TBD vs TBI vs OD vs HCs | Stimulated Whole-Blood Supernatants | IL-1RA, TNF-α (both higher in TB) | AUCs individually assessed in the full cohort: 0.92 (IL-1RA), 0.70 (TNF-α) |
| *Dhanasekaran, 2013 [147]* | Luminex; TBI vs uninfected controls | Stimulated Whole-Blood Supernatants | IL-2, IL-13 (both higher in TB infection) | Not evaluated |
| *Kumar, 2025 [148]* | Luminex; TBD vs OD | Blood Plasma | MMP-1, MMP-2, MMP-9 (all higher in TB) | AUC of 1.00 in the full cohort |
| *Druszczynska, 2021 [149]* | Luminex; TBD vs TBI | Serum | APRIL/TNFSF13, IL-2, IL-8, LIGHT/TNFSF14, MMP-1, MMP-3, osteocalcin (higher in TB); sCD30/TNFRSF8, IFN-α2, IFN-γ, sIL-6Rα, IL-11, IL-29/IFN-λ1, MMP-2, osteopontin, TSLP, and TWEAK/TNFSF12 (lower in TB) | Not evaluated |

## 5.2 Meso Scale Discovery V-Plex Assays for the Validation of Proteomic Signatures

Similar to Luminex, MSD assays rely on the principle that underlies traditional ELISA-based methods of protein quantification: utilizing antibody binding to detect proteins of interest. Unlike Luminex or ELISAs, however, MSD assays differ primarily in how samples are bound to the assay-specific plate and how detection antibodies produce a signal which is read by the instrument. For sample binding, MSD assays rely on capture antibodies which bind the protein of interest and which are immobilized on patterned carbon electrode surfaces localized to a small area at the bottom of each well. This configuration offers more stable protein immobilization than the polystyrene surfaces typically used in conventional ELISAs. This localized, strong attachment within each well allows there to be multiple "islands" of distinct capture antibodies, targeting different proteins, placed within each well. This, in turn, enables multiplexing, as multiple proteins can be captured at once—the instrument simply must know where to look in each well for each protein of interest.

After the proteins of interest (the "analytes") are captured, a detection antibody is added which also binds to the analytes. As the detection antibodies bind directly to the analytes, the number that remain after a washing step correlates to the concentration of the analyte itself. To measure the amount of bound detection antibodies, proprietary electrochemiluminescent labels are conjugated to each detection antibody such that they emit light when the assay plate's electrodes are charged by the MSD instrument; it is this process that is most unique to MSD. As the electrodes are charged, the total light emitted

from each section of the well is measured by the MSD instrument and converted into analyte concentration measurements using a standard curve [150].

These alterations to the basic ELISA methodology yield distinct advantages. MSD assays generally have a much broader dynamic range than traditional ELISAs and can use a smaller volume of sample. MSD assays are also thought to exhibit lower background signal than ELISAs, as only detection antibodies near the electrodes' surface will be excited [150]. Additionally, simply by using a different technology, testing different protein targets using different antibodies, we gain further benefits from diversifying the technology used for protein measurement and subsequent signature development [151].

With these technological benefits in mind, we sought to use MSD assays to both validate previously reported proteomic signatures of paediatric TB and to derive novel signatures of paediatric TB disease in our cohort. Additionally, using samples collected at a visit conducted two weeks after enrolment, we sought to evaluate whether changes in concentrations of the assessed protein biomarkers differed between TB-treated individuals with and without TB disease. To enable these aims, we utilized the most comprehensive set of MSD biomarker assays commercially available: a 54-plex test, split across 7 distinct panels, which detected a range of protein biomarkers, most of which were not related to any of the genes analysed in **Chapters 3 or 4**.

## 5.3 Methods

*Sample Selection*

If available, a blood plasma sample from the baseline and two-week visits was tested from all participants previously selected for the first round of the NanoString project. The selection criteria for these participants are further detailed in **Section 4.4.** In short, however, participants were selected for the NanoString project from among those with confirmed TB and unlikely TB. While all samples available at the time from individuals with confirmed TB were selected, samples from children with unlikely TB were selected to enrich for malnutrition, CLHIV, and participants who had been selected for other NOD-pedFEND projects. Additionally, after selecting all of the limited number of available samples from individuals with confirmed TB, CLHIV with a positive LAM test were also selected up until the total number of eMRS+ individuals selected matched the number of eMRS- individuals selected. For the MSD project, any available baseline and/or two-week plasma sample from these participants was utilized, to enable harmonization with the first-round NanoString data.

*Plasma preparation and MSD V-Plex Protocol*

Plasma samples were obtained as described in **Chapter 2**. For this test, the frozen plasma samples were thawed at room temperature for one hour, after which they were either immediately centrifuged at 2,000xg for 3 minutes or stored at 4-8$^{o}$ C for centrifugation on the same day, performed with the same settings. Once centrifuged, samples were loaded onto one of two sets of seven MSD V-Plex plates, each detecting

between four and ten proteins of interest. The panels used, along with the proteins they detect, are detailed in **Table 2.**

**Table 2.** The MSD V-Plex Panels used for this work and the proteins they each detect.

| MSD V-Plex Panels | Proteins Assayed | Sample Dilution Used |
|---|---|---|
| *Cytokine Panel 1 (Human)* | GM-CSF, IL-1α, IL-5, IL-7, IL-12/IL-23p40, IL-15, IL-16, IL-17A, TNF-β, VEGF-A | 1/2 |
| *Cytokine Panel 2 (Human)* | IL-17A/F, IL-17B, IL-17C, IL-17D, IL-1RA, IL-3, IL-9, TSLP | 1/4 for 8 samples, then 1/2 for remainder (dilution changed to optimize use of the assays' dynamic range) |
| *Vascular Injury Panel 2* | CRP, ICAM-1, SAA, VCAM-1 | 1/1000 |
| *Metabolic Panel 1* | C-Peptide, GIP (active), GLP-1 (active), Glucagon, Insulin, Leptin, PP | 1/4 |
| *Proinflammatory Panel 1* | IFN-γ, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, IL-12p70, IL-13, TNF-α | 1/2 |
| *Angiogenesis Panel 1* | FGF (basic), PlGF, Tie-2, VEGF-A, VEGF-C, VEGF-D, VEGFR-1/Flt-1 | 1/5 |
| *Chemokine Panel 1* | Eotaxin, Eotaxin-3, IL-8 (High Abundance), IP-10, MCP-1, MCP-4, MDC, MIP-1α, MIP-1β, TARC | 1/4 |

Standards of known concentration were prepared and loaded onto the plates in accordance with manufacturer protocols. For the first plate of each panel, standards were run in duplicate. For the second, standards were run in triplicate. The first plate was run as a partial plate twice, on distinct days, using unused wells each time. Standards were not

remeasured on the second round of testing with the first plates, an approach which was previously validated by others in the laboratory.

Panels were run in accordance with manufacturer protocols on a MESO QuickPlex SQ 120MM, except DPP-IV inhibitor was not added to plasma samples immediately after collection, as I utilized biobanked specimens which had already been collected via a previously established protocol. As the addition of this inhibitor would typically prevent the degradation of active GLP-1 in collected blood plasma, measured active GLP-1 levels were not utilized in my analysis.

*Data Processing*

Recorded electrochemiluminescent signal intensity was converted to protein concentration using a standard curve constructed from the eight standards of known concentration using the Meso Scale Discovery Workbench software v4.0.13. The standard curve was constructed using a four-parameter logistic fit, with $1/y^2$ weighting, in accordance with manufacturer recommendations. Standards were only used for standard curve construction if they met pre-determined QC thresholds adapted from published guidelines [152, 153]. In line with standard practice, QC thresholds were dictated by the standard's position in the standard curve. The outer standards, which formed the top and bottom points on the standard curve, were included if their calculated concentration yielded a mean recovery of between 65-135% compared to their known concentration, and if the coefficient of variation (CV) between technical replicates' calculated concentrations was ≤30%. For inner standards, thresholds tightened, restricting standards to between 75-125% mean recoveries and a CV ≤25%. The exclusion of standards which did not meet

these thresholds was a systematic process which proceeded iteratively, for each assay, from the outside-in, with the standard curve's parameters, and resultant QC values for remaining standards, re-estimated after each exclusion. Assays were only included in the analysis if at least four of the assay's standards, on the plate being evaluated, met QC criteria. Accordingly, results for samples on one plate (not both) performing the following assays were excluded: Eotaxin-3, IL-2, IL-4, IL-13, IL-12p70, IFN-γ, MIP-1β, and TNF-α. 8/108 (7%) total plates were excluded.

Of the remaining assays, the lower limit of quantification (LLOQ) for calculated concentration values was determined using the lowest concentration standard which had a calculated concentration (based upon the standard curve) within 70%-130% of the standard's known concentration, and <25% CV in the calculated concentration between all technical replicates of the same standard. This LLOQ was then used as the lowest allowable concentration value for the analyte in tested samples; concentrations for analytes below this value were adjusted to $\frac{LLOQ}{\sqrt{2}}$, in line with established protocols [154]. After this adjustment was made, and data from one well on one plate was excluded due to a noted technical error in assay processing, assay values were qualitatively reviewed for outliers.

Three sets of outliers were identified and removed from any subsequent analysis. The first set of outliers consisted of those from a set of samples physically located in adjacent locations on one Cytokine Panel 2 plate. These 12 samples consisted of both baseline and 2-week samples from 2 participants with confirmed TB and 4 with unlikely TB. Samples in this region of the plate all had analyte values that consistently fell below the

LLOQ, including for cytokines—such as IL-1RA—for which all other samples had concentrations above the LLOQ. The second set of excluded measurements consisted of all those obtained for the IL-8 (high-affinity) assay, as the high-affinity assay yielded relatively low signal and the concentration values measured by the standard-affinity IL-8 assay were, for all samples, below the manufacturer's reported LLOQ for the high-affinity assay. Finally, Eotaxin-3 values for all samples were also excluded, as while one plate was already excluded due to not having four or more standards which met QC thresholds, the other plate had no samples with calculated Eotaxin-3 concentrations which exceeded the assay's calculated LLOQ. All remaining assay results were included in my analysis.

*Statistical Analysis and Machine Learning*

Compiled results were imported into and analysed in R v4.5.2 using tidyverse v2.0 for data wrangling and ggplot2 v4.0.3 for data visualization. pROC v1.19.0.1 was used for construction of the ROC curves, UpSetR v1.4 for the UpSet plot, and DESeq2 v1.50.2 for calculation of the principal component analysis. Statistics were performed as described in **Section 2.8**. When indicated, corrections for multiple hypothesis testing followed the Benjamini-Hochberg (BH) procedure.

Given that two plates were needed to test all samples on each of the seven panels, batch correction was performed on $\log_{10}$-transformed concentration values using the ComBat algorithm as implemented in the sva package v3.58.0 [114], with the outcome of interest set as eMRS status. As the ComBat algorithm requires all datapoints to be known, any excluded assay datapoints were imputed using k-nearest neighbours, based upon the

10 most similar samples. Imputed values were removed after batch correction and were not used for any subsequent analysis, unless otherwise indicated.

The ability of the full panel of 52 biomarkers (post exclusions) to identify SRS+ children was determined by fitting various machine learning models to the log-transformed MSD dataset using the AutoML function encoded in the H2O package v3.46.0.10. Due to the potential for data leakage, batch corrections were not performed on the dataset used for any machine learning. Deep-learning models were also avoided due to the relatively small number of samples tested. Additionally, given the small sample size, no training/test split was performed; rather, models were trained and evaluated using six-fold cross-validation. The resultant AUCs calculated for the trained models, then, are acknowledged to likely be overestimates of what diagnostic performance would be in an independent cohort.

**Figure 1** presents a visual flow chart detailing the broad experimental approach described above, including all analyses presented within the remainder of this chapter.

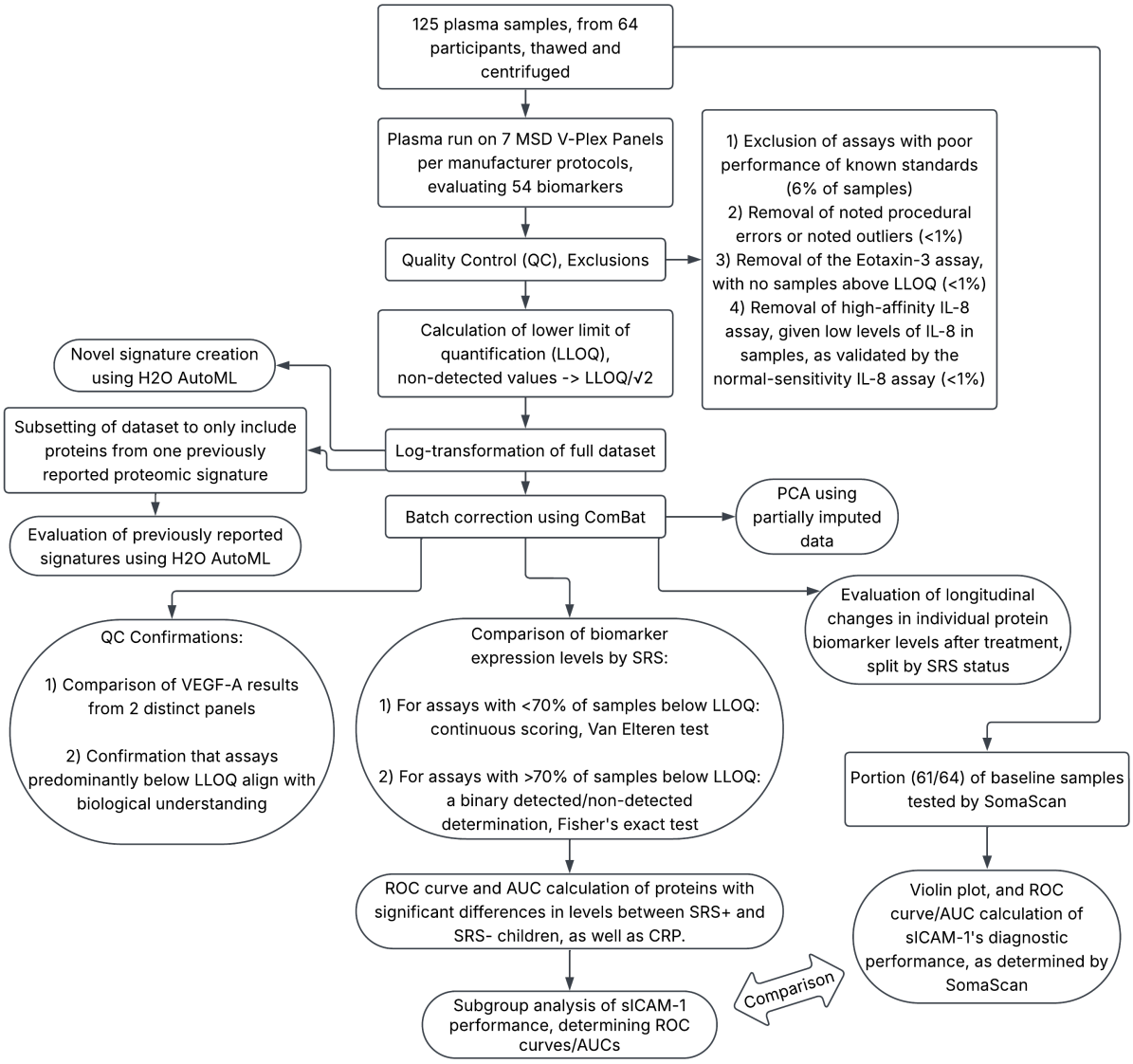



**Figure 1.** A flow chart detailing the experimental design of the project presented within this chapter. Ovals indicate results, while rectangles designate processes. Abbreviations: Lower Limit of Quantification (LLOQ), Meso Scale Discovery (MSD), Strict Reference Standard (SRS), Principal Component Analysis (PCA).

# Results

## 5.4 Performance of Meso Scale Assays in NOD-pedFEND Samples

*Sample Characteristics*

125 samples, 64 from baseline and 61 from the two-week visits of the same NOD-pedFEND participants, were run on the 7 Meso Scale panels. Participant characteristics are summarized in **Table 3**.

**Table 3.** Characteristics of NOD-pedFEND participants whose samples were utilized for this project. Three participants with unlikely TB did not have a two-week sample tested. For continuous variables, p values were calculated using the Kruskal-Wallis Rank Sum Test. For categorical variables, chi-squared tests with simulated p-values (from 2000 replicates) were used. Significance was calculated for the entire category across participant classifications; significance values, therefore, are the same for each percentage given in the same category. Interquartile ranges for median values are given in parentheses, with the 25th and 75th quartiles separated by a hyphen. Case classifications were determined as detailed in **Section 2.5**. Abbreviations: Lipoarabinomannan (LAM), Children Living with HIV (CLHIV), Severe Acute Malnutrition (SAM), Moderate Acute Malnutrition (MAM).

| | **Confirmed TB (n=28)** | **Unconfirmed TB (all LAM+ CLHIV) (n=4)** | **Unlikely TB (n=32)** | **P Value** |
|---|---|---|---|---|
| Age (in months) | 17.0 (13.0-36.5) | 15.0 (11.2-18.0) | 14.5 (11.8-36.0) | 0.55 |
| **Sex Distribution** | | | | 0.74 |
| % Female | 14/28 (50%) | 3/4 (75%) | 16/32 (50%) | |
| % Male | 14/28 (50%) | 1/4 (25%) | 16/32 (50%) | |
| **HIV Status** | | | | ≤0.001 |
| % HIV Positive | 4/28 (14%) | 4/4 (100%) | 0/32 (0%) | |
| % HIV Exposed-Uninfected | 4/28 (14%) | 0/4 (0%) | 6/32 (19%) | |
| % HIV Negative | 20/28 (71%) | 0/4 (0%) | 26/32 (81%) | |
| % HIV Status Unknown | 0/28 (0%) | 0/4 (0%) | 0/32 (0%) | |
| **Malnutrition Status** | | | | 0.41 |
| % with SAM | 9/28 (32%) | 3/4 (75%) | 15/32 (47%) | |

| | | | | |
|---|---|---|---|---|
| % with MAM | 4/28 (14%) | 1/4 (25%) | 7/32 (22%) | |
| % Without Malnutrition | 14/28 (50%) | 0/4 (0%) | 9/32 (28%) | |
| % Malnutrition Status Unknown | 1/28 (4%) | 0/4 (0%) | 1/32 (3%) | |
| **Disease Severity** | | | | ≤0.001 |
| % with Severe Disease | 13/28 (46%) | 4/4 (100%) | 3/32 (9%) | |
| % with Non-Severe Disease | 11/28 (39%) | 0/4 (0%) | 0/32 (0%) | |
| % with Unclassifiable Disease Severity | 4/28 (14%) | 0/4 (0%) | 29/32 (91%) | |

*Technical Validation of the MSD Assay*

Prior to evaluating potential biomarkers, several approaches were utilized for technical calibration and validation of the MSD assays. First, for standards above the calculated LLOQ in all non-excluded assays, the mean CVs and percent recoveries (from the standards' known concentration values) were calculated for each assay. These values are given in **Appendix 3**. In short, however, mean percent recovery values were between 98%-105%, and mean CVs were between 3%-16%.

As described in **Section 5.3**, a standard curve was fit to the data from the standards of known concentration which passed QC, enabling the recorded electrochemiluminescent signal to be related to target protein concentration. As an additional QC measure, the fit of this standard curve was evaluated for all assays, on both tested plates. All non-excluded assays had $R^2$ values of greater than 0.997, with an average $R^2$ value of 0.999. As 0.98 had been chosen as the cutoff QC threshold for an assay's $R^2$

value, no additional assays were excluded from subsequent analysis based upon the fit of their standard curve.

Given that VEGF-A was measured on two MSD panels, I also compared the VEGF-A concentrations determined by each panel for the same sample (**Figure 2**). By doing so, I sought to evaluate the precision of the VEGF-A concentration measurement along with, somewhat by proxy, the other assays on the "Cytokine Panel 1" and "Angiogenesis" panels. It was observed that, while the concentrations given by the two panels were well correlated with each other for all but two outlier samples ($R^2$ = 0.81), Cytokine Panel 1 tended to report slightly lower concentrations of VEGF-A than the Angiogenesis Panel, for the range of concentration values within which the tested samples fell. While these results provided some validation for the ordinal ranking of sample concentrations within a panel, these results do engender some uncertainty in the precise concentration values determined.

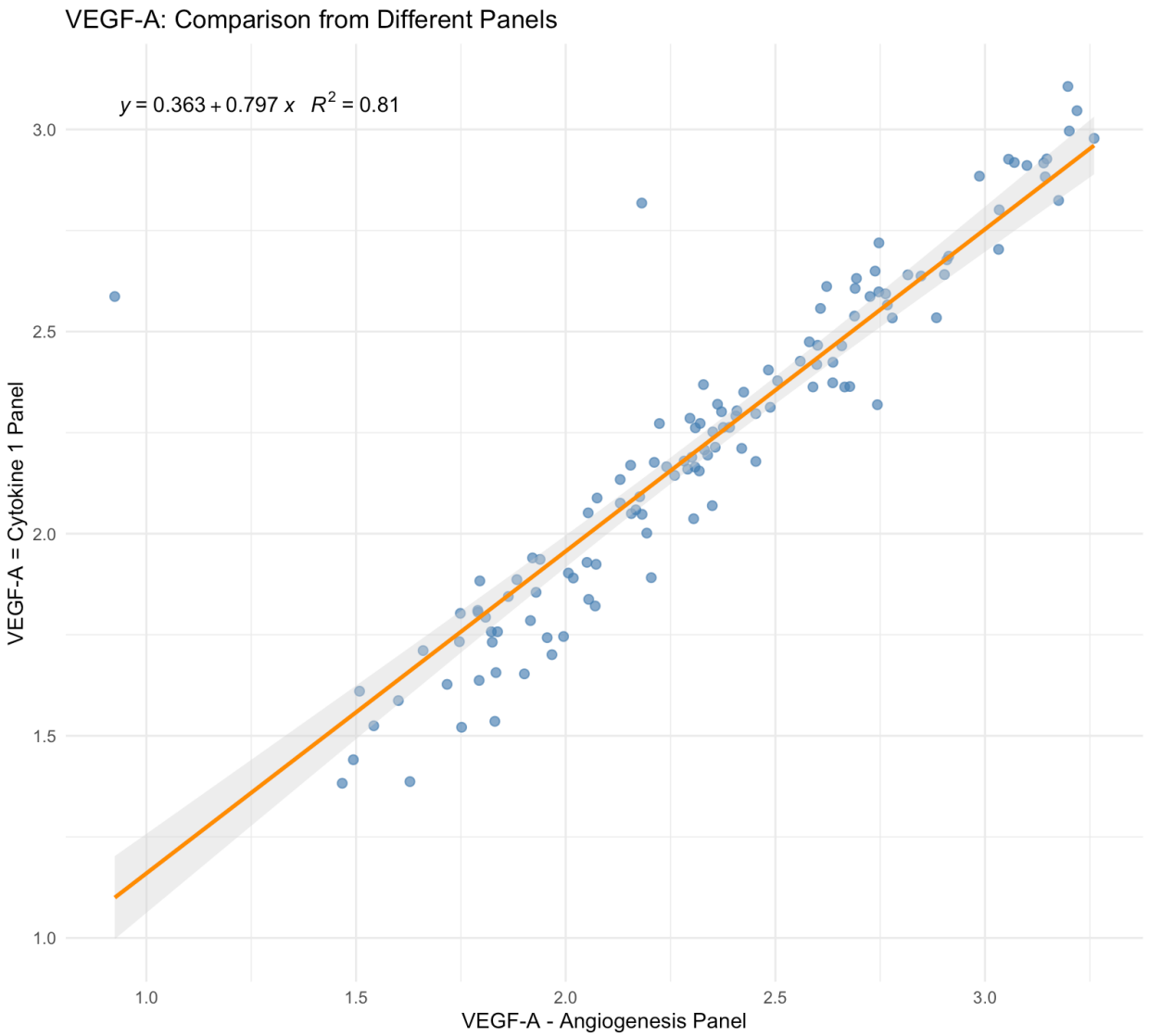



**Figure 2.** A scatter plot comparing the VEGF-A concentrations determined by the Angiogenesis and Cytokine 1 Panels for all samples tested.

Finally, I also evaluated the number of samples which fell below the LLOQ for each assay, with the LLOQ determined as described above. In short, the number of samples with analyte concentrations below the LLOQ (summarized in **Table 4**) varied dramatically by assay and aligned with the expected levels of that protein in the blood. For example,

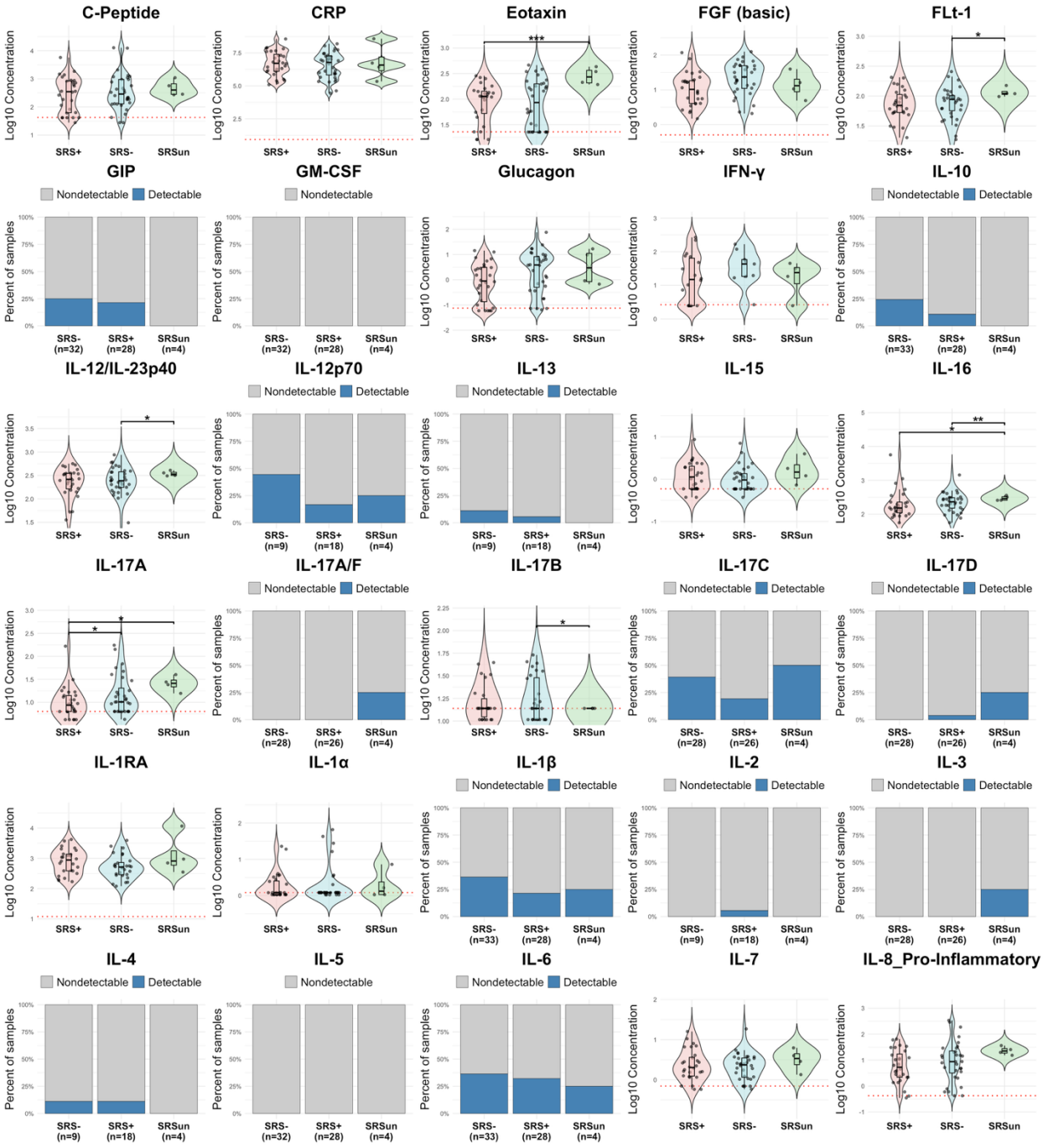
adaptive immune cytokines like IL-2, IL-4, and TNF-β were below quantifiable levels in at least 95% of measured samples, which aligns with the short half-life and paracrine nature of the proteins. In contrast, ubiquitous acute phase proteins like C-reactive protein (CRP) and SAA were detected in all samples. With this alignment to expected biology, I moved forward to evaluate differences in biomarker levels between participants of differing TB statuses.

**Table 4.** The number and proportion of samples (baseline and two-week) which fell below the LLOQ for each assay performed. Abbreviation: Lower Limit of Quantification (LLOQ).

| Assay | Number below LLOQ | Total Samples | Proportion below LLOQ |
|---|---|---|---|
| C-Peptide | 18 | 124 | 0.15 |
| CRP | 0 | 125 | 0.00 |
| Eotaxin | 24 | 125 | 0.19 |
| FGF (basic) | 0 | 125 | 0.00 |
| Flt-1 | 0 | 125 | 0.00 |
| GIP | 94 | 124 | 0.76 |
| GM-CSF | 122 | 125 | 0.98 |
| Glucagon | 25 | 124 | 0.20 |
| IFN-γ | 22 | 62 | 0.35 |
| IL-10 | 105 | 125 | 0.84 |
| IL-12/IL-23p40 | 0 | 125 | 0.00 |
| IL-12p70 | 46 | 62 | 0.74 |
| IL-13 | 57 | 62 | 0.92 |
| IL-15 | 48 | 125 | 0.38 |
| IL-16 | 0 | 125 | 0.00 |
| IL-17A | 44 | 125 | 0.35 |
| IL-17A/F | 111 | 113 | 0.98 |
| IL-17B | 71 | 113 | 0.63 |
| IL-17C | 83 | 113 | 0.73 |
| IL-17D | 107 | 113 | 0.95 |
| IL-1RA | 0 | 113 | 0.00 |
| IL-1α | 84 | 125 | 0.67 |
| IL-1β | 95 | 125 | 0.76 |
| IL-2 | 61 | 62 | 0.98 |
| IL-3 | 109 | 113 | 0.96 |
| IL-4 | 59 | 62 | 0.95 |
| IL-5 | 123 | 125 | 0.98 |
| IL-6 | 92 | 125 | 0.74 |
| IL-7 | 12 | 125 | 0.10 |
| IL-8 (Pro-Inflammatory Panel) | 5 | 125 | 0.04 |
| IL-9 | 73 | 113 | 0.65 |
| IP-10 | 27 | 125 | 0.22 |
| Insulin | 2 | 124 | 0.02 |
| Leptin | 41 | 124 | 0.33 |
| MCP-1 | 0 | 125 | 0.00 |
| MCP-4 | 7 | 125 | 0.06 |
| MDC | 0 | 125 | 0.00 |
| MIP-1α | 32 | 125 | 0.26 |
| MIP-1β | 19 | 62 | 0.31 |
| PIGF | 19 | 125 | 0.15 |
| PP | 0 | 124 | 0.00 |
| SAA | 0 | 125 | 0.00 |
| TARC | 7 | 125 | 0.06 |
| TNF-α | 16 | 62 | 0.26 |
| TNF-β | 119 | 125 | 0.95 |
| TSLP | 105 | 113 | 0.93 |
| Tie-2 | 1 | 125 | 0.01 |
| VEGF-A (Angiogenesis Panel) | 0 | 125 | 0.00 |
| VEGF-A (Cytokine Panel 1) | 0 | 125 | 0.00 |
| VEGF-C | 2 | 125 | 0.02 |
| VEGF-D | 1 | 125 | 0.01 |
| sICAM-1 | 0 | 125 | 0.00 |
| sVCAM-1 | 0 | 125 | 0.00 |

## 5.5 Differences in Individual Plasma Biomarkers in Children with TB

After validating the performance of the MSD assays, I evaluated the concentrations of each tested protein, in isolation, between NOD-pedFEND participants of different TB statuses, as determined by the strict reference standard (**Figure 3**).


C-Peptide
CRP
Eotaxin
FGF (basic)
FLt-1
GIP
GM-CSF
Glucagon
IFN-γ
IL-10
IL-12/IL-23p40
IL-12p70
IL-13
IL-15
IL-16
IL-17A
IL-17A/F
IL-17B
IL-17C
IL-17D
IL-1RA
IL-1α
IL-1β
IL-2
IL-3
IL-4
IL-5
IL-6
IL-7
IL-8_Pro-Inflammatory
Log10 Concentration
Percent of samples
Nondetectable
Detectable
SRS+
SRS-
SRSun
SRS- (n=32)
SRS+ (n=28)
SRSun (n=4)
SRS- (n=33)
SRS- (n=9)
SRS+ (n=18)
SRS- (n=28)
SRS+ (n=26)

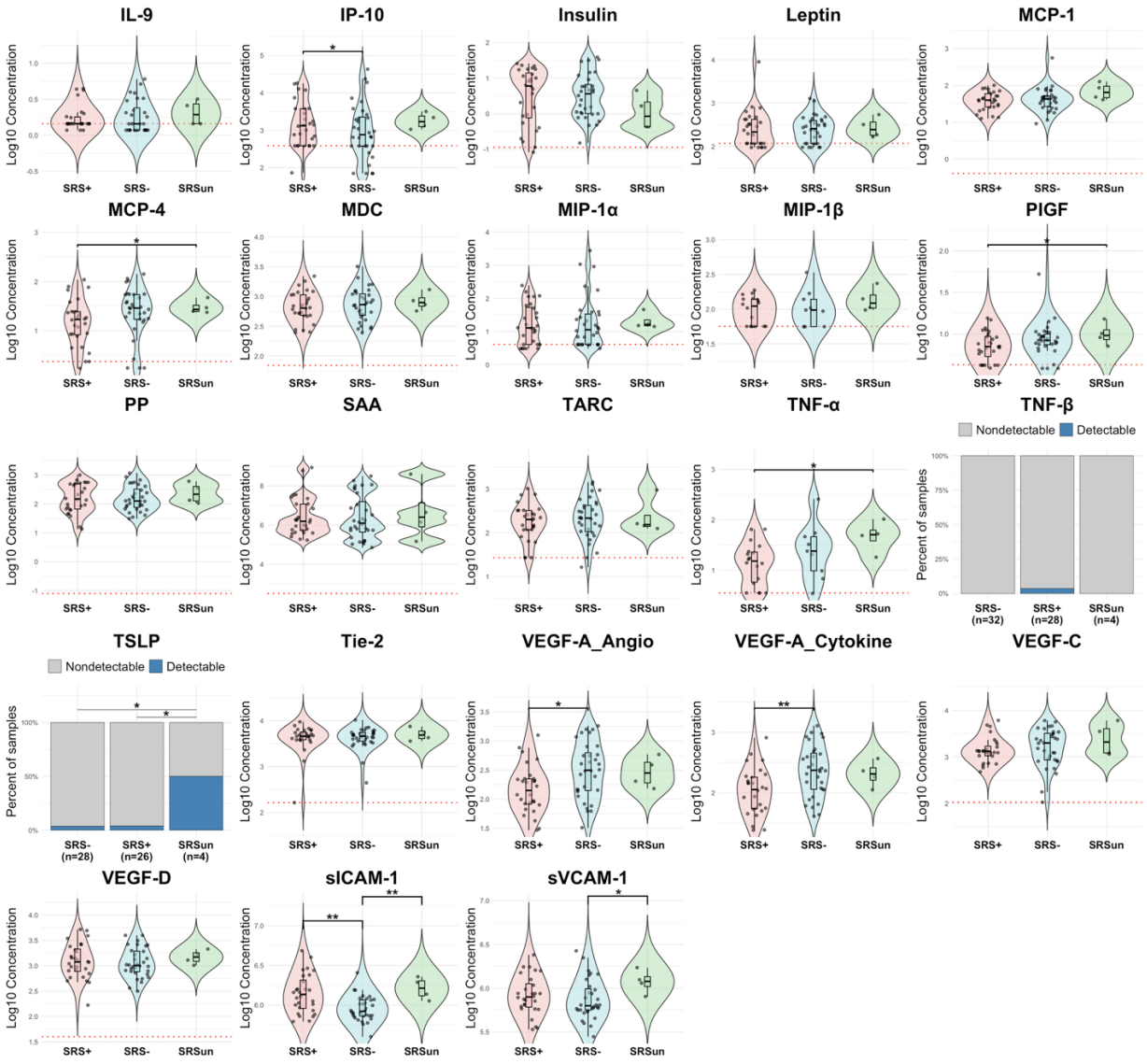



**Figure 3**. Plots demonstrating the measured expression levels for baseline samples across all biomarkers tested, differentiated by SRS. For assays with >70% of datapoints below the lower limit of quantification (LLOQ), a binary detection/no-detection determination was plotted, with significance determined using Fisher's exact test. For assays with <70% of datapoints below the LLOQ, significant differences in the measured concentrations between TB statuses were determined using a Van Elteren test set to control for the two distinct plates samples could be run on, utilizing a Monte-Carlo approximation by resampling with 10,000 iterations. Concentrations are given in log-transformed values of pg/mL. The LLOQ for each assay is denoted with a dotted red line. If not present, concentrations for all plotted samples lie above the LLOQ. Reference standards were defined as outlined in **Section 2.6**. Abbreviations: Strict Reference Standard (SRS), Unclassifiable (Un).

Levels of IL-17A, IP-10, VEGF-A (as determined by both panels), and sICAM-1 were significantly different between SRS+ and SRS- participants. Seeing this, I evaluated the diagnostic performance of each protein, using ROC curves based on the SRS (**Figure 4**). The protein sICAM-1 performed best in differentiating between SRS+ and SRS- participants, with an AUC of 0.74. VEGF-A performed similarly, albeit slightly less well, with AUCs of 0.69 (when measured by the angiogenesis panel) and 0.72 (when measured by the cytokine panel). And while IL-17A performed worse, with an AUC of 0.64, IP-10 was least predictive, with an AUC of 0.60 against the SRS. Given that sICAM-1 performed best, I also evaluated it against the eMRS, where it had an AUC of 0.76.

From the eMRS-based analysis, it was apparent that sICAM-1 levels, if used as a TB diagnostic, would have identified 5/32 (16%) of eMRS+ individuals using a cutoff (6.43) that gave perfect specificity, or 17/32 (53%) of eMRS+ individuals using a cutoff (6.13) determined by the Youden's index that yielded 91% specificity (misidentifying only three eMRS- individuals). Of note, the Youden's index also identified an alternative cutoff, 6.09, that detected 19/32 (59%) of eMRS+ individuals with a specificity of 84%.

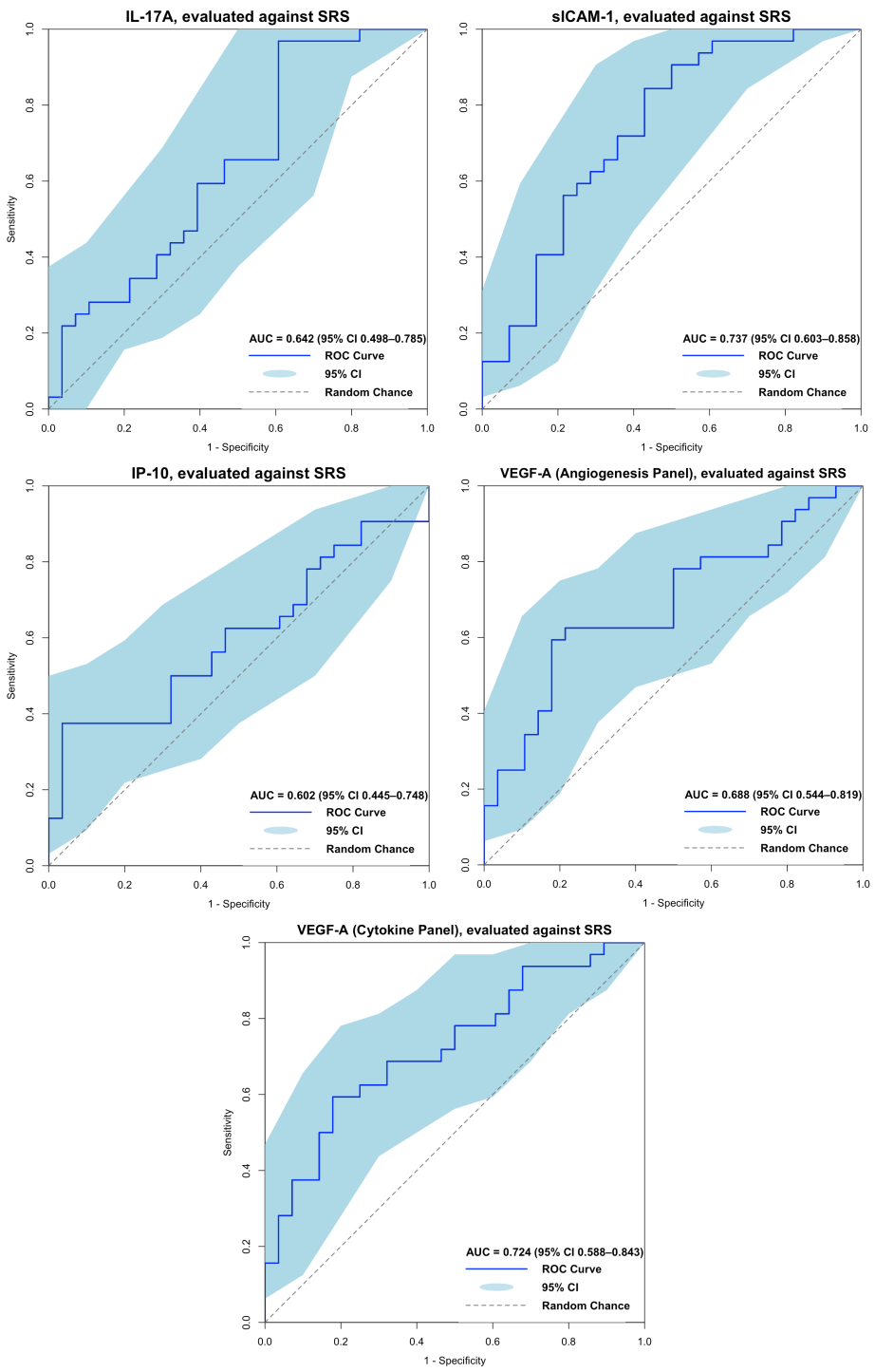



**Figure 4.** ROC curves for the four proteins (with VEGF-A measured twice) which were present at significantly different concentrations between Strict Reference Standard (SRS)+ and SRS- participants. ROC curves were also constructed using the SRS, with a 95% confidence interval determined by bootstrapping with 2000 iterations. Reference standards were defined as outlined in **Section 2.6**.

CRP has also been suggested as a potential diagnostic biomarker for TB [155]. To evaluate its performance in the NOD-pedFEND cohort, I constructed a ROC curve using CRP concentrations and the SRS (**Figure 5**). CRP exhibited worse diagnostic performance than any of sICAM-1, VEGF-A (by both assays), IL-17A, or IP-10, with an AUC of 0.46 (95% CI 0.32-0.61) when evaluated against the SRS. Interestingly, however, while 16/28 (57%) of SRS+ children had CRP values elevated above the normal range (<3mg/L), an additional 17/32 (53%) of SRS- children also had elevated CRP values. As such, it appears that the poor diagnostic performance of CRP in our cohort can be primarily attributed to elevated CRP levels in SRS- children, rather than the lack of elevation in SRS+ children.

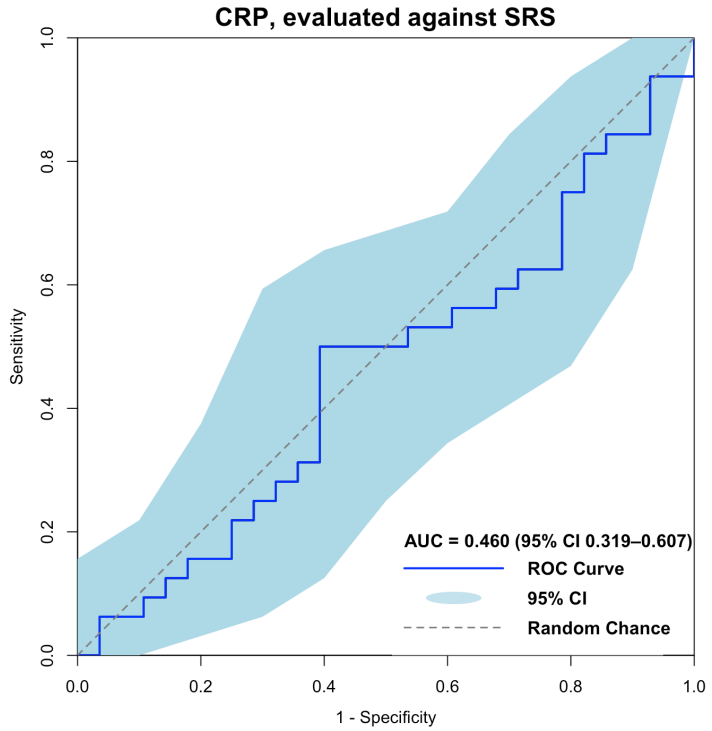



**Figure 5**. A ROC curve demonstrating the ability of log-transformed CRP concentrations to differentiate between Strict Reference Standard (SRS)+ and SRS- children. The 95% confidence interval was determined using bootstrapping with 2000 iterations. Reference standards were defined as outlined in **Section 2.6**.

## 5.6 Diagnostic Accuracy of an Ensemble Model Composed of All Tested Proteins

Given the modest diagnostic performance of individual biomarkers, particularly VEGF-A and sICAM-1, I sought to determine if diagnostic power could be enhanced by using combinations of biomarkers. To explore this, I trained 50 machine learning models of various architectures to predict SRS classifications based upon the log-transformed data from all 52 measured proteins—including VEGF-A and sICAM-1—using the H2O package in R, as I also did with our NanoString data. Datasets used for all machine learning tasks were not batch-corrected to prevent data leakage. Models were ranked by their AUC, which was determined through six-fold cross-validation. The performance of the top three models is shown in **Table 5**. The top-performing model—a Gradient Boosting Machine (GBM)—demonstrated an AUC of 0.80, which was not substantially different than the performance of sICAM-1 alone. Further, as demonstrated in **Figure 6**, the most important variable in the top-performing models was sICAM-1 or VEGF-A concentration. The limited value of additional biomarkers was also demonstrated by a PCA plot derived from all assay data from baseline samples (**Figure 7**), which did not demonstrate clear differentiation between SRS+ and SRS- children by visual inspection.

**Table 5.** Performance of the top three best-performing machine learning models, as determined by AUC, out of the 50 models fit to our dataset and the strict reference standard. Reference standards were defined as outlined in **Section 2.6**. Abbreviation: Gradient Boosting Machine (GBM).

| Model ID | AUC | Logarithmic Loss | Area Under the Precision-Recall Curve | Mean Per Class Error | Root Mean Squared Error | Mean Square Error |
|---|---|---|---|---|---|---|
| *GBM Model #1* | 0.80 | 0.55 | 0.77 | 0.25 | 0.43 | 0.18 |
| *GBM Model #2* | 0.78 | 0.59 | 0.73 | 0.31 | 0.45 | 0.20 |
| *GBM Model #3* | 0.76 | 0.58 | 0.76 | 0.35 | 0.45 | 0.20 |

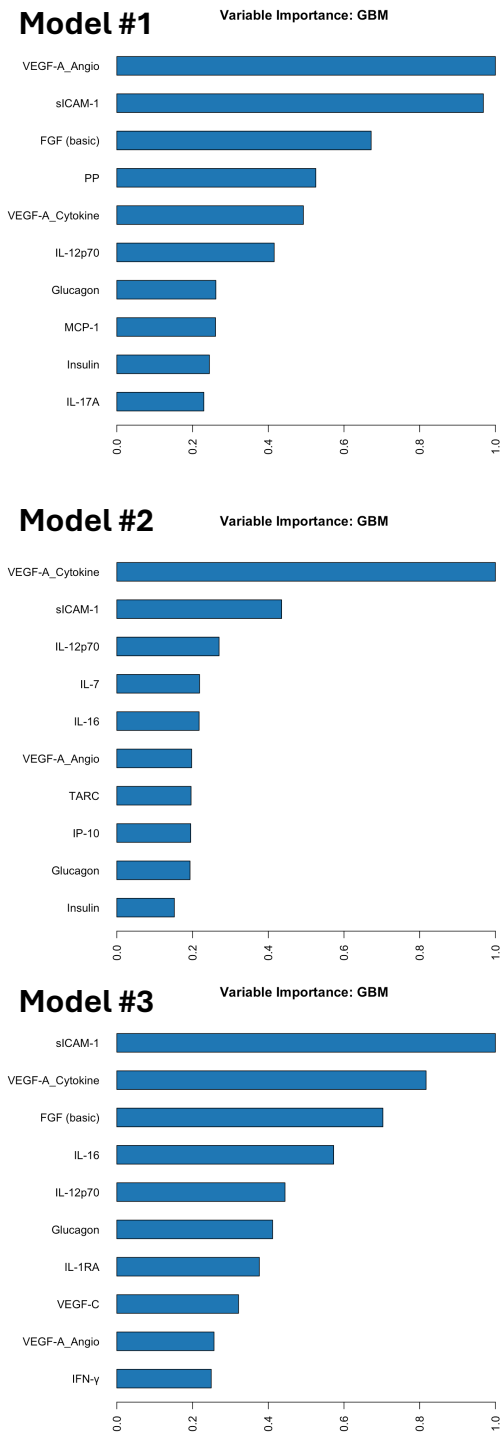



**Figure 6.** The variable importances of the top 10 most important proteins to the top models derived from baseline plasma protein levels from NOD-pedFEND participants. Variable importance is normalized to one for the protein of greatest importance in each model. Models were developed using the strict reference standard, which was defined as outlined in **Section 2.6**.

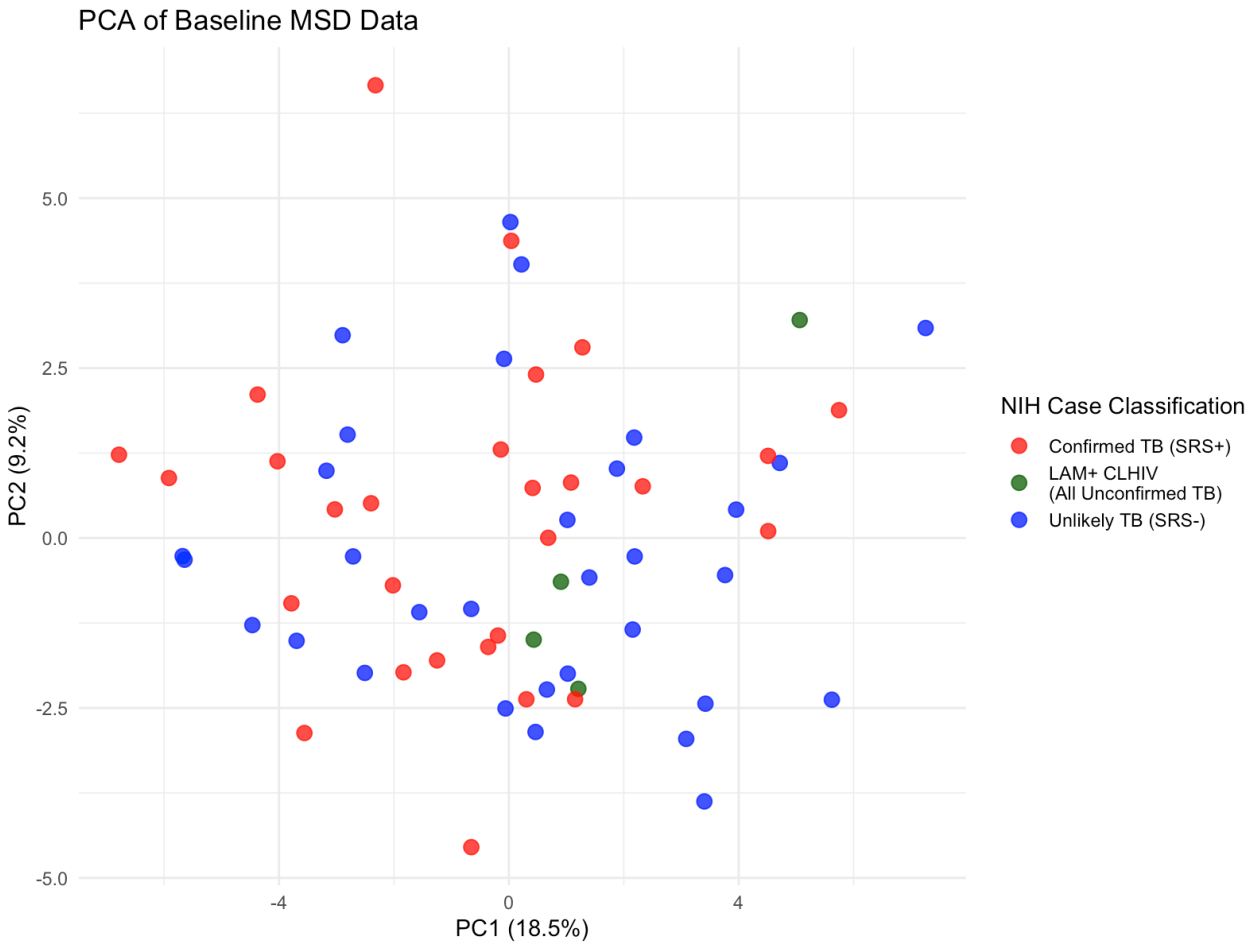



**Figure 7**. A PCA plot of all baseline samples, with samples from individual participants represented as points coloured by their case classification. Given PCA's need for a complete dataset, any data for the 52 analysed proteins which had been excluded for a given sample was imputed using 10 k-nearest neighbours. Classifications were defined as outlined in **Section 2.5** and **Section 2.6**. Abbreviations: Strict Reference Standard (SRS), Lipoarabinomannan (LAM), Children Living with HIV (CLHIV), Meso Scale Discovery (MSD).

Given these results, along with the simplicity of a single-protein biomarker, the minimally increased performance of a multi-protein signature of disease, and the potential for overfitting given the lack of a dedicated test set used for model evaluation, I elected not to move forward with further interrogation of any of the derived machine learning models.

## 5.7 Diagnostic Performance of Previously Reported Proteomic Signatures of TB

I did, however, attempt to independently validate previously reported signatures of TB by deriving machine learning models based solely upon the proteins used in the reported signatures. Unfortunately, precise replication of all reported signatures was not possible due to inconsistently reported model structures and parameters. As a result, I instead opted to evaluate signatures via a uniform approach: using H2O's AutoML function to fit a variety of machine learning models to a subset of our dataset which only included data from proteins in the reported signature. By fitting a model directly to our data in this manner, parameters and model structure were optimized for performance in our cohort, providing a best-case, likely over-estimate, of diagnostic performance. By comparing these optimistic estimates to those of my own derived signatures, comparative performance under optimal parameter tuning could be assessed.

After training 50 models to the SRS using a data subset which included only data from the proteins in the reported signatures, the resultant models were ranked by their AUC, as measured during six-fold cross-validation. The AUC of the top-ranked model, for each signature, was recorded (**Table 6)**. Samples missing data for one or more of the proteins in the signature were not included in the dataset used for model fitting or evaluation.

**Table 6.** Diagnostic performance of the best of 50 models trained only on data from the proteins in a previously reported paediatric signature of TB. Signatures were evaluated against the strict reference standard (SRS), which was defined as detailed in **Section 2.6**. Abbreviation: Gradient Boosting Machine (GBM).

| **Signature** | **Composed of** | **Best-fit model, AUC** |
|---|---|---|
| *Kumar 2021 [138]* | TNF-α, IL-17A, IL-2 | GBM, 0.67 |
| *Tebruegge 2015 [140]* | TNF-α, IL-1RA, IL-10 | GBM, 0.68 |
| *Gourgouillon 2012 [142]* | TNF-α, IL-2 | GBM, 0.64 |
| *López-Suárez 2024 [143]* | IFN-γ, IL-2, MCP-1 | GBM, 0.67 |
| *Togun 2020 [145]* | IL-1RA, IL-7, IP-10 | GBM, 0.67 |
| *Sudbury 2019 [146]* | IL-1RA, TNF-α | GBM, 0.70 |

While a better-performing model might be identified if additional models were trained beyond these first fifty, the calculated AUCs for the best-performing model out of those trained were considered a fit-for-purpose approximation of signature performance. As no signature exhibited diagnostic performance which exceeded that of sICAM-1 (which had an AUC of 0.74 against the SRS) or my derived ensemble model (which had an AUC of 0.80 against the SRS), they were not investigated further.

## 5.8 Subgroup-Specific Performance of Proteomic Signatures

With these overall results, I opted to focus on interrogating the diagnostic performance of the marker which best balanced simplicity with diagnostic performance:

sICAM-1 alone. To further investigate its utility, I assessed whether its performance as a diagnostic marker differed in various subgroups of interest (**Table 7**).

**Table 7**. Performance of sICAM-1 in differentiating between SRS+ and SRS- NOD-pedFEND participants, split by subgroups. Confidence intervals were determined using bootstrapping with 2000 replicates. Reference standards were defined as detailed in **Section 2.6**. Abbreviations: Strict Reference Standard (SRS), Severe Acute Malnutrition (SAM), Moderate Acute Malnutrition (MAM).

| **Subgroup** | **Number of SRS Positive** | **Number of SRS Negative** | **sICAM-1 AUC (95% CI)** |
|---|---|---|---|
| *Overall* | 28 | 32 | 0.74 (0.60-0.86) |
| *HIV Uninfected, Unexposed* | 20 | 26 | 0.73 (0.57-0.86) |
| *≥ two years old* | 12 | 10 | 0.68 (0.43-0.88) |
| *< two years old* | 16 | 22 | 0.77 (0.59-0.91) |
| *Malnourished (SAM or MAM)* | 13 | 22 | 0.80 (0.64-0.93) |
| *Severe TB* | 13 | 3 | 0.92 (0.74-1.00) |
| *Female* | 14 | 16 | 0.73 (0.52-0.90) |

Performance appeared to be best in children who were malnourished or who met criteria for severe TB, with AUCs of 0.80 and 0.92, respectively. Given the small sample size of my exploratory analysis, however, confidence intervals for all subgroups were wide and overlapped with one another. Additionally, I was unable to control for other subgroups of which a child might be a member. As such, participant overlap between subgroups could meaningfully influence these results, making it difficult to determine which participant characteristics are the primary driver behind any changes in biomarker performance observed. While I was unable to directly account for subgroup overlap in my analysis, to

contextualize the reported AUCs, UpSet charts were constructed to demonstrate the overlap in participants' subgroup membership (**Figure 8**).

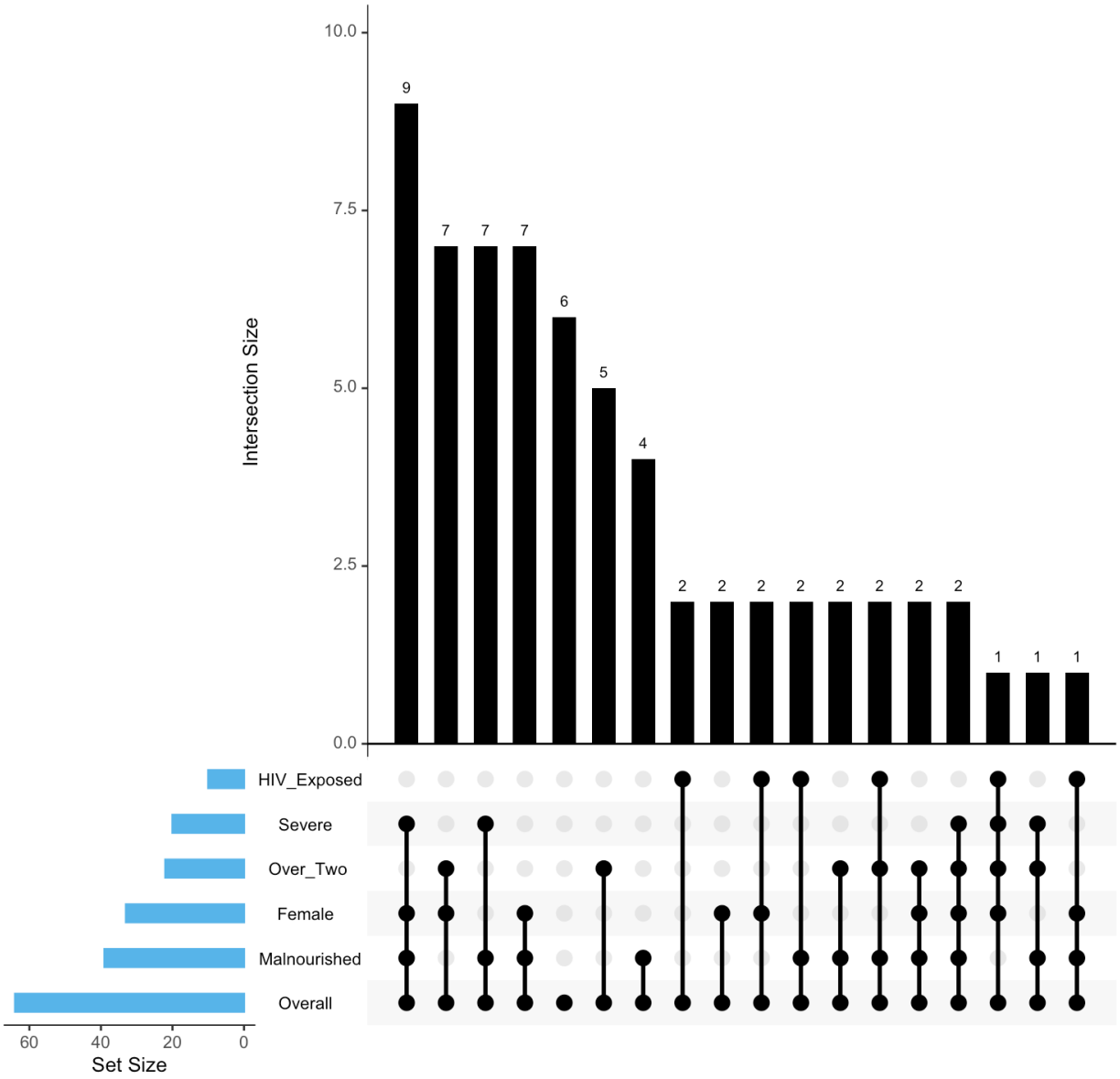
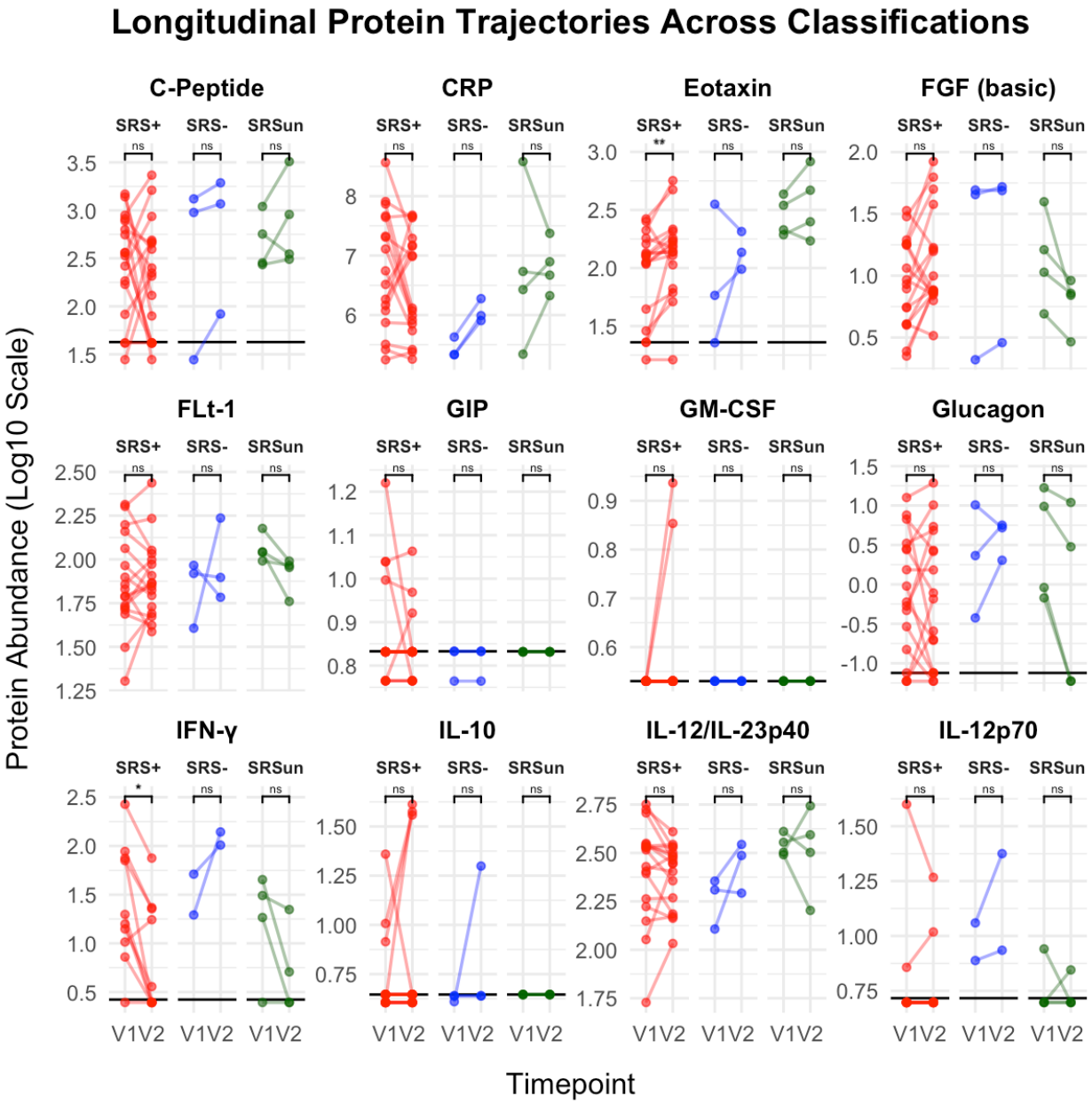
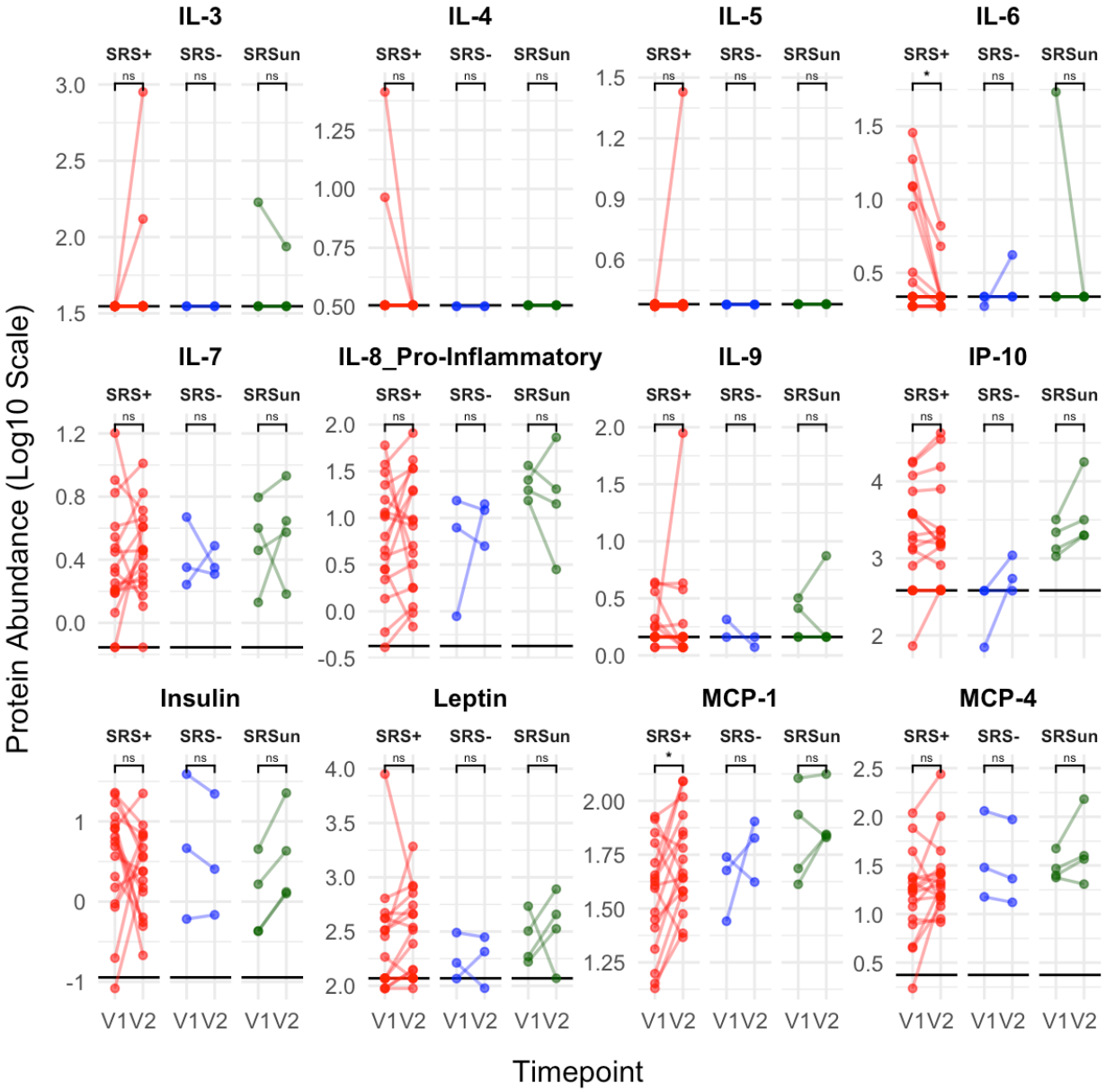


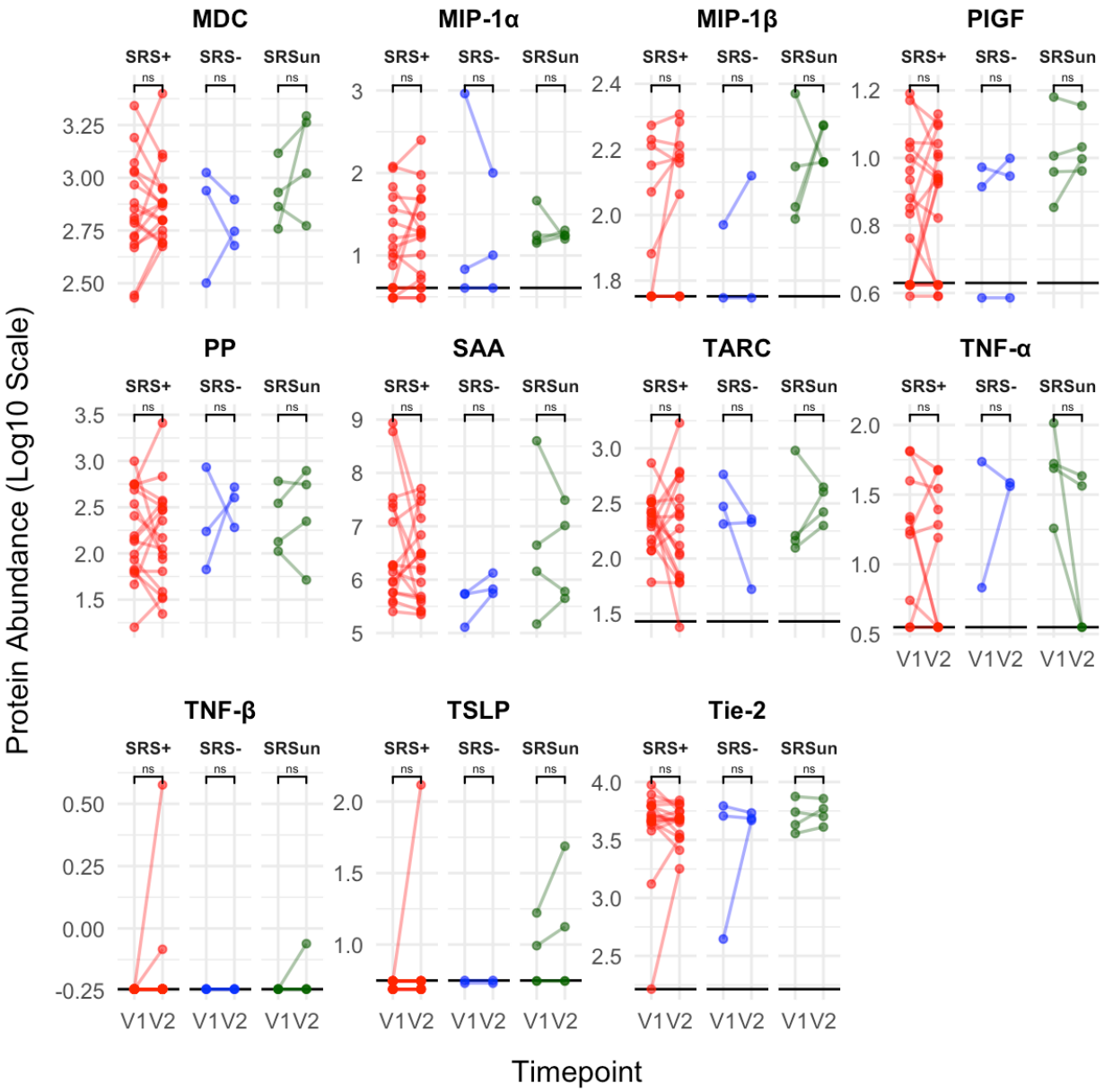

**Figure 8.** UpSet plots demonstrating participant overlaps between multiple evaluated subgroups. The "overall" group includes all participants analysed by MSD. The "malnourished" category includes children with both severe and moderate acute malnutrition.

## 5.9 The Use of Longitudinal Changes in Biomarker Levels to Assist in Retrospective TB Diagnosis

I also evaluated how levels of sICAM-1, along with the other 51 proteins measured, changed after two weeks of anti-TB treatment. For this, I calculated the change in measured protein levels for 24 participants who did not have DR-TB and who were given anti-TB treatment at baseline. These 24 participants consisted of 17 with confirmed TB, 4 with unconfirmed TB (who were also LAM+ CLHIV) and 3 with unlikely TB. Of the children with confirmed TB, 16/17 (94%) exited the study successfully, with one child dying prior to study conclusion, and 13/17 (76%) exhibiting successful treatment response by the end-of-study visit. While sICAM-1 and CRP levels did not significantly change between samples collected at baseline and two-weeks after treatment initiation, four proteins changed significantly ($p<0.05$): Eotaxin ($p<0.01$) and MCP-1 (which increased), and IL-6 and IFN-γ (which decreased). When correcting for multiple comparisons using a Benjamini-Hochberg correction, however, none remained statistically significant. That being said, IL-6 demonstrated a universal drop in concentration for the SRS+ participants who had baseline concentrations above the LLOQ (**Figure 9**). Of proteins with significant changes between the baseline and two-week visits, the following numbers of SRS+ participants were above the LLOQ at baseline: IL-6 (7/17 participants), IFN-γ (9/10 participants), Eotaxin (16/17 participants), and MCP-1 (17/17 participants).


Longitudinal Protein Trajectories Across Classifications
C-Peptide
CRP
Eotaxin
FGF (basic)
FLt-1
GIP
GM-CSF
Glucagon
IFN-γ
IL-10
IL-12/IL-23p40
IL-12p70
SRS+
SRS-
SRSun
ns
*
**
V1V2
Protein Abundance (Log10 Scale)
Timepoint


IL-3
IL-4
IL-5
IL-6
IL-7
IL-8_Pro-Inflammatory
IL-9
IP-10
Insulin
Leptin
MCP-1
MCP-4
SRS+
SRS-
SRSun
ns
*
Protein Abundance (Log10 Scale)
V1V2
Timepoint


MDC
MIP-1α
MIP-1β
PIGF
PP
SAA
TARC
TNF-α
TNF-β
TSLP
Tie-2
SRS+
SRS-
SRSun
ns
V1V2
Protein Abundance (Log10 Scale)
Timepoint

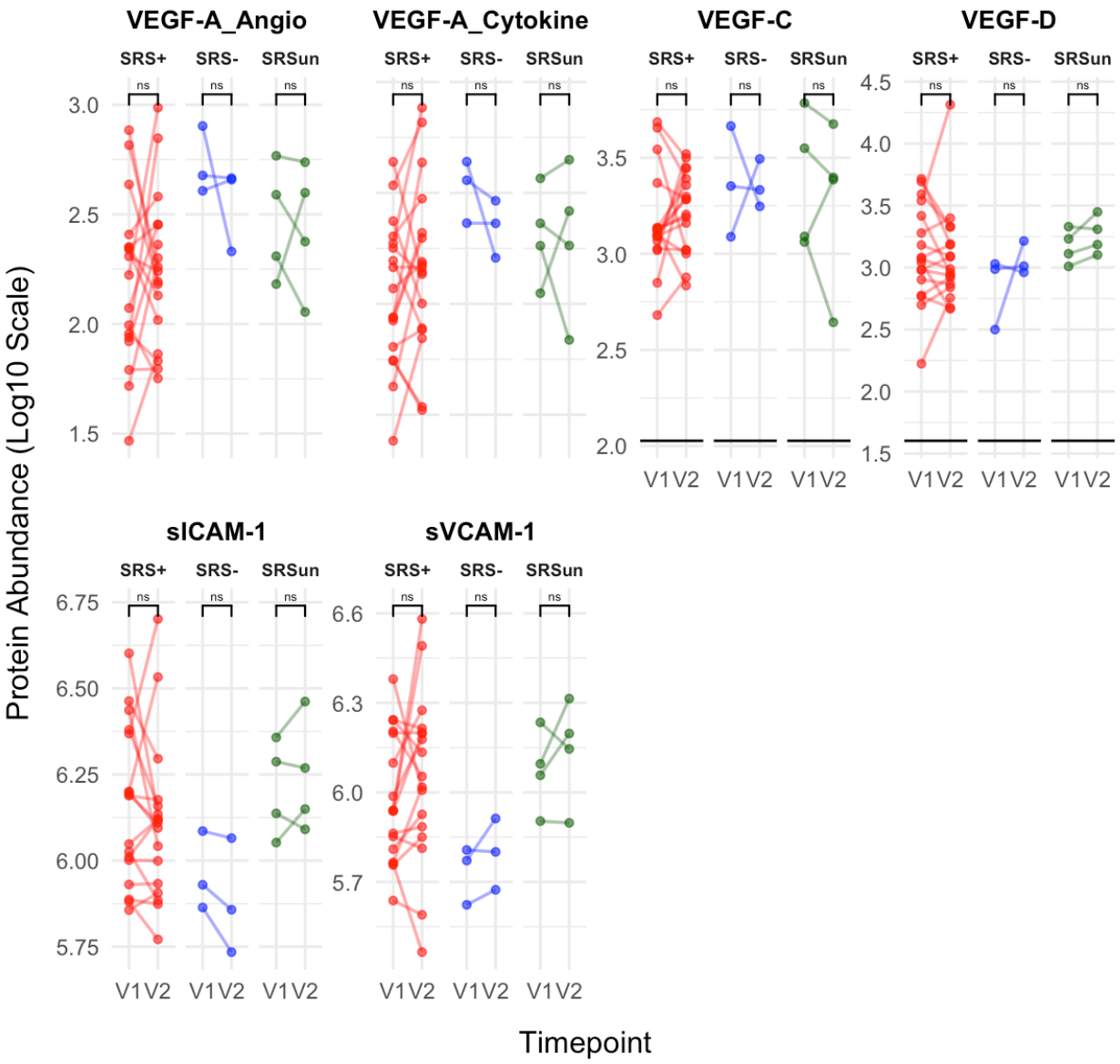



**Figure 9.** Changes in measured protein levels after two weeks for each participant, grouped by strict reference standard (SRS) classification. Significance was calculated using the Wilcoxon signed-rank test and is shown prior to the correction for multiple hypotheses being applied. A black horizontal line demarcates the lower limit of quantification (LLOQ). If not displayed for a particular protein, it lies below the displayed Y axis. V1 signifies the baseline visit, while V2 signifies the two-week visit. Reference standards were defined as outlined in **Section 2.6**.

## 5.10 Confirmation of sICAM-1 Performance using SomaScan Data

All baseline samples from SRS-classifiable participants who had been selected for the MSD analysis (60/64, 94%) were also sent for proteome-wide analysis using the SomaScan platform. While the full analysis of the SomaScan results remains preliminary and lies outside the scope of this thesis, the resultant dataset presents an opportunity to evaluate whether the positive results of sICAM-1 found with the MSD V-Plex platform are replicated on measurements taken using the SomaScan platform. As such, I performed a targeted analysis of only the sICAM-1 levels measured by SomaScan. In so doing, I found results to align with those reported by MSD. Levels of sICAM-1 were also found to be significantly elevated in SRS+ participants in the SomaScan data (**Figure 10**), which yielded an AUC of 0.75 (95% CI 0.62-0.87) against the SRS for sICAM-1 (**Figure 11**). While these results were obtained from the same cohort, validation in a second, independent technical platform supports the previously described MSD results for the protein.

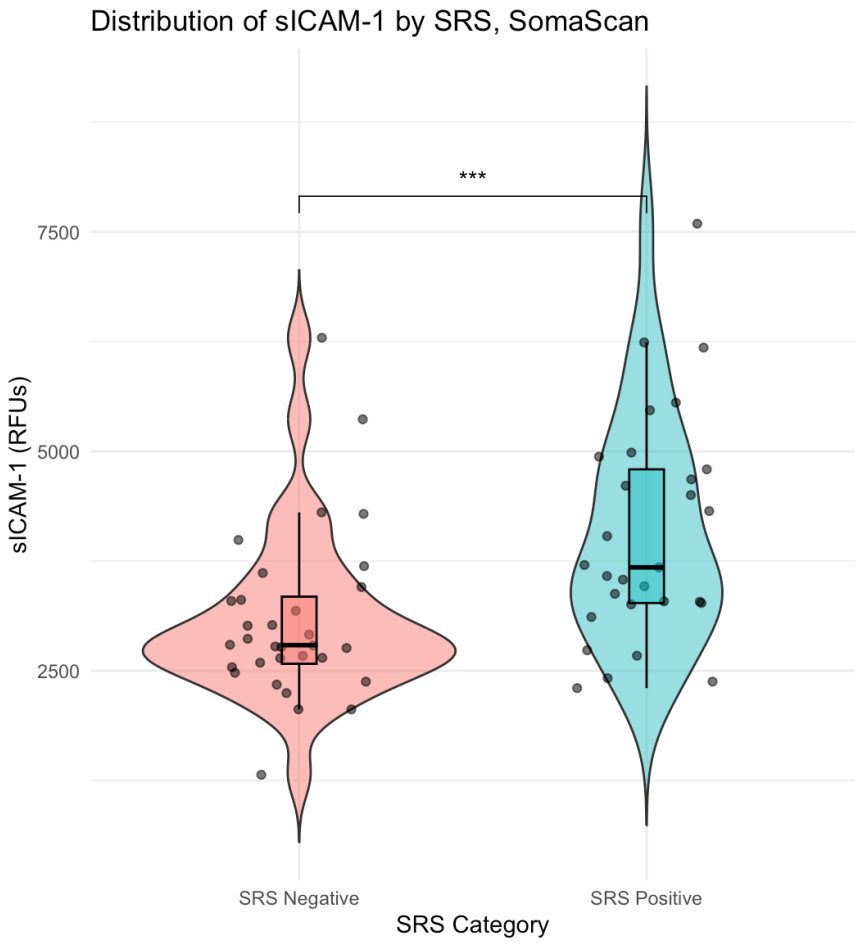



**Figure 10.** A violin plot showing the distribution of sICAM-1 levels in SRS+ and SRS- individuals, as measured by SomaScan. sICAM-1 levels are presented as relative fluorescent units (RFUs), which are directly proportional to protein concentration. Significance was determined using a Wilcoxon Rank-Sum Test. Reference standards were defined as outlined in **Section 2.6**.

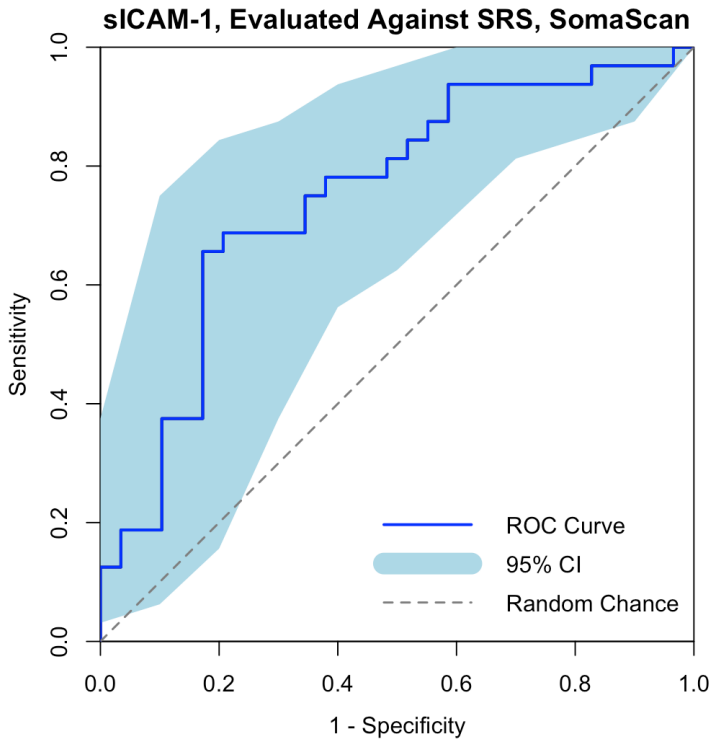



**Figure 11.** A ROC curve showing the ability of sICAM-1 levels, as measured by SomaScan in RFU units, to differentiate between strict reference standard (SRS)+ and SRS- NOD-pedFEND participants. Confidence intervals were determined using bootstrapping with 2000 iterations. Reference standards were defined as outlined in **Section 2.6**.

## 5.11 Discussion

Using samples collected during a non-interventional TB diagnostic trial, I measured and analysed the levels of 52 protein biomarkers to search for diagnostic signatures of TB disease in the blood plasma of children with presumptive TB. In so doing, I identified a single protein—sICAM-1—which could differentiate between symptomatic children with and without TB with an AUC of between 0.74 and 0.76, depending on the technology and reference standard used.

*sICAM-1*

That a single protein not originating from the pathogen exhibited this modest, but meaningful, performance is significant, particularly given that this performance was exhibited in children under five, for whom TB diagnosis is notoriously difficult. For context, the best-performing established test of TB performed in the NOD-pedFEND cohort—Xpert Ultra of a gastric aspirate sample—only identified 35% of eMRS+ children on its own [156], which is in line with data reported by similar studies [157]. sICAM-1, in comparison, had a sensitivity of 16% when set to a specificity of 100% in detecting eMRS+ individuals, or 53% sensitivity and 91% specificity using a lower cutoff determined by the Youden's index. sICAM-1, then, performed at a level approaching that of established TB tests like Xpert Ultra, despite utilizing a simple blood sample. It should, however, be acknowledged that these results utilize an optimal cutoff for sICAM-1 levels derived from only the NOD-pedFEND cohort, with knowledge of participants' TB statuses and which cutoff would maximize diagnostic performance. Nevertheless, the performance of sICAM-1 in paediatric TB diagnosis appears promising.

Despite numerous studies having developed protein-based signatures of paediatric TB, it is interesting that sICAM-1 has rarely been mentioned in this context. This may in part be a reflection of sICAM-1 not being included in the more widely used Luminex panels, which has been the technology utilized by most paediatric TB studies. In fact, none of the reported studies in **Table 1** utilized Luminex panels capable of measuring sICAM-1 levels.

In contrast, the only published paediatric TB research, to my knowledge at the time of writing, which utilized a nontargeted approach—that by Fossati et al. [136]—did detect

differential levels of sICAM-1 in their cases versus controls, though sICAM-1 wasn't utilized for any of their derived signatures. While no other paediatric studies of pulmonary TB, to my knowledge, have identified sICAM-1 as a diagnostic biomarker, it has, however, been reported to be raised in circulating blood in adult patients [158-163], and in the CSF of children with TB meningitis [164]. Some studies in adults do disagree, however. While sICAM-1 was suggested for use as a sarcoidosis diagnostic in 2015 by Du et al., for TB, it was not significantly elevated, with p=0.19 [165]. In other cohorts, sICAM-1 has also been reported as a marker of miliary and advanced pulmonary tuberculosis but not in milder stages of disease [166, 167]. Other groups have proposed to use sICAM-1 levels in blood serum for treatment monitoring, as they observed the protein's concentration to change after time on anti-TB treatment [158, 168]; however, one of these studies, by Bai et al., reported sICAM-1 to be lower in patients with pulmonary TB and to increase with treatment, which is contrary to the opposite trend and relative concentration reported by others.

*Biological Significance of (s)ICAM-1*

ICAM-1, in its membrane-bound form, has also been implicated in TB mechanistically, with the protein thought to be exploited by *Mtb* to support the bacterium's invasion into host macrophages [169]; further, monocytes have been found to upregulate ICAM-1 in the presence of *Mtb* [170]. Levels of CD14+ ICAM-1+ cells have been reported to decrease in tuberculosis, however, though membrane-bound ICAM-1 is notably distinct from the soluble form of the glycoprotein measured in this analysis [171]. Expression of ICAM-1 on γδ T cells and monocytes has also been shown to differentiate between

individuals with TB with or without concurrent underlying disease, such as type 2 diabetes or cancer [163].

Independent of TB, ICAM-1, also known as CD54, is a glycoprotein which functions as an adhesion receptor and is particularly implicated in driving leukocyte recruitment [172]. In addition to the mechanisms previously mentioned, ICAM-1 has been implicated in the adhesion between CD4+ T cells and B cells and the subsequent regulation of B cell development and activation, particularly with regards to IgG production [173]. Other immunological synapses, such as those between dendritic cells and CD8+ T cells, can also be aided by ICAM-1 [172, 174]. Relatedly, ICAM-1 has also been shown to play a role in outside-in signalling and therefore the transcriptional landscape of the cells on which it is expressed [172].

sICAM-1 is the soluble form of ICAM-1 which lacks its cytoplasmic domain and is typically formed from alternative splicing or proteolytic cleavage of ICAM-1 [175-177]. Like ICAM-1, it is a marker of inflammation, and it has been proposed as a biomarker for a number of conditions, including venous thromboembolisms, sepsis, and some cancers [178-180]. Activation of alveolar macrophages by sICAM-1 has also been observed in a murine model [181]. In sum, elevated levels of sICAM-1 in children with TB are consistent with the literature and the expected pathology of the disease, further supporting its potential use as a diagnostic marker in this population.

*Longitudinal and Subgroup-Specific Performance of sICAM-1*

Unlike other studies, I did not observe a significant decrease in sICAM-1 levels in SRS+ individuals treated for TB. However, while most prior reports evaluated changes in

sICAM-1 levels after multiple months of TB treatment, the samples utilized for this project were collected only two weeks after TB treatment initiation. It is possible that this was simply insufficient time on treatment for a significant difference in sICAM-1 levels to be observed; further study at later timepoints would help to inform whether sICAM-1 could also be used as a treatment monitoring tool in young children.

Interestingly, I found the diagnostic performance of sICAM-1 to improve in children with malnutrition or severe disease, although the sample size for the latter evaluation was especially small, with only three SRS- cases able to be included. This observation does, however, align with previous reports in adults that levels of sICAM-1 were most elevated in severe and miliary forms of TB disease [166, 167]. As sICAM-1 has also been proposed as a marker of acute lung injury [182], however, it is possible that the underlying biology being detected as sICAM-1 elevation is primarily lung injury and not specific sequelae of TB, which would make it unlikely that sICAM-1 levels would be sufficient to diagnose TB. Analysing sICAM-1 levels in a greater number of children with presumptive severe TB disease who are eventually determined not to have TB would be helpful for elucidating the specificity of sICAM-1 elevation for TB diagnosis amongst children with presumptive severe TB disease. Validations in independent cohorts—particularly those with a range of other inflammatory diseases —would be especially useful for determining the generalizability and use cases for sICAM-1 as a TB diagnostic.

*Diagnostic Performance of CRP*

CRP has also been proposed as an alternative blood-based biomarker for TB [183]; however, in our cohort, it performed worse than sICAM-1, with an AUC of 0.46 for detecting

SRS+ individuals. Like sICAM-1, CRP also did not show a significant decrease in measured concentrations after two weeks of anti-TB treatment, though I did not evaluate levels later in the treatment course. That CRP, a known marker of inflammation, did not perform significantly better than a random classifier in identifying TB is interesting, but may be due to the low specificity commonly reported for the test [184]. This evaluation compared symptomatic individuals with and without an eventual TB diagnosis, meaning that all children in both the case and control categories likely had some underlying illness. As CRP is elevated in many other conditions [185], it is possible that elevated CRP levels in both children with and without TB prevented the protein from functioning as an effective TB differentiator. This hypothesis was supported by our data showing normal baseline CRP levels in only 47% of tested SRS- participants. Many children (61%) were also malnourished or living with HIV (13%), conditions which themselves could contribute to raised CRP levels, including through increased susceptibility to other infections. It might be thought, then, that CRP levels could still be utilized as a "rule-out" test, particularly in children without concomitant illness, but our data showing normal CRP levels in 43% of SRS+ children make this a difficult prospect as well.

*Longitudinal Changes in Protein Biomarkers May Differentiate TB Statuses*

I did, however, observe increases in the levels of Eotaxin and MCP-1, and decreases in IL-6 and IFN-γ, after two weeks of TB treatment in SRS+ individuals. These differences were each statistically significant prior to correcting for multiple hypotheses. Of particular interest is the universal decrease in IL-6 levels observed in SRS+ children who had quantifiable levels of IL-6 at baseline. As samples from the same participant were assayed

in adjacent wells, and only one other cytokine (IFN-γ) on the same MSD panel exhibited a similar decrease in concentration after treatment amongst SRS+ individuals, it is unlikely that this decrease is a technical artifact. Given this, it is possible that dynamic changes in plasma levels of these or other proteins may be useful as a treatment monitoring tool or for retrospective diagnosis. However, further study of how the concentrations of these protein biomarkers fluctuate during and after TB treatment is necessary to determine whether this finding replicates in other independent paediatric cohorts and to better quantify the biomarkers' performance for the aforementioned use cases.

The observed alterations in biomarker concentrations integrate well with the existing literature, however, as IL-6 was previously theorized to be a potential biomarker for TB treatment monitoring from work in an *in vitro* macrophage infection model [186]. Additionally, elevated levels of IL-6 and IFN-γ have been previously associated with TB treatment failure in adults [187]. Interestingly, too, *Mtb* has been reported to exploit IL-6 induction to inhibit IFN-γ [188, 189]. Similarly, our findings of elevated Eotaxin and MCP-1 levels after treatment initiation align with those previously reported [190], with the MCP-1/CCR2 axis thought to contribute to immune evasion by *Mtb* [191]. That MCP-1 levels increase post-treatment is particularly interesting mechanistically, as higher levels are often associated with poorer TB outcomes, though this phenomenon appears to be related to a polymorphism in the *MCP-1* gene [192]. As the NOD-pedFEND cohort was not genotyped, the relation of participant genetics to changes in MCP-1 levels in our dataset is unclear but would be an interesting avenue for further investigation.

*Limitations*

While this work spanned over 50 proteins of high interest for TB diagnosis, it is inherently limited by the specific proteins assayed and low basal levels of some cytokines in the absence of stimulating antigens. Repeating this testing with samples stimulated with TB antigens—such as those used for IGRA testing—would be an interesting approach to attempt to raise the signal-to-noise ratio of any TB-specific biomarkers, or to identify samples which exhibit strong cytokine responses to TB antigens, as has been done in other cohorts [144, 146]. Further, our reliance upon only baseline and two-week testing timepoints constrained my ability to detect alterations to the proteome which occur on a longer time horizon; future work should consider trending measured protein levels across later timepoints.

Non-targeted approaches, such as mass spectrometry or SomaScan, may also help to identify other proteins which could supplement or replace sICAM-1 or the longitudinal markers for even greater diagnostic performance. As alluded to with my technical validation of the sICAM-1 findings, the baseline samples presented here have also recently been characterized using SomaScan, and the analyses of those data are ongoing. By utilizing this broader, near-proteome-wide approach, in conjunction with targeted MSD assays that yield absolute concentration values for proteins of particular interest (perhaps even in samples stimulated with TB antigens), we hope to build a comprehensive understanding of alterations to the plasma proteome in young children with TB.

## 5.12 Conclusion

I performed a targeted analysis of 52 proteins of interest in the blood plasma of 64 children with and without TB disease using the Meso Scale Discovery platform. In so doing, I identified one protein, sICAM-1, which had an AUC for detecting TB in our cohort of between 0.74 and 0.76, depending on the reference standard utilized. This result was technically validated using preliminary data from measurements of the same samples performed using SomaScan. When set to an optimized cutoff which yielded 91% specificity, sICAM-1 identified more children—including MRS- LAM+ CLHIV—than the best-performing NAAT- or culture-based TB test. Additionally, sICAM-1's diagnostic performance appeared to be elevated in children with malnutrition or severe disease, although small sample sizes require caution in interpreting these results.

I also explored alternative approaches to TB diagnosis using the other 51 proteins assayed. I developed novel signatures of TB derived from all measured proteins and evaluated these signatures along with those which had been previously reported in the literature, finding none to meaningfully exceed the diagnostic performance of sICAM-1 alone. Additionally, I evaluated the diagnostic performance of CRP, finding its performance to be worse than that of sICAM-1, VEGF-A, IL-17A, IP-10, or any of the multi-protein signatures of TB that I evaluated.

Alterations in biomarker concentrations after the initiation of TB treatment were also investigated. While none of sICAM-1, VEGF-A, IL-17A, IP-10, or CRP exhibited statistically significant changes in concentration following two weeks of TB treatment in children with TB, significant changes were observed for Eotaxin, MCP-1, IL-6, and IFN-γ; however, after

adjusting for multiple hypotheses, these longitudinal changes were no longer statistically significant. It should be noted, though, that IL-6 concentrations dropped in the two-week samples from all seven of the SRS+ individuals who had baseline IL-6 concentrations above the LLOQ, while IFN-γ levels dropped in all but one of nine participants with baseline levels above the LLOQ. While low sample sizes limit precise calculation of the frequency of this outcome, the near-universal nature of the responses in our cohort particularly supports their future study as potential biomarkers of treatment response or retrospective TB diagnosis in larger, independent cohorts.

In sum, using blood plasma samples from symptomatic children under five with and without TB, I identified a marker of TB disease at baseline, along with potential indicators of TB treatment response. These results reinforce the promise of protein-based biomarkers for paediatric TB disease and identify multiple promising biomarkers for validation in external cohorts.

# Chapter 6: Discussion and Future Directions

## 6.1 Paediatric TB—An Ongoing Challenge

Current diagnostic tests are unable to detect TB disease in the majority of children who are believed to have it [13]. With an estimated 174,000 children dying from TB each year, 96% of whom are reported to not be receiving TB treatment, improved diagnostics are desperately needed to lower the global burden of disease [1, 193].

Here, I sought to determine if measurements of host biology could, in young children, be used as a TB diagnostic. While the generally subdued immune response young children mount to *Mtb* makes such an approach challenging, it is this same dynamic that makes dedicated studies in young children so critical. While host-based TB diagnostics have not yet entered standard practice for adult patients, adult-type TB disease substantially differs in pathology from paediatric disease [36], so it is possible that approaches that do not work in adults may work for children, and vice versa. Further, while diagnostics for adults could still be improved, molecular tests are estimated to detect nearly 90% of adult cases of TB disease [194], compared to about a quarter of cases in children [31]. As such, even tests which perform poorly when compared to the standard of care for adult disease might represent meaningful improvements to the current standard of care for children. Given the paucibacillary nature of paediatric TB disease, host-based assays present a promising theoretical avenue to allow the body itself to amplify the signal of disease, versus pathogen-based approaches which require detection of a remarkably small signal amidst large volumes of noise. In summary, the distinct immune response,

performance of existing diagnostics, and paucibacillary nature of paediatric TB each support a host-based approach to TB diagnosis in children.

## 6.2 Summary of Work Presented

Given the theoretical advantages of a host-based approach to paediatric TB diagnostics, I sought to evaluate it through three projects. Each project utilized samples collected during NOD-pedFEND, a non-interventional diagnostic cohort trial enrolling children under five years of age with presumptive TB in Uganda. I discuss each project in turn.

*Performance of the Host Response Cartridge*

First, I evaluated the performance of a late-stage diagnostic prototype based upon a single, well-researched transcriptomic signature of TB disease originally proposed by Sweeney et al. [61]. Using 100 μl of capillary blood, the test, designed by Cepheid, utilizes the GeneXpert platform to quantify levels of three mRNA transcripts—*KLF2* or *TBP*, along with *GBP5* and *DUSP3*—using RT-qPCR. In our cohort, I did not find the test to exhibit useful diagnostic performance, with AUCs ranging from 0.55-0.57, depending on the scoring approach and reference standard used for evaluation. Further, while I also evaluated alternative HRC scoring methodologies, as well as the use of changes in HRC scores for treatment monitoring or retrospective diagnosis, no approach raised the technology's performance much above that of a random classifier. I also could not identify any subgroups in which the HRC performed meaningfully better than in the NOD-pedFEND cohort overall; this evaluation was, however, inherently limited by a relatively small number

of children with confirmed TB tested by the HRC (n=19). I did, however, observe a statistically significant trend towards lower *TBP*-based HRC scores with increasing age (up to our cohort's maximum age of five) when looking at the larger sample size of children with unconfirmed TB, which speaks to the continued importance of evaluating host-based technologies specifically in young children.

*Transcriptomic Signatures of TB*

With these poor results from one signature, as implemented in a near-commercial technology, I broadened my focus to evaluate 16 previously reported transcriptomic signatures of TB using NanoString. Using RNA extracted from PAXgene tubes collected at baseline and roughly two weeks later, I quantified the expression, in whole blood, of 90 genes in pre-existing signatures of TB diagnosis, prognosis, or treatment response, along with 11 other housekeeping genes. Using the expression data obtained, I determined the diagnostic performance of each signature by scoring each sample for enrichment of the signature's genes using the PLAGE algorithm. From these scores, and the known case classification of each participant, the three-gene signature reported by Roe et al. emerged having the best performance, with a calculated AUC of 0.69 (95% CI 0.60-0.82) when evaluated against the mSRS [100]. Other signatures performed similarly, however, with the signatures with the next-highest AUCs being those developed by Penn-Nicholson et al. (AUC=0.68; 95% CI 0.55-0.78) and Kaul et al. (AUC=0.66; 95% CI 0.55-0.75) [94, 99]. Interestingly, all three of these signatures were developed using data from cohorts which included adolescents, which contrasts with the majority of the other tested signatures, which relied entirely on adult cohorts. Relatedly, signature performance appeared to be

enhanced in children who were two years old or older, as well as in those with malnutrition or severe disease, for 6/6, 6/6, and 4/6 of the top 6 signatures (ranked by AUC), respectively. It should be noted, however, that my evaluation was underpowered to identify statistically significant differences in signature performance between subgroups of participants.

From the NanoString data, I also derived a novel signature of TB disease, which performed similarly to the other gene signatures evaluated, with the top-performing model having an AUC of 0.66. I also explored how all signatures performed as longitudinal markers of treatment response by analysing the difference in signature scores between a participant's samples collected at the baseline and two-week visits. While this evaluation was limited by the small number of children with unlikely TB who were started on TB treatment at baseline, changes in signature scores after treatment initiation did appear to differ between children with and without confirmed TB. In fact, the change in signature scores for all signatures but two—Kulkarni_HIV_2 and Sloot_HIV_2—differentiated between SRS+ and SRS- participants noticably better than a random classifier, with the other two signatures having an AUC CI that was lower-bounded precisely at 0.50. The notion that this statistical significance persisted for nearly all signatures, despite being reliant on an SRS- group which consisted of only four children, is exciting and merits further research to validate my findings and further explore the longitudinal dynamics of TB signatures in children. In addition to being used as markers of treatment response, our data suggest that changes in TB signatures during treatment may assist in retrospectively identifying which treated children truly had TB. While of limited utility for clinical diagnosis

and decision-making prior to treatment initiation, such an approach could help to better inform the reference standards used for paediatric TB research.

*Proteomic Biomarkers of TB*

Finally, I selected a subset of the same participants chosen for the NanoString evaluation to quantify levels of 52 protein biomarkers of interest in their blood plasma samples, which were also collected at the baseline and two-week visits. In evaluating those samples, I identified one protein—sICAM-1—which had respectable diagnostic performance at baseline, with an AUC of between 0.74 and 0.76, for the SRS and eMRS, respectively. I technically validated this result using preliminary data from a proteome-wide analysis, performed using SomaScan, of the same baseline samples; from these data, I calculated a near-identical AUC for sICAM-1 (0.75 against the SRS). However, while it is promising that sICAM-1 could identify SRS+ children even amongst a fully symptomatic cohort, it should be noted that the non-specific nature of the sICAM-1 protein itself suggests it may be of greatest use as a triage test versus a true diagnostic. Further validation of the protein's diagnostic performance for TB is warranted; ideally, these studies should also be adequately powered to identify the optimal population(s) for its use as a diagnostic or triage test. In our cohort, the diagnostic performance of sICAM-1 appeared to be elevated in children with malnutrition or those with more severe disease, but these also may be the populations in which the pre-test probability of other diseases likely to raise sICAM-1 levels—such as sepsis—is also higher [195].

From the protein biomarker data, I also developed a novel signature of TB disease which had an AUC of 0.80. However, as the single most important variable to this signature

was sICAM-1 concentration, and the signature had a similar AUC to that protein alone, I opted not to move forward with the derived signature, believing the near-equivalent AUC to not justify the signature's increased complexity over a single-protein marker. Additionally, I evaluated the performance of previously reported protein-based signatures of TB by training machine learning models only on data from the proteins used in each signature. In so doing, I found no signature to rise above the performance of sICAM-1 alone. Moreover, as the performance of these signatures was likely overestimated by not being evaluated in an independent test set, and was derived from models far more complicated than the simple linear metric of sICAM-1 levels, I also opted not to study any of the previously reported signatures further. I did, however, find promising results in how protein levels changed over two weeks of treatment in SRS+ children, with four proteins—Eotaxin and MCP-1 (increasing), and IL-6 and IFN-γ (decreasing)—exhibiting significant changes in concentration prior to correcting for multiple comparisons. While no changes remained significant after performing that statistical correction, study of these four proteins for treatment monitoring or retrospective diagnosis in children may still be warranted. This is particularly true for IL-6, which decreased in all SRS+ participants who had baseline levels above the LLOQ.

*In Summary*

Taken together, my work spanned a number of distinct approaches to host-based TB diagnostics and identified several promising avenues for future research. Using samples collected from NOD-pedFEND, I observed poor performance of a late-stage diagnostic prototype built off of a transcriptomic signature. Performance improved, but was still below

the WHO TPPs, when I looked to alternative transcriptomic signatures of disease, as measured by NanoString. Performance appeared, however, to be particularly improved in older children (2-5 years old) and in those with severe disease, and perhaps also when signatures were evaluated longitudinally during TB treatment, although confirmatory studies are needed. Finally, I identified promising protein biomarkers for diagnostic testing at baseline (sICAM-1), along with others which appear to alter in concentration after two weeks of TB treatment (IL-6 and others). Validating these results in independent cohorts will be critical to understanding their generalizability and to further elucidate the potential use cases for host-based biomarkers in TB diagnosis.

## 6.3 Limitations of the Work and Continued Challenges for the Field

Our work is limited, however, by many of the same challenges which have vexed the paediatric TB diagnostics field writ large. In particular, a small proportion of confirmed TB cases amongst all participants, imperfect reference standards, and a geographically limited cohort may reduce the generalizability and interpretability of my findings.

*Low Rates of Microbiological Confirmation*

In NOD-pedFEND overall, only 43/640 (7%) of all Ugandan participants tested positive for TB on any of the microbiological tests performed and were therefore SRS+. Expanding our definition of TB-positive cases to include CLHIV with a positive LAM test (so-called eMRS+ participants) increased this number to 59/640 (9%) of participants, but this still represented a small fraction of the full cohort. Additionally, only 144/640 (23%) of

participants were classified as “unlikely TB” cases (1 of whom was eMRS+, and 4 of whom were eMRS unclassifiable), with the vast majority (383/640, 60%) being classified as “unconfirmed TB” cases, with 347/640 (54%) being eMRS- unconfirmed TB. Children with unconfirmed TB are particularly challenging for TB diagnostic studies, as given the low sensitivity of existing microbiological tests, their TB status is the least certain. The proportion of participants with unconfirmed TB in our cohort was notably higher than that of confirmed or unlikely TB and diverged from proportions reported by other recent studies, such as the RaPaed-TB trial, which enrolled 282/965 (29%) children who were eventually classified as having unconfirmed TB [30]. This discrepancy may stem from, at least in part, the relatively low microbiological confirmation rate observed in the NOD-pedFEND cohort, although this itself may simply be a reflection of NOD-pedFEND’s exclusive focus on children under five, which are the hardest age group to microbiologically diagnose [196]. Relatedly, sputum testing was not performed in NOD-pedFEND due in part to the difficulty of getting young children to spontaneously expectorate sputum, and this may have further limited confirmation rates and raised levels of unconfirmed TB.

*The Proportion of Children with Unconfirmed TB*

A larger proportion of children with unconfirmed TB could influence the reported results in distinct ways which largely depend upon the characteristics of the unconfirmed TB cohort. If it is assumed that the additional children with unconfirmed TB were primarily children with TB lacking microbiological confirmation and not children with a non-TB respiratory disease, analyses which utilize the CRS would not be substantially impacted, while analyses based upon the MRS or eMRS will misclassify those additional children with

TB as reference standard negative. The SRS would also be impacted, but primarily by losing power in the SRS+ group; as children with unconfirmed TB are not included in the SRS- group, the classification of more children with TB as unconfirmed TB would simply lead more children to be SRSun, minimally influencing results based upon the SRS (this is, in fact, the primary advantage of the SRS). In contrast, if the additional children with unconfirmed TB primarily did not have TB, and would have otherwise been classified as "unlikely TB", the MRS and eMRS would not be as impacted, as these children would still be considered MRS- and eMRS-. However, analyses against the CRS would be significantly altered, as these children would move from being CRS- to CRS+. Finally, and as before, the SRS would be minimally impacted, primarily by decreased statistical power, though in this case that would largely be caused by a decrease in the number of participants in the SRS- group.

Potential factors which could have caused children who would have received microbiological confirmation in other studies from not receiving it in ours, beyond simple differences in study populations, include errors in specimen collection or microbiological test processing, data entry errors, or the lack of sputum or other microbiological testing. These concerns are attenuated, however, by testing having been conducted in an accredited clinical laboratory, with strict sample and data handling practices. Additionally, it has been reported that Xpert testing of stool is well correlated with that of induced sputum [197]; further, gastric aspirate has been reported to have higher diagnostic yield than sputum in children under 10 [198], though prior work comparing induced sputum to gastric lavage is contradictory [199].

In contrast, factors which could have caused children without TB disease to be classified as unconfirmed TB at a higher rate than in other studies, beyond simple differences in the study population, include the potential for overestimation of TST induration, the measurement of which has been shown to often be variable [200]. Additionally, the NOD-pedFEND study's thorough documentation of TB exposure history, particularly its follow-up on contacts reported to have a chronic cough, could also increase rates of unconfirmed TB. This would not be a result of misclassification, as symptomatic children with a known source of TB exposure would be correctly classified as having unconfirmed TB by the NIH criteria [65]. However, if exposure history was not similarly taken by other studies, NOD-pedFEND's practice could still have contributed to a higher rate of unconfirmed TB in our cohort, with unclear influences on classification accuracy.

Finally, NOD-pedFEND's relatively high rate of TB treatment (73%) also highlights one of the most likely contributors to the increased levels of unconfirmed TB: higher rates of TB treatment yielding more children considered to have "responded to treatment", and therefore to be classified as having unconfirmed TB by the NIH criteria, despite TB treatment not having contributed to their resolution of symptoms. As many respiratory diseases can self-resolve [201], the reliance on symptom resolution while on treatment as a proxy for "treatment response" could certainly label some children as "treatment responders" despite having not benefitted from anti-TB treatment in reality. Given the high rates of TB treatment in NOD-pedFEND, this possibility could significantly influence rates of unconfirmed TB by reclassifying children who would have otherwise been classified as "unlikely TB" due to an observed "treatment response".

*Imperfect Reference Standards*

These factors highlight how the retrospectively determined TB status of participants in our cohort can be influenced by clinical decision-making, as well as by deliberate choices made in the classification process. The subjectivity introduced by these factors comprises another limitation of my results, and those of many other studies in the field more broadly, which also requires explicit consideration.

While the broad approach to TB case classification in diagnostic studies has largely coalesced around foundational definitions developed by a NIH-convened expert panel and reported in a 2015 paper by Graham et al. [65], these guidelines are not comprehensive and still leave room for subjective interpretation that can differ between studies. The criteria we used for assigning classifications in line with the definitions of Graham et al. are given in **Section 2.5**. As alluded to above, however, the criterion with perhaps the largest potential for subjectivity and discrepancy between studies is the determination of "response to treatment". Per the Graham criteria, any symptomatic children who demonstrate a "response to treatment" in the absence of microbiological confirmation of disease are to be classified as having "unconfirmed TB". What it means to demonstrate that response, and when it should be assessed, however, is primarily left to the study clinicians and researchers to determine. For this work, we evaluated response to treatment at two time points: two months into treatment and at the end of treatment. We defined a participant as having responded to treatment if, for any of four critical symptoms— weight loss, cough, fever, or fatigue—present at the time of enrolment, all had resolved at the evaluated timepoint. Additionally, the site clinician had to also believe that a response to

treatment had been demonstrated. Other studies, however, likely arrived at slightly different definitions, and how those definitions differ could classify the same child as having unconfirmed TB in one study and unlikely TB in another. I attempted to attenuate these concerns by focusing my analysis on the SRS and eMRS—standards for which positivity is not dependent on the differentiation between unlikely and unconfirmed TB—and by implementing definitions which increased the stringency of the "unlikely TB" classification.

The primary method we employed to increase confidence in the unlikely TB group was the creation of an "unclassifiable" group, which is not explicitly called for by the 2015 guidelines. Children could be considered unclassifiable if there was insufficient information available to assign them to a definitive classification, such as when a CXR or clinical data needed to determine response to treatment was missing and had the potential to change the child's classification. Children started on tuberculosis preventative therapy (TPT) were also automatically categorized as "unclassifiable", given the unknown influence of TPT on TB disease or the host response to it. Additionally, we utilized a three-member CXR review panel to ensure that CXR interpretations were agreed upon by more than one reviewer, fortifying my analysis against the known inter-rater variability in evaluating paediatric CXRs for TB [202]. Together, these approaches increased my confidence in participants' assigned classifications.

For this work, classifications were decided algorithmically, using an internally agreed-upon interpretation of the 2015 consensus case definitions further described in **Section 2.5.** Moving forward, we are also working to compare this strict application of the

2015 case definitions to one based upon the consensus opinion of three experts in childhood TB, an approach which permits greater clinical discretion in a participant's ultimate case classification. In comparing the classifications derived via the algorithm used here with those determined by clinical experts, we hope to engender further confidence in the reference standards utilized for my analysis. However, these actions are in no way a comprehensive fix for the more foundational concerns about reference standard accuracy. As is also the case for other paediatric diagnostic studies, my analyses are inherently limited by the uncertainty present in the reference standards utilized.

*Geographic Limitations*

Finally, it should be noted the research presented here originates only from samples collected in Uganda from a single study: NOD-pedFEND. The study spanned two distinct sites in southern Uganda. The first, based in Uganda's capital of Kampala, was an academic medical facility specializing in paediatric and HIV/AIDS care. The second, based 54 miles east of Kampala in Uganda's second largest urban district, was set in the paediatric annex of the region's referral hospital. As a result, while our research may be reflective of diagnostic performance in children in southern Uganda, it is as of yet unclear how these results would generalize to broader populations. Additionally, while a strength of NOD-pedFEND was its focus solely on the children most difficult to diagnose—those children under five years of age—this also means that our data do not shed light on diagnostic performance in older children. NOD-pedFEND's exclusive focus on pulmonary TB should also be noted, as well as its relative high rate of malnutrition (51%) and, perhaps

relatedly, TST negativity (53%) amongst children with confirmed TB, which may have been indicative of a more muted host immune response.

## 6.4 Future Directions

These limitations, however, also reveal promising avenues for future exploration. Given our limited number of microbiologically confirmed TB cases, meta-analyses and dedicated validation studies in independent cohorts would help to build statistical power, generalizability, and to enable a more robust interrogation of subgroup-level performance. Further, the continued efforts of an expert review committee to assign NOD-pedFEND participants with consensus classifications will assist in better understanding the strengths and limitations of the reference standards utilized in the work presented here. Our ongoing expansion of the protein-based biomarker work to a non-targeted, proteome-wide approach using SomaScan is also exciting, as it may help us to further develop novel protein-based signatures of paediatric TB disease as well as to validate the signatures proposed by other groups which included proteins that were not measured by the MSD assays. A similar omics-level approach may also be justifiable for the continued study of transcriptomic biomarkers, particularly in samples which have been stimulated with TB antigens.

*Suggested Next Steps for the Field: Improving the Ground Truth Used for Diagnostic Studies*

More generally, the field of paediatric TB diagnostics continues to be challenged by the lack of a clear ground truth against which novel diagnostics can be evaluated. Samples from many children—all those who are classified as having unconfirmed TB—are typically

difficult to include in at least first-stage diagnostic evaluations, as the assumed heterogeneity of the unconfirmed TB group makes it unclear what testing result would be the "correct" one for a diagnostic to give. As such, work to better clarify this ground truth would be beneficial to paediatric TB research efforts. My results suggest that measuring the influence of TB treatment on signatures of disease may help to better define reference standards, as children with confirmed TB may be more likely to experience dramatic changes in TB biomarker levels over the treatment course. It may even be possible to develop specific, dedicated signatures for the retrospective classification of TB disease. With this knowledge of which children had biomarkers of disease that changed after they received TB treatment, it may be possible to better identify "true positive" cases for future paediatric TB research. Not only would this improve the accuracy of the reference standards utilized, but it would also allow diagnostic evaluations to be targeted at those children for whom better tests are most needed: the "unconfirmed TB" group undetected by current diagnostics. If researchers could gain a better understanding of who in the unconfirmed group truly has TB, they could evaluate novel tests against these more difficult cases. This would also allow for more efficient studies by minimizing the number of enrolled children with uncertain diagnoses, whose samples are of limited use in early-stage diagnostic evaluations. An alternative approach to strengthening the ground truth used in diagnostic evaluations could be to look for TB via post-mortem testing and autopsies; however, using longitudinal biomarkers instead presents a less resource-intensive and ethically complex approach to better defining reference standards, one which is deserving of further study.

*Additional Paths Worth Exploring*

Beyond this, updates to the 2015 NIH consensus case classifications would also assist research in paediatric TB diagnostics, particularly in light of recent technological developments such as the advent of Xpert Ultra and the molecular testing of stool. More explicitly defined and operationalized case definitions would be conducive to cross-study comparisons, helping to catalyse the aforementioned meta-analyses which could in turn compensate for the relatively small numbers of children with confirmed TB in many individual studies. It would also be interesting to evaluate diagnostics against the treatment recommendation given by the WHO's treatment decision algorithms (TDAs) for paediatric TB. Using the TDAs as a comparative reference standard would help to elucidate the added value of TDAs to laboratory testing, and vice versa, and the ideal use cases for each. Ultimately, diagnostic testing is but one facet of a large system of tuberculosis identification, prevention, and treatment; understanding how each aspect of a system complements the others is crucial to developing diagnostics which are truly helpful and implementable.

*In Summary*

While mixed, my results demonstrate the continued promise of host-based diagnostics for paediatric TB. While not all attempted approaches were successful, many showed promise, particularly the use of changes to biomarkers after treatment initiation, as well as sICAM-1 levels measured at baseline. Moreover, my distinctly different results for the diagnostic performance of the HRC and other TB signatures, as compared to their reported performance in adults, emphasized the importance of dedicated evaluations in

young children; just as many host-based diagnostics appear to perform better in adults than they do in children, it is also possible that some perform better in children than they do in adults. Through dedicated study, in young children, of a diversity of approaches to host-based diagnostics, it is possible that true progress on paediatric TB diagnosis can be made. And with over one million children battling TB disease each year [1], such progress is sorely needed.

# Chapter 7: References


1. *Global Tuberculosis Report 2025*. 2025, World Health Organization.
2. Herzog, H., *History of Tuberculosis.* Respiration, 1998. **65**(1): p. 5-15.
3. *Tuberculosis*. Available from: https://www.who.int/news-room/fact-sheets/detail/tuberculosis.
4. Centers for Disease Control and Prevention. *Ending the Global Tuberculosis Epidemic*. 2024 June 13, 2024 March 18, 2026]; Available from: https://www.cdc.gov/global-hiv-tb/php/our-approach/combatingglobaltb.html.
5. Zumla, A., et al., *Inequities underlie the alarming resurgence of Tuberculosis as the world's top cause of death from an Infectious Disease - Breaking the silence and addressing the underlying root causes.* IJID Reg, 2025. **14**(Suppl 2): p. 100587.
6. Martinez, L., et al., *The risk of tuberculosis in children after close exposure: a systematic review and individual-participant meta-analysis.* The Lancet, 2020. **395**(10228): p. 973-984.
7. Courtwright, A. and A.N. Turner, *Tuberculosis and stigmatization: pathways and interventions.* Public Health Rep, 2010. **125 Suppl 4**(Suppl 4): p. 34-42.
8. Jenkins, H.E., et al., *Mortality in children diagnosed with tuberculosis: a systematic review and meta-analysis.* Lancet Infect Dis, 2017. **17**(3): p. 285-295.
9. Seddon, J.A., et al., *The Wonder Years: What Can Primary School Children Teach Us About Immunity to Mycobacterium tuberculosis?* Front Immunol, 2018. **9**: p. 2946.
10. DiNardo, A.R., et al., *Culture is an imperfect and heterogeneous reference standard in pediatric tuberculosis*. Tuberculosis, 2016. **101**: p. S105-S108.
11. Kunkel, A., et al., *Smear positivity in paediatric and adult tuberculosis: systematic review and meta-analysis.* BMC infectious diseases, 2016. **16**(1): p. 282.
12. World Health Organization, *WHO consolidated guidelines on tuberculosis: module 3: diagnosis*. 2025.
13. Signorino, C., et al., *Diagnostic accuracy of Xpert ultra for childhood tuberculosis: A preliminary systematic review and meta-analysis.* Pediatr Allergy Immunol, 2022. **33 Suppl 27**(Suppl 27): p. 80-82.
14. Steingart, K.R., et al., *Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults.* Cochrane database of systematic reviews, 2013(1).
15. World Health Organization, *Lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis in people living with HIV, 2019 Update*. 2019.
16. García-Basteiro, A.L., et al., *Radiological findings in young children investigated for tuberculosis in Mozambique.* PLoS One, 2015. **10**(5): p. e0127323.
17. Navuluri, N., et al., *Clinician and radiologist chest radiography interpretation variability in a high-TB-burden setting.* IJTLD Open, 2026. **3**(1): p. 53-55.
18. Abraham, Y., et al., *Facilitators and barriers to implementing chest radiography in tuberculosis systematic screening of clinically high-risk groups in Ethiopia: A qualitative study.* SAGE Open Med, 2024. **12**: p. 20503121241233232.

19. Sofyan, D. and R. Shabariah, *The Relationship between Mantoux Sensitivity and The Diagnosis of Childhood Tuberculosis: A Comprehensive Systematic Review*. The International Journal of Medical Science and Health Research, 2026. **33**(1): p. 1-62.
20. Pouchot, J., et al., *Reliability of Tuberculin Skin Test Measurement*. Annals of Internal Medicine, 1997. **126**(3): p. 210-214.
21. Lu, P., et al., *Interferon-Gamma Release Assays for the Diagnosis of Tuberculosis: A Systematic Review and Meta-analysis*. Lung, 2016. **194**(3): p. 447-458.
22. Volkman, T., et al., *QuantiFERON Gold-In-Tube for the diagnosis of mycobacterial tuberculosis infection in children under 5 years of age: A systematic review and meta-analysis*. Plos one, 2024. **19**(1): p. e0295913.
23. Dagnew, A.F., et al., *Interferon-gamma release assay positivity in populations at high risk of TB infection*. IJTLD Open, 2026. **3**(4): p. 232-240.
24. Menzies, N.A., et al., *Time Since Infection and Risks of Future Disease for Individuals with Mycobacterium tuberculosis Infection in the United States*. Epidemiology, 2021. **32**(1): p. 70-78.
25. Cohen, A., et al., *The global prevalence of latent tuberculosis: a systematic review and meta-analysis*. Eur Respir J, 2019. **54**(3).
26. Houben, R.M. and P.J. Dodd, *The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling*. PLoS Med, 2016. **13**(10): p. e1002152.
27. Frydenberg, A.R. and S.M. Graham, *Toxicity of first-line drugs for treatment of tuberculosis in children: review*. Tropical Medicine & International Health, 2009. **14**(11): p. 1329-1337.
28. Randremanana, R.V., et al., *A multicountry evaluation of the Xpert MTB/RIF assay for the diagnosis of intrathoracic tuberculosis in children using alternative specimens (nasopharyngeal aspirate and stool): A prospective cohort study conducted in Madagascar, Ivory Coast and Cameroon (TB KIDS project)*. International Journal of Infectious Diseases, 2025. **151**: p. 107366.
29. World Health Organization, *Diagnosis of tuberculosis and detection of drug-resistance: rapid communication*. 2024.
30. Olbrich, L., et al., *Sequential and parallel testing for microbiological confirmation of tuberculosis disease in children in five low-income and middle-income countries: a secondary analysis of the RaPaed-TB study*. The Lancet Infectious Diseases, 2025. **25**(2): p. 188-197.
31. Olbrich, L., et al., *Parallel use of low-complexity automated nucleic acid amplification tests on respiratory and stool samples with or without lateral flow lipoarabinomannan assays to detect pulmonary tuberculosis disease in children*. Cochrane Database of Systematic Reviews, 2025(6).
32. Nicol, M.P., et al., *Accuracy of the Xpert MTB/RIF test for the diagnosis of pulmonary tuberculosis in children admitted to hospital in Cape Town, South Africa: a descriptive study*. Lancet Infect Dis, 2011. **11**(11): p. 819-24.
33. Hernanz-Lobo, A., et al., *Diagnostic performance of chest radiography for pediatric tuberculosis across high- and low-burden settings*. Front Pediatr, 2025. **13**: p. 1704149.

34. Thomas, T.A., *Tuberculosis in Children*. Pediatr Clin North Am, 2017. **64**(4): p. 893-909.
35. Marais, B.J., *Childhood Tuberculosis: Epidemiology and Natural History of Disease*. The Indian Journal of Pediatrics, 2011. **78**(3): p. 321-327.
36. Basu Roy, R., et al., *Tuberculosis susceptibility and protection in children*. The Lancet Infectious Diseases, 2019. **19**(3): p. e96-e108.
37. *HIV - Elimination of mother-to-child transmission*. Available from: https://data.unicef.org/topic/hivaids/emtct/.
38. Madhi, S.A., et al., *Primary isoniazid prophylaxis against tuberculosis in HIV-exposed children*. N Engl J Med, 2011. **365**(1): p. 21-31.
39. Cotton, M.F., et al., *Tuberculosis exposure in HIV-exposed infants in a high-prevalence setting*. Int J Tuberc Lung Dis, 2008. **12**(2): p. 225-7.
40. Olbrich, L., et al., *Untangling the complex relationship between HIV-exposure and tuberculosis in children: a narrative review*. The Lancet Global Health, 2026. **14**(5): p. e817-e828.
41. Getahun, H., et al., *Prevention, diagnosis, and treatment of tuberculosis in children and mothers: evidence for action for maternal, neonatal, and child health services*. Journal of Infectious Diseases, 2012. **205**(suppl_2): p. S216-S227.
42. Bekker, A., et al., *High tuberculosis exposure among neonates in a high tuberculosis and human immunodeficiency virus burden setting*. International journal of tuberculosis and lung disease, 2012. **16**(8): p. 1040-1046.
43. Pillay, T., et al., *Vertical transmission of Mycobacterium tuberculosis in KwaZulu Natal: impact of HIV-1 co-infection*. International Journal of Tuberculosis and Lung Disease, 2004. **8**(1): p. 59-69.
44. Kukarkina, V.A., A.A. Golubkova, and A.S. Podymova, *Prevalence of tuberculosis among perinatally exposed to HIV and HIV-infected children and predictors of its development*. Journal of microbiology, epidemiology and immunobiology, 2021. **98**(3): p. 351-361.
45. Cranmer, L.M., et al., *High incidence of tuberculosis infection in HIV-exposed children exiting an isoniazid preventive therapy trial*. The Pediatric infectious disease journal, 2018. **37**(10): p. e254-e256.
46. Marquez, C., et al., *Tuberculosis infection in early childhood and the association with HIV-exposure in HIV-uninfected children in rural Uganda*. The Pediatric infectious disease journal, 2016. **35**(5): p. 524-529.
47. Mecha, J., et al., *Maternal HIV status and risk of infant mycobacterium tuberculosis infection as measured by tuberculin skin test*. The Pediatric infectious disease journal, 2024. **43**(3): p. 250-256.
48. Moore, D.P., et al., *The etiology of pneumonia in HIV-uninfected South African children: findings from the pneumonia etiology research for child health (PERCH) study*. The Pediatric infectious disease journal, 2021. **40**(9S): p. S59-S68.
49. Mahomed, H., et al., *The impact of a change in bacille Calmette-Guérin vaccine policy on tuberculosis incidence in children in Cape Town, South Africa*. Pediatr Infect Dis J, 2006. **25**(12): p. 1167-72.

50. da Costa, F.B.P., et al., *Mycobacterium tuberculosis infection and tuberculosis disease in the first decade of life: a South African birth cohort study.* Lancet Child Adolesc Health, 2024. **8**(12): p. 891-899.
51. Marais, B.J., et al., *The natural history of childhood intra-thoracic tuberculosis: a critical review of literature from the pre-chemotherapy era.* Int J Tuberc Lung Dis, 2004. **8**(4): p. 392-402.
52. Moyo, S., et al., *Age-related tuberculosis incidence and severity in children under 5 years of age in Cape Town, South Africa*. The International journal of tuberculosis and lung disease, 2010. **14**(2): p. 149-154.
53. Martinez, L., et al., *Paediatric tuberculosis transmission outside the household: challenging historical paradigms to inform future public health strategies.* Lancet Respir Med, 2019. **7**(6): p. 544-552.
54. le Roux, S.M., et al., *Infectious morbidity of breastfed, HIV-exposed uninfected infants under conditions of universal antiretroviral therapy in South Africa: a prospective cohort study.* The Lancet Child & Adolescent Health, 2020. **4**(3): p. 220-231.
55. Bentley, T.G., A. Catanzaro, and T.G. Ganiats, *Implications of the impact of prevalence on test thresholds and outcomes: lessons from tuberculosis.* BMC Res Notes, 2012. **5**: p. 563.
56. Basile, F.W., et al., *Acceptability and feasibility of oral swabs for tuberculosis diagnosis in young children: a qualitative study from Uganda and Peru*. Frontiers in Tuberculosis, 2026. **Volume 4 - 2026**.
57. Broger, T., et al., *Diagnostic yield as an important metric for the evaluation of novel tuberculosis tests: rationale and guidance for future research*. The Lancet Global Health, 2024. **12**(7): p. e1184-e1191.
58. *Target product profile for tuberculosis diagnosis and detection of drug resistance*. 2024: World Health Organization.
59. *Target product profiles for tuberculosis screening tests*. 2025: World Health Organization.
60. Fischer, M.H., *Fischerisms: Being a Sheaf of Sundry and Divers Utterances Culled from the Lectures of.* Southern Medical Journal, 1937. **30**(10): p. 1038.
61. Sweeney, T.E., et al., *Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis.* Lancet Respir Med, 2016. **4**(3): p. 213-24.
62. Song, R., et al., *Novel and optimized diagnostics for pediatric TB in endemic countries: NOD-pedFEND study protocol.* BMC Pediatr, 2025. **25**(1): p. 647.
63. *World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants.* Jama, 2025. **333**(1): p. 71-74.
64. Harris, P.A., et al., *Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support.* Journal of Biomedical Informatics, 2009. **42**(2): p. 377-381.
65. Graham, S.M., et al., *Clinical Case Definitions for Classification of Intrathoracic Tuberculosis in Children: An Update.* Clin Infect Dis, 2015. **61Suppl 3**(Suppl 3): p. S179-87.

66. *Child growth standards*. June 18, 2026]; Available from: https://www.who.int/tools/child-growth-standards.
67. Seid, G., et al., *Value of urine-based lipoarabinomannan (LAM) antigen tests for diagnosing tuberculosis in children: systematic review and meta-analysis.* IJID Reg, 2022. **4**: p. 97-104.
68. Wright, V.J., et al., *Diagnosis of Kawasaki Disease Using a Minimal Whole-Blood Gene Expression Signature.* JAMA Pediatrics, 2018. **172**(10): p. e182293-e182293.
69. Andre, F., et al., *Biomarkers for adjuvant endocrine and chemotherapy in early-stage breast cancer: ASCO guideline update.* Journal of Clinical Oncology, 2022. **40**(16): p. 1816-1837.
70. Na, R., et al., *Clinically available RNA profiling tests of prostate tumors: utility and comparison.* Asian J Androl, 2016. **18**(4): p. 575-9.
71. Scicluna, B.P., et al., *A consensus blood transcriptomic framework for sepsis.* Nature Medicine, 2025. **31**(12): p. 4119-4130.
72. Sebzda, E., et al., *Transcription factor KLF2 regulates the migration of naive T cells by restricting chemokine receptor expression patterns.* Nature Immunology, 2008. **9**(3): p. 292-300.
73. Fagerberg, E., et al., *KLF2 maintains lineage fidelity and suppresses CD8 T cell exhaustion during acute LCMV infection*. Science, 2025. **387**(6735): p. eadn2337.
74. Richardson, M.W., et al., *Kruppel-like factor 2 modulates CCR5 expression and susceptibility to HIV-1 infection.* J Immunol, 2012. **189**(8): p. 3815-21.
75. Singh, P., et al., *DUSP3 Genetic Deletion Confers M2-like Macrophage–Dependent Tolerance to Septic Shock.* The Journal of Immunology, 2015. **194**(10): p. 4951-4962.
76. Krapp, C., et al., *Guanylate Binding Protein (GBP) 5 Is an Interferon-Inducible Inhibitor of HIV-1 Infectivity.* Cell Host & Microbe, 2016. **19**(4): p. 504-514.
77. Kim, B.-H., et al., *IFN-Inducible GTPases in Host Cell Defense.* Cell Host & Microbe, 2012. **12**(4): p. 432-444.
78. Moreira, F.M.F., et al., *Blood-based host biomarker diagnostics in active case finding for pulmonary tuberculosis: a diagnostic case-control study.* EClinicalMedicine, 2021. **33**.
79. Seifert, M., et al., *Prospective Exploratory Evaluation of Cepheid Xpert Mycobacterium tuberculosis Host Response Cartridge: A Focus on Adolescents and Young Adults.* Clinical Infectious Diseases, 2025. **80**(1): p. 180-188.
80. Organization, W.H. *High priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting, 28-29 April 2014, Geneva, Switzerland*. in *High priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting, 28-29 April 2014, Geneva, Switzerland*. 2014.
81. Sutherland, J.S., et al., *Performance of Two Fingerstick Blood Tests to Triage Adults with Symptoms of Pulmonary TB: A Prospective Multisite Diagnostic Accuracy Study.*
82. Wu, X., et al., *Assessment of the Cepheid 3-gene Host Response Fingerstick Blood Test (MTB-HR) on rapid diagnosis of tuberculosis.* Emerg Microbes Infect, 2023. **12**(2): p. 2261561.

83. Li, M., et al., *Evaluation of the Cepheid 3-gene host response blood test for tuberculosis diagnosis and treatment response monitoring in a primary-level clinic in rural China*. J Clin Microbiol, 2023. **61**(11): p. e0091123.
84. Olbrich, L., et al., *Diagnostic accuracy of a three-gene Mycobacterium tuberculosis host response cartridge using fingerstick blood for childhood tuberculosis: a multicentre prospective study in low-income and middle-income countries*. The Lancet Infectious Diseases, 2024. **24**(2): p. 140-149.
85. Sweetser, B., et al., *A prospective evaluation of a three-gene host response signature to classify tuberculosis severity in children*. Journal of the Pediatric Infectious Diseases Society, 2025: p. piaf041.
86. Poore, H., et al., *The Three-Gene Xpert Host Response Signature for Pediatric Tuberculosis Screening: A Prospective Diagnostic Accuracy Study*. Clin Infect Dis, 2025. **81**(5): p. e424-e431.
87. Sun, L., et al., *A blood-based Xpert host response assay for diagnosis and differentiation of active and latent tuberculosis in children and adolescents: A multi-center, prospective study*. Journal of Infection, 2026. **92**(2): p. 106678.
88. Zimmer, A.J., et al., *A novel blood-based assay for treatment monitoring of tuberculosis*. BMC Res Notes, 2021. **14**(1): p. 247.
89. Darboe, F., et al., *Diagnostic performance of an optimized transcriptomic signature of risk of tuberculosis in cryopreserved peripheral blood mononuclear cells*. Tuberculosis, 2018. **108**: p. 124-126.
90. Zak, D.E., et al., *A blood RNA signature for tuberculosis disease risk: a prospective cohort study*. Lancet, 2016. **387**(10035): p. 2312-2322.
91. Francisco, N.M., et al., *Diagnostic accuracy of a selected signature gene set that discriminates active pulmonary tuberculosis and other pulmonary diseases*. Journal of Infection, 2017. **75**(6): p. 499-510.
92. Jacobsen, M., et al., *Candidate biomarkers for discrimination between infection and disease caused by Mycobacterium tuberculosis*. J Mol Med (Berl), 2007. **85**(6): p. 613-21.
93. Kaipilyawar, V., et al., *Development and Validation of a Parsimonious Tuberculosis Gene Signature Using the digital NanoString nCounter Platform*. Clin Infect Dis, 2022. **75**(6): p. 1022-1030.
94. Kaul, S., et al., *Latent Tuberculosis Infection Diagnosis among Household Contacts in a High Tuberculosis-Burden Area: a Comparison between Transcript Signature and Interferon Gamma Release Assay*. Microbiol Spectr, 2022. **10**(2): p. e0244521.
95. Kulkarni, V., et al., *A Two-Gene Signature for Tuberculosis Diagnosis in Persons With Advanced HIV*. Front Immunol, 2021. **12**: p. 631165.
96. Leong, S., et al., *Cross-validation of existing signatures and derivation of a novel 29-gene transcriptomic signature predictive of progression to TB in a Brazilian cohort of household contacts of pulmonary TB*. Tuberculosis, 2020. **120**: p. 101898.
97. Lopez, K., et al., *Development of a four-gene host signature for paucibacillary TB among symptomatic individuals with sputum Xpert MTB/RIF Ultra very low and trace results*. International Journal of Infectious Diseases, 2026. **169**: p. 108763.

98. Maertzdorf, J., et al., *Concise gene signature for point-of-care classification of tuberculosis*. EMBO Mol Med, 2016. **8**(2): p. 86-95.
99. Penn-Nicholson, A., et al., *RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response*. Scientific Reports, 2020. **10**(1): p. 8629.
100. Roe, J., et al., *Blood Transcriptomic Stratification of Short-term Risk in Contacts of Tuberculosis*. Clin Infect Dis, 2020. **70**(5): p. 731-737.
101. Sambarey, A., et al., *Unbiased Identification of Blood-based Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks*. EBioMedicine, 2017. **15**: p. 112-126.
102. Sloot, R., et al., *Biomarkers Can Identify Pulmonary Tuberculosis in HIV-infected Drug Users Months Prior to Clinical Diagnosis*. EBioMedicine, 2015. **2**(2): p. 172-9.
103. Suliman, S., et al., *Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis*. Am J Respir Crit Care Med, 2018. **197**(9): p. 1198-1208.
104. Roe, J.K., et al., *Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis*. JCI Insight, 2016. **1**(16).
105. Geiss, G.K., et al., *Direct multiplexed measurement of gene expression with color-coded probe pairs*. Nature biotechnology, 2008. **26**(3): p. 317-325.
106. Reis, P.P., et al., *mRNA transcript quantification in archival samples using multiplexed, color-coded probes*. BMC biotechnology, 2011. **11**(1): p. 46.
107. Speranza, E., et al., *Comparison of transcriptomic platforms for analysis of whole blood from Ebola-infected cynomolgus macaques*. Scientific Reports, 2017. **7**(1): p. 14756.
108. Bruker Spatial Biology, I. *Custom Solutions Overview*. [cited 2026 June 25]; Photo, used with permission.]. Available from: https://brukerspatialbiology.com/products/ncounter-assays-panels/ncounter-custom-solutions/custom-solutions-overview/.
109. Bontonou, M., et al., *A comparative analysis of gene expression profiling by statistical and machine learning approaches*. Bioinformatics Advances, 2025. **5**(1).
110. Wang, X., et al., *Comparison of gene set scoring methods for reproducible evaluation of tuberculosis gene signatures*. BMC Infect Dis, 2024. **24**(1): p. 610.
111. Tomfohr, J., J. Lu, and T.B. Kepler, *Pathway level analysis of gene expression using singular value decomposition*. BMC Bioinformatics, 2005. **6**: p. 225.
112. Johnson, W.E., et al., *Comparing tuberculosis gene signatures in malnourished individuals using the TBSignatureProfiler*. BMC Infect Dis, 2021. **21**(1): p. 106.
113. Spencer, T.F.a.E.L.a.N.G.a., et al., *h2o: R Interface for the 'H2O' Scalable Machine Learning Platform*. 2024.
114. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments*. Bioinformatics, 2012. **28**(6): p. 882-3.
115. Bruker Spatial Biology, I. *Gene Expression Data Analysis Guidelines*. 2024; Available from: https://university.nanostring.com/gene-expression-data-analysis-guidelines/894768.
116. Davies, L.R.L., et al., *Age and sex influence antibody profiles associated with tuberculosis progression*. Nat Microbiol, 2024. **9**(6): p. 1513-1525.

117. Berry, M.P., et al., *An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis*. Nature, 2010. **466**(7309): p. 973-7.
118. Turner, C.T., et al., *Blood transcriptional biomarkers for active pulmonary tuberculosis in a high-burden setting: a prospective, observational, diagnostic accuracy study*. Lancet Respir Med, 2020. **8**(4): p. 407-419.
119. Sivakumaran, D., et al., *Combining host-derived biomarkers with patient characteristics improves signature performance in predicting tuberculosis treatment outcomes*. Commun Biol, 2020. **3**(1): p. 359.
120. Mulenga, H., et al., *The effect of host factors on discriminatory performance of a transcriptomic signature of tuberculosis risk*. EBioMedicine, 2022. **77**: p. 103886.
121. Gjøen, J.E., et al., *Novel transcriptional signatures for sputum-independent diagnostics of tuberculosis in children*. Scientific Reports, 2017. **7**(1): p. 5839.
122. Chen, L., et al., *Assessment of ferroptosis-associated gene signatures as potential biomarkers for differentiating latent from active tuberculosis in children*. Microb Genom, 2023. **9**(5).
123. Schibler, U., *The daily timing of gene expression and physiology in mammals*. Dialogues Clin Neurosci, 2007. **9**(3): p. 257-72.
124. An, P., et al., *Regulatory polymorphisms in the cyclophilin A gene, PPIA, accelerate progression to AIDS*. PLoS pathogens, 2007. **3**(6): p. e88.
125. Zhu, Q., et al., *RAB13 regulates macrophage polarization in sepsis*. Scientific Reports, 2024. **14**(1): p. 20400.
126. Salleron, L., et al., *DERA is the human deoxyribose phosphate aldolase and is involved in stress response*. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2014. **1843**(12): p. 2913-2925.
127. Tornheim, J.A., et al., *Transcriptomic Profiles of Confirmed Pediatric Tuberculosis Patients and Household Contacts Identifies Active Tuberculosis, Infection, and Treatment Response Among Indian Children*. J Infect Dis, 2020. **221**(10): p. 1647-1658.
128. Vito, O., et al., *Novel RNA biomarkers improve discrimination of children with tuberculosis disease from those with non-TB pneumonia after in vitro stimulation*. Frontiers in Immunology, 2024. **Volume 15 - 2024**.
129. Schaaf, H.S. and B.J. Marais, *The Role of Xpert MTB/RIF Ultra in Diagnosing Pulmonary Tuberculosis in Children*. Am J Respir Crit Care Med, 2019. **200**(12): p. 1464-1465.
130. Pandey, D. and D. Ghosh, *Proteomics-based host-specific biomarkers for tuberculosis: The future of TB diagnosis*. Journal of Proteomics, 2024. **305**: p. 105245.
131. *Meso Scale Discovery*.
132. *Luminex*. Available from: https://us.diasorin.com/en/luminex.
133. Assarsson, E., et al., *Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability*. PLoS One, 2014. **9**(4): p. e95192.
134. Gold, L., et al., *Aptamer-based multiplexed proteomic technology for biomarker discovery*. PLoS One, 2010. **5**(12): p. e15004.

135. Li, J., et al., *Characterization of plasma proteins in children of different Mycobacterium tuberculosis infection status using label-free quantitative proteomics*. Oncotarget, 2017. **8**(61): p. 103290-103301.
136. Fossati, A., et al., *Plasma proteomics for biomarker discovery in childhood tuberculosis*. Nature Communications, 2025. **16**(1): p. 6657.
137. Kumar, N.P., et al., *Plasma chemokines as immune biomarkers for diagnosis of pediatric tuberculosis*. BMC Infectious Diseases, 2021. **21**(1): p. 1055.
138. Kumar, N.P., et al., *Discovery and Validation of a Three-Cytokine Plasma Signature as a Biomarker for Diagnosis of Pediatric Tuberculosis*. Frontiers in Immunology, 2021. **Volume 12 - 2021**.
139. Chiappini, E., et al., *Potential role of M. tuberculosis specific IFN-γ and IL-2 ELISPOT assays in discriminating children with active or latent tuberculosis*. PLoS One, 2012. **7**(9): p. e46041.
140. Tebruegge, M., et al., *Mycobacteria-Specific Cytokine Responses Detect Tuberculosis Infection and Distinguish Latent from Active Tuberculosis*. Am J Respir Crit Care Med, 2015. **192**(4): p. 485-99.
141. Lighter-Fisher, J., C.H. Peng, and D.B. Tse, *Cytokine responses to QuantiFERON® peptides, purified protein derivative and recombinant ESAT-6 in children with tuberculosis*. Int J Tuberc Lung Dis, 2010. **14**(12): p. 1548-55.
142. Gourgouillon, N., et al., *TNF-α/IL-2 ratio discriminates latent from active tuberculosis in immunocompetent children: a pilot study*. Pediatric Research, 2012. **72**(4): p. 370-374.
143. López-Suárez, A., et al., *Diagnostic potential of combining plasma biomarkers of tissue damage and inflammation in pediatric TB*. Journal of Microbiology, Immunology and Infection, 2024. **57**(6): p. 937-946.
144. Jenum, S., et al., *Approaching a diagnostic point-of-care test for pediatric tuberculosis through evaluation of immune biomarkers across the clinical disease spectrum*. Scientific Reports, 2016. **6**(1): p. 18520.
145. Togun, T., et al., *A three-marker protein biosignature distinguishes tuberculosis from other respiratory diseases in Gambian children*. EBioMedicine, 2020. **58**: p. 102909.
146. Sudbury, E.L., et al., *Mycobacterium tuberculosis-specific cytokine biomarkers for the diagnosis of childhood TB in a TB-endemic setting*. J Clin Tuberc Other Mycobact Dis, 2019. **16**: p. 100102.
147. Dhanasekaran, S., et al., *Identification of biomarkers for Mycobacterium tuberculosis infection and disease in BCG-vaccinated young children in Southern India*. Genes & Immunity, 2013. **14**(6): p. 356-364.
148. Pavan Kumar, N., et al., *Plasma Matrix Metalloproteinases Signature as Biomarkers for Pediatric Tuberculosis Diagnosis: A Prospective Case-Control Study*. Diseases, 2025. **13**(6).
149. Druszczynska, M., et al., *Cytokine Biosignature of Active and Latent Mycobacterium Tuberculosis Infection in Children*. Pathogens, 2021. **10**(5).
150. *Our Technology*. [cited 2026 July 2]; Available from: https://www.mesoscale.com/en/technical_resources/our_technology.

151. Ravindranath, M.H., et al., *Nature and Clonality of the Fluoresceinated Secondary Antibody in Luminex Multiplex Bead Assays Are Critical Factors for Reliable Monitoring of Serum HLA Antibody Levels in Patients for Donor Organ Selection, Desensitization Therapy, and Assessment of the Risk for Graft Loss*. J Immunol, 2017. **198**(11): p. 4524-4538.
152. Azadeh, M., et al., *Calibration Curves in Quantitative Ligand Binding Assays: Recommendations and Best Practices for Preparation, Design, and Editing of Calibration Curves*. Aaps j, 2017. **20**(1): p. 22.
153. Kelley, M., et al., *Large molecule run acceptance: Recommendation for best practices and harmonization from the Global Bioanalysis Consortium Harmonization Team*. Aaps j, 2014. **16**(2): p. 221-5.
154. Hornung, R.W. and L.D. Reed, *Estimation of average concentration in the presence of nondetectable values*. Applied occupational and environmental hygiene, 1990. **5**(1): p. 46-51.
155. Kagujje, M., et al., *The accuracy of point-of-care C-Reactive Protein as a screening test for tuberculosis in children*. PLOS Global Public Health, 2024. **4**(10): p. e0003725.
156. Mudrak, N., et al., *Incremental Yield of Multiple TB Tests in Children <5: a Retrospective Evaluation of WHO Concurrent Testing Recommendations*, in *The Union World Conference on Lung Health*. 2026.
157. Zar, H.J., et al., *Tuberculosis diagnosis in children using Xpert Ultra on different respiratory specimens*. American journal of respiratory and critical care medicine, 2019. **200**(12): p. 1531-1538.
158. Bai, X.-j., et al., *Cytokine and soluble adhesion molecule profiles and biomarkers for treatment monitoring in Re-treated smear-positive patients with pulmonary tuberculosis*. Cytokine, 2018. **108**: p. 9-16.
159. Chen, T., et al., *Profiling the human immune response to Mycobacterium tuberculosis by human cytokine array*. Tuberculosis, 2016. **97**: p. 108-117.
160. Mukae, H., et al., *Elevated levels of circulating adhesion molecules in patients with active pulmonary tuberculosis*. Respirology, 2003. **8**(3): p. 326-331.
161. Hamzaoui, A., et al., *Levels of soluble VCAM-1, soluble ICAM-1, and soluble E-selectin in patients with tuberculous pleuritis*. Mediators of inflammation, 1996. **5**(4): p. 276-279.
162. Lai, C.K., et al., *Circulating adhesion molecules in tuberculosis*. Clin Exp Immunol, 1993. **94**(3): p. 522-6.
163. Behr-Perst, S.I., et al., *Phenotypically Activated γδ T Lymphocytes in the Peripheral Blood of Patients with Tuberculosis*. The Journal of Infectious Diseases, 1999. **180**(1): p. 141-149.
164. Manyelo, M.C., *Evaluation of host biomarkers for early diagnosis of tuberculosis disease in children*. 2018, Stellenbosch: Stellenbosch University.
165. Du, S.-S., et al., *Screening for Differentially Expressed Proteins Relevant to the Differential Diagnosis of Sarcoidosis and Tuberculosis*. PLOS ONE, 2015. **10**(9): p. e0132466.

166. Shijubo, N., et al., *Elevated Concentrations of Circulating ICAM-1 in Far Advanced and Miliary Tuberculosis*. American Review of Respiratory Disease, 1993. **148**(5): p. 1298-1301.
167. Kim, M., et al., *The Changes of Serum Level of Tumor Necrosis Factor-Alpha, Gamma-Interferon and Soluble-Intercellular Adhesion Molecule-1 Relating to the Progression and Treatment of Patients with Pulmonary Tuberculosis*. Tuberculosis and Respiratory Diseases, 1998. **45**(6): p. 1167-1177.
168. Valiollahpour Amiri, M., et al., *SICAM-1 as a serum marker for follow-up of pulmonary tuberculosis therapy*. TANAFFOS (Respiration), 2004. **3**(3 (summer)): p. 55-63.
169. Bhalla, K., et al., *Host ICAMs play a role in cell invasion by Mycobacterium tuberculosis and Plasmodium falciparum*. Nature Communications, 2015. **6**(1): p. 6049.
170. Ramírez, G.M.L., et al., *Mycobacterium tuberculosis alters expression of adhesion molecules on monocytic cells*. Infection and Immunity, 1994. **62**(6): p. 2515-2520.
171. Schierloh, P., et al., *NK cell activity in tuberculosis is associated with impaired CD11a and ICAM-1 expression: a regulatory role of monocytes in NK activation*. Immunology, 2005. **116**(4): p. 541-552.
172. Bui, T.M., H.L. Wiesolek, and R. Sumagin, *ICAM-1: A master regulator of cellular responses in inflammation, injury resolution, and tumorigenesis*. J Leukoc Biol, 2020. **108**(3): p. 787-799.
173. Liu, S., et al., *ICAM-1 mediated cell-cell adhesion exerts dual roles on human B cell differentiation and IgG production*. iScience, 2023. **26**(12).
174. Sapoznikov, A., et al., *Dendritic cell ICAM-1 strengthens synapses with CD8 T cells but is not required for their early differentiation*. Cell Rep, 2023. **42**(8): p. 112864.
175. Champagne, B., et al., *Proteolytic Cleavage of ICAM-1 by Human Neutrophil Elastase*. The Journal of Immunology, 1998. **161**(11): p. 6398-6405.
176. Wakatsuki, T., et al., *A distinct mRNA encoding a soluble form of ICAM-1 molecule expressed in human tissues*. Cell adhesion and communication, 1995. **3**(4): p. 283-292.
177. Whiteman, S.C., et al., *Human Rhinovirus Selectively Modulates Membranous and Soluble Forms of Its Intercellular Adhesion Molecule–1 (ICAM-1) Receptor to Promote Epithelial Cell Infectivity **. Journal of Biological Chemistry, 2003. **278**(14): p. 11954-11961.
178. Dzikowska-Diduch, O., et al., *E-selectin and sICAM-1, biomarkers of endothelial function, predict recurrence of venous thromboembolism*. Thrombosis Research, 2017. **157**: p. 173-180.
179. Chen, C., et al., *The clinical significance of sICAM-1 in differentiating benign breast lesions from breast cancer*. Annals of Clinical & Laboratory Science, 2020. **50**(5): p. 650-656.
180. Moskalets, O., *Diagnostic and prognostic value of assessment of soluble adhesion molecules sICAM-1 and sVCAM-1 in septic patients*. Kazan medical journal, 2018. **99**(4): p. 645-650.

181. Seagrave, J., et al., *Effects of cigarette smoke exposure and cessation on inflammatory cells and matrix metalloproteinase activity in mice.* Exp Lung Res, 2004. **30**(1): p. 1-15.
182. Calfee, C.S., et al., *Soluble intercellular adhesion molecule-1 and clinical outcomes in patients with acute lung injury.* Intensive Care Med, 2009. **35**(2): p. 248-57.
183. Tenforde, M.W., et al., *C-reactive protein (CRP), interferon gamma-inducible protein 10 (IP-10), and lipopolysaccharide (LPS) are associated with risk of tuberculosis after initiation of antiretroviral therapy in resource-limited settings.* PLoS One, 2015. **10**(2): p. e0117424.
184. Yoon, C., et al., *Diagnostic accuracy of C-reactive protein for active pulmonary tuberculosis: a meta-analysis.* Int J Tuberc Lung Dis, 2017. **21**(9): p. 1013-1019.
185. Landry, A., et al., *Causes and outcomes of markedly elevated C-reactive protein levels.* Can Fam Physician, 2017. **63**(6): p. e316-e323.
186. Singh, P.P., *Interleukin-6: a potential biomarker of the success of tuberculosis treatment.* International Journal of Infectious Diseases, 2016. **45**: p. 413.
187. Gupte, A.N., et al., *Baseline IL-6 is a biomarker for unfavourable tuberculosis treatment outcomes: a multisite discovery and validation study.* European Respiratory Journal, 2022. **59**(4): p. 2100905.
188. Martinez, A.N., S. Mehra, and D. Kaushal, *Role of interleukin 6 in innate immunity to Mycobacterium tuberculosis infection.* J Infect Dis, 2013. **207**(8): p. 1253-61.
189. Pai, R.K., et al., *Inhibition of IFN-gamma-induced class II transactivator expression by a 19-kDa lipoprotein from Mycobacterium tuberculosis: a potential mechanism for immune evasion.* J Immunol, 2003. **171**(1): p. 175-84.
190. Yang, Q., et al., *IP-10 and MIG Are Compartmentalized at the Site of Disease during Pleural and Meningeal Tuberculosis and Are Decreased after Antituberculosis Treatment.* Clinical and Vaccine Immunology, 2014. **21**(12): p. 1635-1644.
191. Zhao, G., et al., *MCP-1/CCR2 axis regulates M2 macrophage polarization and immunosuppression in mycobacterium tuberculosis.* Front Immunol, 2025. **16**: p. 1698369.
192. Flores-Villanueva , P.O., et al., *A functional promoter polymorphism in monocyte chemoattractant protein–1 is associated with increased susceptibility to pulmonary tuberculosis.* Journal of Experimental Medicine, 2005. **202**(12): p. 1649-1658.
193. Dodd, P.J., et al., *The global burden of tuberculosis mortality in children: a mathematical modelling study.* Lancet Glob Health, 2017. **5**(9): p. e898-e906.
194. Zhang, M., M. Xue, and J.-q. He, *Diagnostic accuracy of the new Xpert MTB/RIF Ultra for tuberculosis disease: A preliminary systematic review and meta-analysis.* International Journal of Infectious Diseases, 2020. **90**: p. 35-45.
195. de Pablo, R., et al., *Circulating sICAM-1 and sE-Selectin as biomarker of infection and prognosis in patients with systemic inflammatory response syndrome.* European Journal of Internal Medicine, 2013. **24**(2): p. 132-138.
196. Creswell, J. and C. Colvin, *Closing the gaps in childhood TB detection.* IJTLD Open, 2025. **2**(2): p. 83-89.

197. Babo, Y., et al., *Comparison of Xpert MTB/RIF Ultra Results of Stool and Sputum in Children with Presumptive Tuberculosis in Southern Ethiopia*. Trop Med Infect Dis, 2023. **8**(7).
198. Fenn, B., et al., *Evaluation of Induced Sputum against Gastric Juice Aspirate in the Diagnosis of Tuberculosis in Children: A Cross-sectional Study.* JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH, 2024. **18**.
199. Zar, H.J., et al., *Induced sputum versus gastric lavage for microbiological confirmation of pulmonary tuberculosis in infants and young children: a prospective study.* The Lancet, 2005. **365**(9454): p. 130-134.
200. Kendig, E.L., et al., *Underreading of the Tuberculin Skin Test Reaction*. Chest, 1998. **113**(5): p. 1175-1177.
201. Smith, A., et al., *Holistic care in mild upper respiratory tract infections (MURTIs): an approach to individualized care management.* Drugs Context, 2025. **14**.
202. Sakurada, S., et al., *Inter-rater agreement in the assessment of abnormal chest X-ray findings for tuberculosis between two Asian countries.* BMC Infect Dis, 2012. **12**: p. 31.

# Appendices

**Appendix 1.** Case Report Forms (CRFs) used in the NOD-pedFEND study.

While the CRFs used in the NOD-pedFEND study are too numerous to be directly copied into this thesis, a full data dictionary has instead been made available online. Should this link become unusable, the author can be directly contacted to provide the dictionary at mudraknathan@gmail.com.

Link to access the data dictionary: https://bit.ly/4B1LygG or https://app.box.com/s/ft894yibpfomyng3v7j0ys50fbx5hims

**Appendix 2.** Genes measured by the NanoString CodeSet used for this analysis.

| Gene | Gene Type |
|---|---|
| *AAK1* | Endogenous |
| *ABCF1* | Housekeeping |
| *ACTB* | Housekeeping |
| *AGAP9* | Endogenous |
| *AIRE* | Endogenous |
| *ALAS1* | Housekeeping |
| *ANKRD22* | Endogenous |
| *APOL1* | Endogenous |
| *APOL4* | Endogenous |
| *APOL6* | Endogenous |
| *ARHGEF9* | Endogenous |
| *ASPHD2* | Endogenous |
| *ATP1A1* | Endogenous |
| *BATF2* | Endogenous |
| *BCL6* | Endogenous |
| *BLK* | Endogenous |
| *CCDC14* | Endogenous |
| *CCDC78* | Endogenous |
| *CCR6* | Endogenous |
| *CD1C* | Endogenous |
| *CD96* | Endogenous |
| *CDKN1C* | Endogenous |
| *CIRBP* | Endogenous |
| *CPVL* | Endogenous |
| *CTSA* | Endogenous |
| *CTSB* | Endogenous |
| *CYB561* | Endogenous |
| *CYP4F3* | Endogenous |
| *DARS2* | Endogenous |
| *DERA* | Endogenous |
| *DUSP3* | Endogenous |
| *EFCAB2* | Endogenous |
| *ENO3* | Endogenous |
| *ETV7* | Endogenous |
| *EXOC2* | Endogenous |
| *FCGR1A* | Endogenous |

| | |
|---|---|
| *FCGR1B* | Endogenous |
| *FCGR1C* | Endogenous |
| *FUT4* | Endogenous |
| *GAPDH* | Housekeeping |
| *GAS6* | Endogenous |
| *GBP1* | Endogenous |
| *GBP2* | Endogenous |
| *GBP4* | Endogenous |
| *GBP5* | Endogenous |
| *GSTA4* | Endogenous |
| *GUSB* | Housekeeping |
| *HK3* | Endogenous |
| *HM13* | Endogenous |
| *HPRT1* | Housekeeping |
| *ID3* | Endogenous |
| *IFI44L* | Endogenous |
| *IFITM3* | Endogenous |
| *IL-13* | Endogenous |
| *IL31RA* | Endogenous |
| *INSL3* | Endogenous |
| *KLF2* | Endogenous |
| *LTF* | Endogenous |
| *LUC7L* | Endogenous |
| *MDN1* | Endogenous |
| *MOB3C* | Endogenous |
| *NAGA* | Endogenous |
| *NEIL1* | Endogenous |
| *OAZ1* | Housekeeping, but excluded from use in normalization due to having a high CV with other housekeeping genes |
| *P2RY14* | Endogenous |
| *PGK1* | Housekeeping |
| *POLB* | Endogenous |
| *PPIA* | Housekeeping |
| *PRSS53* | Endogenous |
| *RAB13* | Endogenous |
| *RAB20* | Endogenous |
| *Rab33A* | Endogenous |
| *RBBP8* | Endogenous |
| *RPL13A* | Housekeeping |

| | |
|---|---|
| *RSAD2* | Endogenous |
| *SCARF1* | Endogenous |
| *SDHA* | Housekeeping |
| *SDR39U1* | Endogenous |
| *SEPT4* | Endogenous |
| *SERPING1* | Endogenous |
| *SH2D1B* | Endogenous |
| *SLC3A1* | Endogenous |
| *SLPI* | Endogenous |
| *SMARCD3* | Endogenous |
| *SMCO4* | Endogenous |
| *SNRNP70* | Endogenous |
| *SPDYE5* | Endogenous |
| *SPSB1* | Endogenous |
| *SRBD1* | Endogenous |
| *STAT1* | Endogenous |
| *TAP1* | Endogenous |
| *TBP* | Housekeeping |
| *TCN2* | Endogenous |
| *TIMM10* | Endogenous |
| *TRAFD1* | Endogenous |
| *TRMT2A* | Endogenous |
| *TUBGCP6* | Endogenous |
| *WARS* | Endogenous |
| *ZNF202* | Endogenous |
| *ZNF419* | Endogenous |
| *ZNRF1* | Endogenous |

**Appendix 3.** Mean QC values for the standards, for each assay, which fell above the LLOQ.

| Assay | Mean % Recovery of Standards above LLOQ | IQR of Mean % Recovery | Mean % CV of Standards above LLOQ | IQR of % CV |
|---|---|---|---|---|
| C-Peptide | 100.3 | 10.0 | 8.5 | 7.6 |
| CRP | 100.0 | 6.7 | 7.8 | 3.7 |
| Eotaxin | 98.0 | 3.9 | 6.6 | 5.1 |
| FGF (basic) | 100.8 | 5.7 | 8.7 | 3.5 |
| Flt-1 | 101.6 | 8.6 | 2.6 | 2.7 |
| GIP | 100.1 | 3.0 | 7.2 | 6.8 |
| GM-CSF | 100.5 | 4.2 | 5.4 | 3.0 |
| Glucagon | 102.3 | 11.3 | 7.1 | 4.0 |
| IFN-γ | 99.1 | 8.1 | 9.2 | 10.4 |
| IL-10 | 100.5 | 4.5 | 8.3 | 6.5 |
| IL-12/IL-23p40 | 100.7 | 4.5 | 6.5 | 4.2 |
| IL-12p70 | 100.3 | 4.0 | 11.4 | 2.7 |
| IL-13 | 100.0 | 0.1 | 5.6 | 6.7 |
| IL-15 | 98.4 | 6.8 | 8.9 | 6.7 |
| IL-16 | 102.8 | 17.2 | 6.2 | 6.0 |
| IL-17A | 105.0 | 21.8 | 5.3 | 3.7 |
| IL-17A/F | 99.7 | 5.0 | 6.7 | 4.5 |
| IL-17B | 101.2 | 10.6 | 9.0 | 3.7 |
| IL-17C | 99.6 | 5.3 | 7.1 | 6.4 |
| IL-17D | 100.2 | 3.4 | 11.1 | 9.5 |
| IL-1RA | 99.3 | 1.8 | 5.6 | 5.6 |
| IL-1α | 100.5 | 14.5 | 9.5 | 5.8 |
| IL-1β | 100.3 | 3.8 | 10.4 | 8.9 |
| IL-2 | 99.5 | 2.5 | 4.6 | 3.2 |
| IL-3 | 100.9 | 9.0 | 9.6 | 6.2 |
| IL-4 | 100.4 | 3.3 | 16.2 | 11.7 |
| IL-5 | 100.5 | 1.0 | 6.7 | 6.8 |
| IL-6 | 100.3 | 1.4 | 8.9 | 3.8 |
| IL-7 | 102.6 | 14.4 | 5.5 | 3.6 |
| IL-8 | 99.4 | 3.9 | 8.9 | 4.1 |

| | | | | |
|---|---|---|---|---|
| IL-9 | 104.0 | 18.5 | 8.4 | 8.1 |
| IP-10 | 101.1 | 11.3 | 14.9 | 4.2 |
| Insulin | 100.6 | 6.8 | 6.6 | 8.5 |
| Leptin | 98.5 | 8.1 | 7.6 | 7.2 |
| MCP-1 | 100.7 | 4.0 | 6.4 | 7.0 |
| MCP-4 | 102.3 | 16.5 | 6.1 | 3.4 |
| MDC | 100.1 | 9.0 | 5.4 | 7.2 |
| MIP-1α | 100.2 | 1.4 | 4.8 | 2.4 |
| MIP-1β | 100.3 | 0.7 | 7.6 | 15.4 |
| PIGF | 101.1 | 5.5 | 5.2 | 1.5 |
| PP | 100.3 | 3.3 | 7.7 | 5.7 |
| SAA | 101.9 | 11.5 | 8.1 | 7.4 |
| TARC | 99.7 | 2.8 | 4.5 | 4.1 |
| TNF-α | 99.3 | 2.6 | 7.9 | 1.4 |
| TNF-β | 102.0 | 10.5 | 5.8 | 6.3 |
| TSLP | 99.9 | 11.3 | 8.7 | 13.5 |
| Tie-2 | 102.6 | 8.4 | 8.1 | 11.1 |
| VEGF-A | 101.8 | 13.2 | 4.8 | 3.4 |
| VEGF-C | 101.6 | 14.5 | 4.8 | 2.9 |
| VEGF-D | 101.4 | 13.9 | 5.9 | 7.3 |
| sICAM-1 | 102.0 | 11.7 | 11.0 | 10.9 |
| sVCAM-1 | 100.4 | 6.3 | 7.2 | 6.4 |