

Validation of New Technologies for the Diagnostic Evaluation of Active Tuberculosis (VANTDET)

Alice Halliday,¹ Pooja Jain,^{1,2\$} Long Hoang,^{1\$} Robert Parker,^{1\$} Mica Tolosa-Wright,^{1\$} Tereza Masonou,¹ Nathan Green,^{3,4,5,6} Aime Boakye,^{1,2} Yemisi Takwoingi,⁹ Shea Hamilton,⁷ Vinay Mandagere,¹ Anastasia Fries,¹ Lachlan Coin,⁸ Jon Deeks,^{10,11} Peter White,^{3,4,5,6} Michael Levin,⁷ Peter Beverley,¹ Onn Min Kon,^{1,2,12} Ajit Lalvani.^{1,2*}

* Corresponding author. Email: a.lalvani@imperial.ac.uk, phone: 020 7594 0883 (address below)

\$ These authors contributed equally

Author order: please note that this final author order has not yet been finalised and agreed by all authors.

Author affiliations:

¹ TB Research Centre, National Heart and Lung Institute, Imperial College London, UK

² National Institute for Health Research (NIHR) Health Protection Research Unit in Respiratory Infection, Imperial College London, UK

³ NIHR Health Protection Research Unit in Modelling Methodology, Imperial College London, UK

⁴ Medical Research Council (MRC) Centre for Outbreak Analysis and Modelling, Imperial College London, UK

⁵ Modelling and Economics Unit, National Infection Service, Public Health England, UK

⁶ Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

⁷ Paediatric Infectious diseases group, Division of Medicine, Imperial College London, London, United Kingdom

⁸ Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia

⁹ Institute of Applied Health Research, University of Birmingham, Birmingham, UK

¹⁰ Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham, Birmingham B15 2TT

¹¹ NIHR Birmingham BRC, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, B15 2TT

¹² St Mary's Hospital, Imperial College Healthcare Trust, London, UK

Keywords: Tuberculosis, *Mycobacterium tuberculosis*, Diagnostic test, Transcriptomic, Proteomics, cellular immune, signatures, validation,

1

2 **Declared competing interests of authors:**

3 Dr. Halliday has a patent “A cellular immune signature for risk stratification of latent
4 tuberculosis infection” pending. Prof Lalvani reports issued patents underpinning IGRA and
5 next-generation IGRA some of which were assigned by the University of Oxford to Oxford
6 Immunotec plc resulting in royalty entitlements for the University of Oxford and Prof.
7 Lalvani. Prof Lalvani is also inventor of issued and pending unlicensed patents underpinning
8 flow-cytometric diagnosis of TB. Prof Deeks reports grants from NIHR during the conduct of
9 the study; and Receipt of NIHR Senior Investigator Emeritus award and supported by the
10 NIHR Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust
11 and University of Birmingham. Prof Kon is chairperson of the UK Joint Tuberculosis
12 Committee. Dr White reports grants from Medical Research Council, grants from National
13 Institute for Health Research, grants from National Institute for Health Research, during the
14 conduct of the study; grants from Otsuka, outside the submitted work. .. Dr. Parker reports
15 having a patent 1719853.2 pending, and a patent 2017904359 pending. Dr. Coin reports
16 having a patent WO2014067943A1 pending, a patent US20150284780A1 pending, and a
17 patent EP2914740B1 issued.

18

Abstract

Background: Tuberculosis (TB) is a devastating disease for which new diagnostic tests are desperately needed.

Objective: To validate promising new technologies (namely whole blood transcriptomics, proteomics, flow cytometry and qRT-PCR) and existing signatures for detection of active TB in samples obtained from individuals suspected of active TB.

Design: Four sub-studies, each of which used the samples from biobank collected as part of the IDEA study, which was a prospective cohort of patients recruited with suspected TB.

Setting: secondary care

Participants: Adults (aged ≥ 16 years old) presenting as inpatients or outpatients at 12 NHS hospital trusts in London, Slough, Oxford, Leicester and Birmingham with suspected active TB.

Interventions: New tests using either: genome-wide gene expression microarray (transcriptomics); SELDI TOF/ LC-MS (proteomics), flow cytometry, qRT-PCR.

Main outcome measures: Area under the curve (AUC), sensitivity and specificity, were calculated to determine diagnostic accuracy. Positive and negative predictive values were calculated in some cases. A decision tree model was developed to calculate the incremental costs and quality-adjusted life-years (QALYs) of changing from current practice to using the novels tests.

Results: The project and 4 sub-studies which assessed the previous published signatures measured using each of the new technologies, and a health economic analysis where the best performing tests were evaluated for cost effectiveness. The diagnostic accuracy of the transcriptomic tests ranged from AUC=0.81-0.84 for detecting all TB in our cohort. The performance for detecting culture confirmed TB or pulmonary TB (PTB) was better than for highly probable TB or extrapulmonary TB (EPTB) respectively, but not high enough to be clinically useful. None of the previously described serum proteomic signatures for active TB provided good diagnostic accuracy, not did the candidate rule-out tests. Four of six previously described cellular immune signatures provided a reasonable level of diagnostic accuracy (AUC = 0.78-0.92) for discriminating all TB from those with other disease (OD)

1 and latent TB infection (LTBI) in HIV- TB suspects. Two of these assays may be useful in
2 the IGRA+ population and can provide high positive predictive value (PPV). None of the
3 new tests for TB can be considered cost effective.

4 **Limitations:** The diagnostic performance of new tests within the HIV+ population was
5 either underpowered or not sufficiently achieved in each sub-study.

6 **Conclusions:** Overall, the diagnostic performance of all previously identified ‘signatures’ of
7 TB was lower than previously reported. This likely reflects the nature of the cohort we used,
8 which includes the harder to diagnose groups, such as culture unconfirmed TB, and EPTB,
9 which were underrepresented in previous cohorts.

10 **Future work:** We are yet to perform our secondary objective of deriving novel signatures of
11 TB using our datasets. This was beyond the scope of this report. We recommend that future
12 studies using these technologies target specific sub-types of TB, specifically those groups
13 where new diagnostic tests are required.

14 **Funding details:** This project was funded by the National Institute for Health Research
15 (NIHR) and Medical Research Council (MRC) Efficacy and Mechanism Evaluation
16 programme and the NIHR Health Protection Research Unit in Respiratory Infections,
17 Imperial College London, UK.

1	Contents	
2		
3	Abstract	3
4	Contents	5
5	Table of tables.....	11
6	Table of figures	16
7	List of abbreviations	19
8	Plan English Summary.....	21
9	Scientific Summary.....	22
10	Chapter 1. Introduction.....	27
11	Background: Tuberculosis	27
12	Current diagnostic tests for TB	28
13	Clinical unmet needs.....	30
14	Rationale for VANTDET	31
15	Aim	33
16	Research Objectives.....	33
17	Primary objectives:	33
18	Secondary objectives:	33
19	Chapter 2. Study overview, management and general methodology	35
20	Study management structure.....	35
21	Overview of study design	35
22	General management and task allocation	36
23	Management and oversight meetings.....	37
24	Issues and problems	37
25	Changes to study protocol and management plan.....	37
26	Study cohorts	39
27	Reference standard.....	40

1	Patient and public engagement and involvement.....	42
2	Statistical analysis and presentation of data.....	43
3	Chapter 3. Transcriptomics.....	45
4	Introduction.....	45
5	Methods.....	46
6	Inclusion criteria	46
7	RNA extraction and quality control	46
8	Whole genome-wide microarray.....	46
9	Microarray QC and normalization	46
10	Statistical analysis	47
11	Results.....	49
12	Final cohort	49
13	Diagnostic accuracy analysis	51
14	Discussion	59
15	Chapter 4. Proteomics.....	62
16	Introduction.....	62
17	Proteomic biomarkers – current methodologies	62
18	Previously published signatures.....	63
19	Aim	65
20	Objectives	65
21	Methods.....	66
22	Cohort selection	66
23	Serum samples	66
24	SELDI-TOF	66
25	LC-MS	67
26	Data processing.....	67
27	Statistical analysis.....	68

1	Results.....	68
2	Cohort characteristics.....	70
3	Bridging study.....	74
4	Performance of previously published signatures using LC-MS	74
5	Discussion	78
6	Chapter 5. Cellular immune signatures	81
7	Introduction.....	81
8	Aim	82
9	Objectives	83
10	Methods.....	83
11	Cohort selection	83
12	Experimental design.....	83
13	Use of PPD-responsive control.....	85
14	Analysis and gating.....	86
15	Determining the inclusion criteria for antigen responsiveness	86
16	Statistical analysis	87
17	Results.....	87
18	Cohort selection	87
19	Cohort characteristics.....	88
20	Recovery and detection of antigen specific responses.....	89
21	Diagnostic performance of signatures in full cohort.....	90
22	Diagnostic performance in key sub-groups of TB	95
23	Discussion	99
24	Chapter 6. Candidate rule-out tests.....	102
25	Introduction.....	102
26	Additional Mtb antigens	102
27	Detection of additional or alternative analytes	103

1	Rationale for candidate rule-out assays	104
2	Aim	104
3	Primary objective:	104
4	Secondary objectives:	105
5	Methods.....	105
6	Cohort selection	105
7	Experimental set-up for molecular rule-out test	106
8	Detection of candidate biomarkers from Quantiferon supernatants	107
9	Results.....	107
10	Cohort characteristics.....	107
11	Validation of molecular rule out tests	109
12	Performance of CXCL10 and IFN γ in QFT-GIT supernatants using MSD	113
13	Discussion	114
14	Chapter 7. Health Economic Analysis.....	116
15	Introduction.....	116
16	Methods.....	116
17	Decision tree model	116
18	Estimation of costs and health impact used in the model	118
19	Selection of tests to evaluate in health economic analysis	119
20	Results.....	120
21	Discussion	125
22	Recommendations.....	126
23	Chapter 8. General Discussion	127
24	Principle findings	127
25	Strengths and limitations.....	128
26	Recommendations for clinical practice.....	129
27	Recommendations for future research	130

1	Conclusion	131
2	Acknowledgements.....	133
3	Independent Scientific Advisory Board.....	133
4	Members of the Study Management Group.....	134
5	Members of Wider Study Management Group.....	134
6	Contribution of authors	134
7	Data Sharing Statement.....	136
8	Funding	136
9	References.....	137
10	Appendix 1 Prevalence data from the IDEA study.....	142
11	Appendix 2 Distribution of disease types in the transcriptomic validation cohort.....	143
12	Appendix 3 Additional information on transcriptomic signatures used for validation	
13	studies.	145
14	Appendix 4 Additional analyses for validation of transcriptomic ‘score’ signatures.....	146
15	Appendix 5 Additional analyses for validation of transcriptomic signatures using LDA	
16		153
17	Appendix 6 Supporting data for proteomic signature validation	158
18	Analysis for bridging study.....	158
19	Abundance of signature proteins	160
20	Missing and imputed data	163
21	Appendix 7 Additional analyses for diagnostic evaluation of serum protein signatures for	
22	detection of active TB using LDA	165
23	Appendix 8 Validation of published proteomic signatures using LC-MS by LDA, with the	
24	removal of Sarcoidosis.....	166
25	Appendix 9 Additional information for flow cytometry introduction and methods.....	167
26	Appendix 10 Additional cohort characteristics of the validation of cellular immune	
27	signatures sub-study.....	171
28	Appendix 11 Viability, cell recovery and cytokine responses to antigens	172

1	Recovery and viability	172
2	Responses to antigen stimulations	172
3	Appendix 12 Additional analyses for the validation of cellular immune signatures for the	
4	diagnostic evaluation of active TB	175
5	Appendix 13 Inclusion criteria of cellular immune signatures from previous publications	
6		179
7	Appendix 14 Performance of signatures after Mtb RD-1-related antigen stimulation	180
8	Performance of cellular immune signatures using Mtb-specific RD1-related antigens	180
9	Correlation between Mtb RD-1-related and PPD antigen responses	181
10	Performance of signatures after ECRR stimulation	181
11	Appendix 15 Additional cohort characteristics for the molecular rule-out test assessment	
12		184
13	Appendix 16 Additional data for the validation of candidate rule out tests	185
14	Appendix 17 Additional tables for health economic analysis – methodology	192
15	Distributional formulation of individual level/sample uncertainty	194
16	Use as rule-out test	195
17	Use of dual testing as rule-in test	196
18		

Table of tables

TABLE 1.	Details of Technologies to be assessed in VANTDET.....	31
TABLE 2.	Dosanjh categorisation for the diagnosis of active TB suspects, as from. ¹⁰	40
TABLE 3.	Demographic characteristics of transcriptomic validation cohort (n=628) of patients selected from the IDEA study cohort.	50
TABLE 4.	Clinical characteristics of transcriptomics validation cohort, stratified by patient final diagnosis.	51
TABLE 5.	Performance of published gene expression score signatures in full cohort, stratified by HIV infection status.....	53
TABLE 6.	Performance of score signatures in the detection of sub-groups of TB cases, within the smear negative population of TB suspects.	56
TABLE 7.	Summary of publications presenting proteomic signatures for detecting of active TB cases.	63
TABLE 8.	Demographic and clinical characteristics of the LC-MS proteomic signature validation cohort.	71
TABLE 9.	Demographic and clinical characteristics of the n=86 participants used for validation of the 4 protein SELDI-TOF signature.	72
TABLE 10.	Demographic characteristics of final cohort for validation of cellular immune signatures sub-study (n=184).....	88
TABLE 11.	Clinical characteristics of the validation of cellular immune signature sub-study cohort (n=184).....	89
TABLE 12.	Performance of published cellular immune signature in diagnosing all TB, stratified by HIV infection status.....	93
TABLE 13.	Performance of cellular immune signatures in detecting TB, stratified by Dosanjh criteria.....	96
TABLE 14.	The performance of the %HLA-DR and %T _{EFF} signatures in detecting sub-groups of TB patients in the smear negative patients, within the IGRA positive cohort used in the validation of flow cytometric signatures sub-study.	98

1	TABLE 15.	Demographic and clinical characteristics for GO/NO GO cohort for	
2		assessment of candidate molecular rule-out test for active TB.....	108
3	TABLE 16.	qPCR detection of CXCL10, CXCL9, and BATF2 in response to Mtb	
4		antigens or positive control in the GO/NO GO cohort of patients selected for validation of	
5		candidate molecular rule-out tests.	109
6	TABLE 17.	Key model parameters.....	117
7	TABLE 18.	Diagnostic test performance and distributions for different tests.	120
8	TABLE 19.	Cost-effectiveness statistics for transcriptomic rule-out scenarios	121
9	TABLE 20.	Cost-effectiveness statistics for transcriptomic rule-in scenarios.	122
10	TABLE 21.	Cost-effectiveness statistics for flow cytometry rule-in scenario.	123
11	TABLE 22.	: Cost-effectiveness statistics for molecular rule-out scenarios	124
12	TABLE 23.	Calculated prevalence data for TB populations, using the IDEA study data....	
13			142
14	TABLE 24.	Distribution of severity and disease in the TB group.....	143
15	TABLE 25.	Distribution of alternative diagnoses in the OD group	144
16	TABLE 26.	Summary of previously identified signatures, including cohort characteristics	
17		and measures of diagnostic performance.	145
18	TABLE 27.	Cross tabulation of previously published transcriptomic score signature	
19		results by the reference standard. The score signatures being assessed are: Anderson ²⁸ ;	
20		Kaforou ¹⁹ ; Sweeney ³¹	146
21	TABLE 28.	Performance of score signatures based on Dosanjh category in the full cohort.	
22		The score signatures being assessed are: Anderson ²⁸ ; Kaforou ¹⁹ ; Sweeney ³¹	149
23	TABLE 29.	Performance of score signatures based on Dosanjh category, HIV- cohort	
24		only/ The score signatures being assessed are: Anderson ²⁸ ; Kaforou ¹⁹ ; Sweeney ³¹	150
25	TABLE 30.	Performance of all previously published transcriptomic signatures in full	
26		cohort using LDA	155
27	TABLE 31.	Performance of all transcriptomic signatures in detecting either culture-	
28		confirmed TB (Dosanjh 1) or culture-unconfirmed TB (Dosanjh 2) using LDA according to	
29		Dosanjh criteria.	157

1	TABLE 32.	Diagnostic performance of previously published signatures for detection of	
2		PTB or EPTB, by LDA.....	157
3	TABLE 33.	Details of proteins within the SELDI-TOF Signature discovered by Hamilton	
4		and Levin et al.	158
5	TABLE 34.	Correlation matrix showing r and p-value (Pearson's) for SELDI-TOF	
6		normalised intensity and LC-MS normalised ratio for protein 1-3 in the Levin Signature...	158
7	TABLE 35.	Abundance of signature proteins in full cohort, and comparisons between key	
8		patient groups.	160
9	TABLE 36.	Performance of LC-MS technology in detecting proteins from key previously	
10		published signatures, including the proportion of missing data in the cohorts.....	163
11	TABLE 37.	Diagnostic performance of Song ⁴⁴ signature in the validation of TB in the full	
12		LC-MS validation cohort (n=168) and in key sub-groups of TB.	164
13	TABLE 38.	Diagnostic performance of serum proteomic signatures in full cohort n=166...	
14		165	
15	TABLE 39.	Diagnostic performance of Xu ⁴⁵ , Levin (unpublished) and Achkar ²⁰ serum	
16		proteomic signatures in key patient sub-groups in full cohort, using LDA.....	165
17	TABLE 40.	Performance of Xu ⁴⁵ , Levin (unpublished) and Achkar ²⁰ signatures for	
18		detecting all TB, using LDA, with patients with Sarcoidosis removed.....	166
19	TABLE 41.	Performance of Xu ⁴⁵ , Levin (unpublished) and Achkar ²⁰ signatures for	
20		detecting key subgroups of TB, using LDA, with Sarcoidosis removed.....	166
21	TABLE 42.	Summary of cellular immune signatures to be validated, with reference to the	
22		original publication and the study design and characteristics of the discovery cohorts.	167
23	TABLE 43.	Details of antibodies used and their final dilutions during staining.	167
24	TABLE 44.	Distribution of severity and disease in TB group of cellular immune signature	
25		validation cohort.	171
26	TABLE 45.	Distribution of alternative diagnoses in OD group (n=92) in the validation	
27		cohort for cellular immune signatures.	171
28	TABLE 46.	Responses to PMA/ionomycin in full cohort.	172
29	TABLE 47.	Responses to PPD, stratified by Dosanjh criteria.....	173
30	TABLE 48.	Responses to PPD, stratified by patient group and HIV infection status.....	174

1	TABLE 49.	Details of samples with positive CD4+ and CD8+ T cell cytokine responses	
2		to Mtb antigens.	174
3	TABLE 50.	Responses to Mtb antigens, stratified by patient group and HIV infection	
4		status.	174
5	TABLE 51.	Performance of cellular immune signatures for the detection of TB vs OD	
6		when Dosanjh 4 is further stratified by risk of latent TB.	176
7	TABLE 52.	Performance of cellular immune signatures for diagnosing either PTB or	
8		EPTB.	177
9	TABLE 53.	Inclusion criteria for ‘responders’ and diagnostic performance analysis from	
10		previous publications.	179
11	TABLE 54.	Performance of published cellular immune signatures in the diagnosis of all	
12		active TB, stratified by HIV infection status.	182
13	TABLE 55.	Distribution of disease types and severity within the TB patients in the	
14		GO/NO GO cohort for the validation of the molecular rule out tests.....	184
15	TABLE 56.	Distribution of other diseases in the OD patients within the TB patients in the	
16		GO/NO GO cohort for the validation of the molecular rule out tests.....	184
17	TABLE 57.	Measurment of IFN γ transcripts in stimulated PBMC samples using qRT-	
18		PCR.	186
19	TABLE 58.	Diagnostic accuracy data for candidate molecular rule out test: all TB vs all	
20		other disease.	187
21	TABLE 59.	Diagnostic accuracy data for candidate molecular rule out test: all TB vs	
22		OTHER DISEASE T-SPOT-. AUC, sensitivity and specificity values are shown, with 95%	
23		CI, and corresponding cut-offs.	187
24	TABLE 60.	Diagnostic accuracy data for candidate molecular rule out test: TB TSPOT+	
25		vs OTHER DISEASE TSPOT-.....	188
26	TABLE 61.	Diagnostic accuracy data for candidate molecular rule out test: ATB T-SPOT	
27		+ vs OD T-SPOT +.	189
28	TABLE 62.	Performance of IFN γ and CXCL10 analytes from Quantiferon supernatants	
29		detected by MSD for the detection of all TB in the GO/NO GO cohort.	190
30	TABLE 63.	Health impact of adverse event due to TB treatment.All values from ⁶³	192

1	TABLE 64.	Time to obtain a result and cost of new tests.	192
2	TABLE 65.	Health care professional consultation visit monetary costs incurred. SE,	
3	standard error		192
4	TABLE 66.	Test and sampling procedure costs for cost-effectiveness calculations. SE,	
5	standard error		193
6	TABLE 67.	Active TB treatment costs for the first 60 days. Final cost is the sum of all	
7	listed 60-day total drug costs.		194
8			
9			

Table of figures

FIGURE 1.	Proposed VANTDET study design, and link with the IDEA study.....	36
FIGURE 2.	Study flow chart for validation of Transcriptomics signatures.	49
FIGURE 3.	The performance of previously published score signatures, stratified by Dosanjh category.	54
FIGURE 4.	Flow diagram for the proteomics validation sub-study.....	70
FIGURE 5.	Performance of 4 protein signature (Hamilton and Levin) on SELDI-TOF, for the classification of TB in the VANTDET SELDI TOF sub-study cohort (n=86).....	73
FIGURE 6.	The diagnostic performance of the Xu, Levin and Achkar protein signatures of, using LDA, on the LC-MS depleted dataset (n=166).	76
FIGURE 7.	Study flow chart of patients included in the validation of cellular immune signatures sub-study.....	88
FIGURE 8.	Performance of published signatures in full cohort, stratified by HIV infection status.	92
FIGURE 9.	Study flow diagram for the validation of the candidate rule-out tests.	108
FIGURE 10.	Diagnostic performance of candidate qRT-PCR detection of CXCL10, CXCL9, BATF2, or IFN γ in the candidate molecular rule-out as assessed in GO/NO GO cohort.	111
FIGURE 11.	Decision tree comparing current practice (“No rule-out test”) with a diagnostic pathway incorporating an initial rule-out test (“Rule-out test”).	118
FIGURE 12.	Cost-effectiveness planes for status-quo against transcriptomic rule-out test scenario, with contour lines.	121
FIGURE 13.	Cost-effectiveness planes for status-quo against transcriptomic rule-in test scenario, with contour lines (right).	122
FIGURE 14.	Cost-effectiveness planes for status-quo against flow cytometry %HLA-DR rule-in test scenario, with contour lines.	123
FIGURE 15.	Cost-effectiveness planes for status-quo against molecular rule-out test scenario, with contour lines.	124

1	FIGURE 16. Performance of published ‘score’ based signatures in the full cohort, stratified	
2	by HIV infection status.	147
3	FIGURE 17. Performance of published SCORE signatures by Dosanjh category in the	
4	HIV- population.	150
5	FIGURE 18. Performance of score signatures in PTB and EPTB, and in OD stratified by	
6	disease types.	151
7	FIGURE 19. Validation of published transcriptomic signatures using LDA in full cohort	
8	(A), and stratified by HIV infection status (B).	155
9	FIGURE 20. Comparison of three proteins from Levin signature with LC-MS data.	159
10	FIGURE 21. Abundance of SERPINA1 (alpha-1-antitrypsin or P01009).	162
11	FIGURE 22. Initial basic gating strategies for the cellular immunological signatures as set	
12	on FlowJo version 10.	169
13	FIGURE 23. Gating strategies for the validating signatures exploring IFN γ + cytokine	
14	responses within the CD4+ population.	169
15	FIGURE 24. Gating strategy for validating Pollock ²⁴ et al and Harari ²³ et al signatures.	170
16	FIGURE 25. The effect of PMA/ionomycin on viability of lymphocytes (A) and CD4+	
17	surface expression (B).	173
18	FIGURE 26. Performance of published cellular immune signatures according to Dosanjh	
19	classification (HIV- patients only).	175
20	FIGURE 27. Performance of cellular immune signatures in different TB and OD/LTBI	
21	disease types.	177
22	FIGURE 28. Performance of published signatures after stimulation with Mtb antigens.	180
23	FIGURE 29. Correlation of %HLA-DR signature using PPD and RD-1 (ECRR) antigens. ...	
24		181
25	FIGURE 30. Relative abundance of CXCL10, CXCL9, IFN γ and BATF2 in key patient	
26	groups within the GO, NO/GO cohort for assessment of the candidate molecular rule-out test.	
27		185
28	FIGURE 31. Selection of cut-off for target genes in response to IFN γ positive control	
29	condition.	186

1	FIGURE 32. Diagnostic performance of CXCL10, CXCL9, IFN γ and BATF2 Mtb-	
2	specific gene expression in all TB vs TSPOT- OD.	187
3	FIGURE 33. Diagnostic performance of CXCL10, CXCL9, IFN γ and BATF2 Mtb-	
4	specific gene expression in TSPOT+ TB vs TSPOT- OD.	189
5	FIGURE 34. The performance of qRT-PCR detection of CXCL10, CXCL9, IFN γ , BATF2	
6	and T-SPOT for discriminating between T-SPOT+ TB cases (n=19-24) and T-SPOT +	
7	OD/LTBI cases (n=10-22).	189
8	FIGURE 35. Comparison of detection of IFN γ or CXCL10 in Mtb-antigen stimulated	
9	samples, using either MSD from QFT supernatants (A) of qRT-PCR from stimulated PBMCs	
10	(B) and comparing between patient groups in the overlapping cohort on n=75 where both	
11	tests were performed.	191
12	FIGURE 36. Decision tree comparing current practice (“No rule-in test”) with a diagnostic	
13	pathway incorporating an initial rule-in test (“Rule-in test”).	196
14	FIGURE 37. Decision tree comparing current practice (“No rule-in test”) with a diagnostic	
15	pathway incorporating an initial dual rule-in test (“Dual rule-in test”).	197
16		
17		

1 List of abbreviations

2

AIDS	Acquired Immuno Deficiency Syndrome
AUC	Area Under the Curve
BATF2	Basic Leucine Zipper ATF-Like Transcription Factor 2
BCG	Bacille Calmette-Guérin
BMI	Body Mass Index
CD4	Cluster of Differentiation IV
cDNA	Complementary Deoxyribonucleic acid
CI	Confidence Interval
CLC	Charcot-Leyden Crystal Galectin
CXCL10	Chemokine (C-X-C motif) ligand 10
CXCL9	Chemokine (C-X-C motif) ligand 9
DRS	Disease Risk Score
DUSP3	Dual Specificity Phosphatase 3
ELISA	Enzyme-Linked Immunoabsorbent Assay
EPTB	Extra Pulmonary Tuberculosis
EU	European Union
FACS	Fluorescence-Activated Cell Sorting
FBS	Foetal Bovine Serum
GBP1	Guanylate Binding Protein 1
GBP5	Guanylate Binding Protein 5
GCP	Good Clinical Practice
HC	Healthy Control
HIV	Human Immunodeficiency Virus
HLA-DR	Human Leukocyte Antigen – antigen D Related
HPRU	Health Protection Research Unit
HTA	Health Technology Assessment
ID3	Inhibitor of DNA binding 3
IFITM3	Interferon Induced Transmembrane Protein 3
IFN γ	Interferon Gamma
IGJ	Immuno Globulin J chain
IGRA	Interferon Gamma Release Assay
IMAC	Immobilised Metal Affinity
IP-10	Interferon gamma-induced protein-10
KLF2	Kruppel Like Factor 2
LC-MS	Liquid Chromatography-Mass Spectrometry
LDA	Linear Discriminant Analysis

LTBI	Latent Tuberculosis Infection
MDR-TB	Multi Drug Resistant TB
Mtb	Mycobacterium tuberculosis
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute of Health Research
NiR	Near InfraRed
NTB	Non-Tuberculosis Mycobacteria (NTM)
OD	Other Diseases
ORD	Other Related Diseases
P2Ry14	Purinergic Receptor P2Y14
PBMCs	Peripheral Blood Mononuclear Cells
PBS	Phosphate Buffered Saline
PERC	Patient Experience Research Centre
PPD	Purified Protein Derivative
PPI	Public and Patient Involvement
PTB	Pulmonary TB
QC	Quality Control
QFT-GIT	QuantiFERON GOLD In-Tube
qRT-PCR	Quantitative Reverse Transcription-Polymerase Chain Reaction
RD1	Region of Difference 1
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristic
SELDI-TOF	Surface-Enhanced Laser Desorption Ionisation Time-Of-Flight
SERPINA1	Serpin family A member 1
SMG	Study Management Group
SSA	Site Specific Assessment
SSC	Study Steering Committee
STARD	Standards for Reporting of Diagnostic accuracy studies
SVML	Support Vector Machine Learning
TB	Tuberculosis
TBS	TB Score
TST	Tuberculin Skin Test
UK	United Kingdom
VANTDET	Validation of New Technologies for the Diagnostic Evaluation of active Tuberculosis
WHO	World Health Organisation

Plan English Summary

Globally Tuberculosis (TB) is the most deadly infectious disease, causing 10 million cases, and 1.3 million deaths each year. TB is caused by bacteria which is transmitted from an individual with TB disease of the lungs to another person by coughing. There are several diagnostic tests for TB, most of which detect that presence of the bacteria in clinical samples. However, these tests fail to detect all TB patients, particularly when there is a low number of bacteria present at the site of disease, and/or it is difficult to get a sample. This means that some wait a long time for a final diagnosis and incur a delay before starting treatment, or else are given TB treatment without a clear diagnosis. New and improved diagnostic tests which allow for the rapid detection of all active TB cases would greatly improve patient care.

Recently, scientists have found several new approaches to testing for TB disease which use new technologies to measure the immune response in blood samples. In previous studies, these new technologies were able to distinguish between TB and other diseases which appear clinically similar to TB. In this project, we aimed to validate these new technologies using samples from patients with the full range of TB disease, including those who test negative on the current tests. Overall, we found that the new technologies worked less well than previously reported. Importantly, they were unable to detect all of the TB patients who tested positive on the current tests. Due to the poor accuracy for diagnosing all TB patients, and the high cost of these new tests, we found that none of the new tests would be cost effective for use on all individuals who have suspected TB. However, in the hard-to-diagnose patient groups, for whom there are currently no rapid tests available, some of the new tests may be useful if used alongside existing tests.

Scientific Summary

Background

Tuberculosis (TB) is an important global disease causing millions of cases, including 1.3 deaths, annually. In order to reduce the burden of TB disease, new tools are needed, including improved diagnostic tests. The currently available toolkit lacks a test which can detect all cases of TB with high accuracy and speed. The gold standard tests for active TB are those which detect the causative bacteria, *Mycobacterium tuberculosis* (Mtb), but these have limitations. The smear test is quick, but lacks sensitivity for all TB cases and does not achieve 100% specificity; Mtb culture is more sensitive than smear, and has greater specificity, but lacks speed. The new nucleic acid amplifications tests are an improvement on current tests; they have a greater sensitivity than smear, are rapid and provide information about drug sensitivity. However, these tests are unable to detect culture negative TB cases. There are two major clinical unmet needs for TB diagnostics: a triage rule-out test for TB, and a rule-in test for hard to diagnosis TB cases (including culture negative TB).

There has recently been an explosion in discovery of biomarkers for TB based on measurement of a wide range of host responses using genome-wide gene-expression microarrays, proteomics techniques and characterisation of functional T cell subsets by flow-cytometry. The resulting biomarker and cellular immunologic signatures have high diagnostic sensitivity and specificity for active TB and distinguish active from latent infection. The feasibility of these novel approaches for diagnosis of TB has now been established and the apparent accuracy of the signatures has the potential to revolutionise TB diagnosis. In addition to the above, we have recently developed a highly sensitive whole blood molecular test that can reliably rule out the diagnosis of TB based on measurement of IFN-gamma-dependent chemokine release in response to Mtb specific antigens.

Thus, there are a number of new, promising biomarkers that now need urgent prospective and rigorous validation in routine clinical practice.

Aim

To validate promising new technologies (namely whole blood transcriptomics, proteomics, flow cytometry and qRT-PCR) and existing signatures for detection of active TB in samples obtained from individuals suspected of active TB.

Objectives

Primary objectives

1. To define the role of previously identified whole blood gene-expression signatures in the diagnostic evaluation of active TB using host transcriptomic microarray;
2. To define the role of previously identified serum proteomic signatures in the diagnostic evaluation of active TB using liquid chromatography mass spectrometry (LC-MS), and SELDI-TOF;
3. To define the role of previously established cellular immunological signatures in the diagnostic evaluation of active TB using flow cytometry;
4. To validate our candidate stimulated whole blood chemokine-based RT PCR assay for rapid patient triage by excluding the diagnosis of active TB;

Secondary objectives

5. To evaluate the use of the previously identified tests in the diagnosis of key sub-populations of active TB, namely: HIV+/-, culture confirmed/highly probable TB, pulmonary/extrapulmonary TB.
6. To define optimal diagnostic algorithms that delineate the role of these new tests in the routine diagnostic assessment of patients with suspected active TB in the NHS;
7. To assess the cost effectiveness of the best performing novel tests validated in this study for the diagnostic work up of TB patients in routine NHS practice.
8. To identify novel signatures of active TB (using the same technologies described in the primary objectives), selected to provide optimal diagnostic performance, i.e. beyond those previously proposed

Methods

This project used the biobank of samples from the Interferon gamma release assays (IGRA) in the Diagnostic Evaluation of Active TB (IDEA) study, where individuals with suspected TB were recruited as part of routine clinical practice in 12 hospital sites in England (2011-2014). The VANTDET project consisted of four laboratory sub-projects, each evaluating proposed tests using the following new technologies/approaches: transcriptomics, proteomics,

1 flow cytometry and candidate rule-out tests. For each of the sub-projects, the reference
2 standard used was the final diagnosis as allocated in the IDEA study, where clinicians used
3 the Dosanjh classification to stratify patients based on the diagnostic and clinical data
4 available (excluding the IGRA results) into the following groups: culture confirmed TB;
5 highly probably TB; clinically indeterminate TB; active TB excluded. Signatures were first
6 assessed according to their ability to detect all active TB cases (i.e culture confirmed and
7 highly probable TB combined) from the TB excluded group (other diseases, OD). Sub
8 analyses assessing the performance of tests/signatures for specific types of TB, i.e culture
9 confirmed, highly probable TB, pulmonary TB and extrapulmonary TB (EPTB) were also
10 performed.

11 For the validation of transcriptomic signatures from whole blood samples, n=628 individuals
12 were selected who reflected the full IDEA cohort, and were profiled using genome wide
13 gene-expression microarray using the Illumina platform. For validation of proteomic
14 signatures, we were restricted to n=90 for SELDI-TOF analysis due to lack of consumable
15 supply, while n=166 HIV negative individuals from the IDEA cohort were profiled using LC-
16 MS. For both of these 'omic' approaches, signatures were evaluated for their discriminatory
17 ability either by calculating the score as previously described, using raw
18 expression/abundance data, or using linear discriminant analysis (LDA).

19 To validate previous reported cellular immune signatures for discrimination of active TB
20 from individuals with latent TB infection (LTBI) and other diseases (OD) (OD/LTBI), we
21 took a nested case control approach to select the cohort. PBMCs from n=184 individuals
22 were stimulated with Mtb antigens overnight, stained with antibodies for function and
23 phenotypic markers and enumerated using flow cytometry.

24 For validation of candidate rule out tests, we took a GO/NO-GO approach to select the
25 cohort. The GO/NO-GO cohort of n=100 patients was selected to artificially inflate the
26 proportion of individuals with TB who are negative with standard IGRA tests. If high test
27 accuracy was achieved using this cohort, we planned to carry out the tests on remaining
28 samples. We developed protocols to evaluate the performance of two candidate immune
29 based rule-out tests: using either pRT-PCR to detect IFN γ -dependent target genes in Mtb-
30 stimulated PBMCs, or using Meso Scale Discovery (MSD) to detect chemokines in
31 Quantiferon supernatants.

For health economic analysis, the best performing tests were evaluated. Costs were estimated based on similar assays currently available in NHS practice. Decision tree models were developed to calculate the incremental costs and incremental health utilities [quality-adjusted life-years (QALYs)] of changing from current practice to using the candidate test as either rule-in or initial rule-out of active TB. The models were parameterised using the IDEA study clinical patient records and relevant current literature.

Results

The project and 4 sub-studies which assessed the previous published signatures measured using each of the new technologies, and a health economic analysis where the best performing tests were evaluated for cost effectiveness.

Principle findings in the transcriptomic sub-study

We assessed 6 previously reported gene expression signatures, using LDA. Three of these signatures could also be valuated using the score methods as previously described. For all TB, the optimal diagnostic accuracy for all signatures ranged from AUC=0.81-0.84; inclusion of HIV infected individuals in the validation cohort reduced diagnostic accuracy. The diagnostic accuracy for detecting culture confirmed TB or pulmonary TB (PTB) was better than for highly probable TB or extrapulmonary TB (EPTB) respectively.

Principle findings in the proteomics sub-study

We assessed the performance of one 4-protein signature using SELDI-TOF and LC-MS, and 3 additional previously reported proteomic signatures using LC-MS. The optimal performance for the 4-protein signature was found using SELDI-TOF, where the AUC for detecting culture confirmed TB was 0.74. Using LDA, the performance of all signatures for detecting all TB cases using LC-MS ranged from AUC=0.62-0.68.

Principle findings in the cellular immune signatures sub-study

Four of six previously described cellular immune signatures provided a reasonable level of diagnostic accuracy (AUC = 0.78-0.92) for discriminating all TB from those with OD/LTBI in HIV- TB suspects. Two of these assays may be useful in the IGRA+ population and can provide high positive predictive value (PPV).

Principle findings in the candidate rule-out test sub-study

While improved diagnostic accuracy for all TB over IGRA was achieved using the qPCR detection of CXCL9, CXCL10 or BATF2, none of the tests was able to provide >95% sensitivity and >65% specificity for active TB, and as such, would not be suitable for a rule-out test. The assay also resulted in a high failure rate and is complex. CXCL9 may provide good discriminatory ability for TB and LTBI in the T-SPOT+ population.

None of the evaluated new tests for TB can be considered cost effective for use in detecting all TB patients with their current cost and diagnostic performance.

Conclusions

None of the previously reported signatures using new technologies were able to provide diagnostic accuracy at a level which could be considered clinically useful for either all TB cases, or for culture positive TB. However, in specific patient groups, there may be a use for the best performing new tests. Specifically, while the diagnostic accuracy for detection of highly probably TB was sub-optimal (AUC=0.8), the Sweeney 4-gene transcriptomic signature may be useful in this the microbiologically unconfirmed population since it could provide a high NPV (0.95) and a result quickly. We identified two cellular immune signatures which could provide high PPV (0.88-0.95) for all TB, and for culture confirmed TB, which may be useful in the IGRA positive population of TB suspects.

In general, the performance of all signatures we evaluated was sub-optimal compared to previous publications, likely due in part to the heterogeneous cohort of TB and OD patients included in this study, which reflects the full complexity of TB and its differential diagnoses presenting in real life clinical settings. The secondary objective of defining novel signatures for active TB using the datasets generated is outstanding, but will build the bases for future studies. We recommend that future studies investigating blood based tests for TB using these technologies focus on specific sub-groups of TB patients and the OD they are clinically confused with. In particular, new tests for culture negative TB and EPTB are most needed. This project was funded by an NIHR EME project grant (Reference No. 12/65/27).

Word count: 1657

Chapter 1. Introduction

Background: Tuberculosis

Tuberculosis (TB) is a devastating disease, resulting in approximately 10 million cases worldwide each year.¹ It is one of the top ten causes of death globally, and is the leading cause of death from a single infectious agent, ahead of HIV/AIDS.¹ Important characteristics of the current global epidemic are the interaction with the HIV/AIDS epidemic (with HIV/TB co-infection driving increased mortality), and the increasing rates of multi-drug resistant TB (MDR-TB) cases. The areas with the greatest TB burden are in the South-East Asia, Africa and the Western Pacific World Health Organisation (WHO) Regions. England has among the highest rates of TB in Western Europe, with 5,664 notifications and a rate of 10.2 cases per 100,000 in 2016.² The majority of TB cases in Europe occur in large cities (especially London, Birmingham and Leicester in England) and the disease disproportionately affects immigrants and socially marginalised and socioeconomically disadvantaged groups.

TB is caused by the *Mycobacterium tuberculosis* (Mtb) bacterium, and is transmitted from person to person via aerosol route, by droplet particles which are produced when an individual with the pulmonary form of TB coughs. Upon exposure to an uninfected individual, a number of eventualities can occur. Approximately 50% of exposed individuals resist the infection, while the remainder become infected, as identified by detection of peripheral cellular immune responses to Mtb antigens. The majority of infected individuals will develop prolonged but controlled latent infection with Mtb (i.e. latent TB infection; LTBI). It is estimated that between one quarter and one third of the world's population has LTBI.³ In a small proportion of infected individuals, the immune response can lose control of Mtb replication, and in such cases TB disease will develop. In most cases this development to active TB occurs within the first 6 months after initial infection,⁴ although progression from LTBI to active TB (also known as reactivation) can occur at any time during an individual's lifetime. Active TB can manifest in many forms and can infect any organ of the body. Due to the mode of transmission, pulmonary TB (PTB) is the typical form of TB, however extrapulmonary TB (EPTB) is becoming increasingly common particularly in cases of reactivation from LTBI.

Reduction of TB cases and working towards elimination of TB as a public health issue are nationally and internationally agreed public health priorities at the level of the United

Kingdom (UK), European Union (EU) and WHO.^{1, 5} Whilst global TB incidence is declining by around 2% per year, improved reduction rates are required to progress towards elimination. Internationally agreed targets for improved global TB control, such as in the Sustainable Development Goals or the End TB Strategy, require that TB incidence rates to fall by up to 5% per year, which is not achievable with existing tools. New, innovative methods for TB prevention, diagnosis and treatment are urgently needed and the WHO's *Global Plan* calls for researchers and industry to deliver an improved, fit-for-purpose anti-TB toolkit, with major emphasis on the need for new diagnostic tests.⁶ Prompt and accurate diagnosis and treatment of TB are essential to improve individual patient outcomes, as well as to prevent onward transmission in the community with its attendant health-care and social costs. However, the currently available diagnostic toolkit lacks the ability to provide results with clinically acceptable speed and diagnostic accuracy.

Current diagnostic tests for TB

Currently available tests recommended by the National Institute for Health and Clinical Excellence (NICE) for diagnosis of active TB in the UK comprise smear microscopy, Mtb culture, radiography, and nucleic acid amplification tests.⁷ Smear microscopy is rapid but has low sensitivity, while radiography has low specificity. Mtb culture is the current microbiological gold standard, but is limited by its lack of speed (requiring 2-6 weeks to generate a result) and is only moderately sensitive. Importantly, both smear microscopy and Mtb culture suffer from reduced sensitivity in a number of key 'hard-to-diagnose' patient populations, including those with EPTB or HIV co-infection.

Nucleic acid amplification tests have been a major advancement in TB diagnostics, and the GeneXpert MTB/RIF test (Cepheid Inc, CA, USA) now performs with higher sensitivity than smear microscopy and provides a result on the same day. GeneXpert is a substantial improvement on conventional microbiological tests, especially now that the improved GeneXpert Ultra has shown greater sensitivity in the smear-negative culture-positive population of TB cases (64% compared with 46% with conventional GeneXpert).⁸ However, culture-negative TB cases (another key 'hard-to-diagnose' patient subgroup) are mostly falsely-negative when tested by GeneXpert. As with microscopy and culture, sensitivity of GeneXpert (and Ultra) is also lower in people EPTB and HIV co-infection. Furthermore,

GeneXpert is expensive, and its cost effectiveness within routine NHS practice is unknown. Thus, within the UK setting, culture remains the gold standard.

The Tuberculin Skin Test (TST) and Interferon Gamma Release Assay(IGRA)s are immunological tests recommended by NICE for the diagnosis of LTBI.^{7, 9} While these tests can be used to detect Mtb infection, none are able to discriminate between LTBI and active TB disease, and all lack sufficient diagnostic sensitivity to exclude a suspected diagnosis of TB.^{10, 11} IGRAs have improved specificity for Mtb infection over the TST, which is compromised by poor specificity due to cross-reactive responses to the test in BCG vaccinated or non-tuberculous Mycobacteria (NTM) exposed individuals.¹² There are two main types of IGRA: the Quantiferon Gold-In-Tube (QFT-GIT) .^{11, 13} Therefore, while the IGRA and TST tests are recommended for detection of LTBI, the utility of IGRAs in the diagnostic work-up of active TB is not currently recommended. However, should there be an upgrade to these immunological tests whereby the sensitivity for all active TB was increased to >95%, they would have the potential to be used as triage tests to rule out active TB. The IDEA study, which provided the cohort and biobank samples for VANTDET, was an observational prospective cohort study which sought to define the role of commercially available IGRAs, as well as a second generation IGRA, in the diagnosis of active TB. IDEA found that while the T-SPOT had significantly greater sensitivity for active TB than QFT-GIT (81.4 and 67.3% respectively), neither the test had sufficient sensitivity to rule out a diagnosis of active TB. However, the second generation IGRA had a sensitivity of 94% for culture confirmed TB, and 89% for all TB, providing the first proof of principle that an immune based test has the potential to practically serve as a rule out test.^{11, 13} Such a test would significantly reduce the time to diagnosis of patients for whom TB is a suspected condition on initial assessment, but are definitely diagnosed with another disease by the end of diagnostic work-up.

IGRAs and the TST are widely used to detect LTBI in those at high risk of having the infection, such as those who have recently been exposed to an active case of pulmonary TB. A recent large prospective cohort study investigating the predictive power of the IGRA and TST tests for detecting progression to active TB disease (amongst new entrant migrants and recent close contacts of TB cases) in the UK found that these tests provide high negative predictive power but very low positive predictive power for identifying those who progress to active TB.¹⁴ Despite culture being the gold standard microbiological test for active TB in the UK and elsewhere, it is standard practice to commence treatment for clinically suspected TB

cases before a positive culture result has confirmed TB. Globally, while a large proportion of cases of TB are diagnosed based on a positive sputum-smear microscopy test, the sensitivity of smear for all TB cases (especially in EPTB, childhood TB and with HIV co-infection) is low and as such, a large proportion of patients are diagnosed clinically and treated without a definitive diagnosis. The unique epidemiology of TB in the UK/England (i.e. high rate of culture negative TB cases and high proportion of active cases with EPTB²) means that a larger proportion of TB cases who start treatment lack microbiological confirmation, and as such the need for tests to detect these hard-to diagnose groups of patients is greater.

Clinical unmet needs

Globally there is a need for improved diagnostic tests for both active TB and LTBI. As identified by the New Diagnostic Working Group (WHO and Stop TB Partnership), there is a need for a ‘rapid biomarker-based non-sputum-based test for detecting TB’.¹⁵ A fast, highly sensitive test for all types of active TB, with improved sensitivity over the current microbiological tests (which provide high specificity but low sensitivity) would improve triage of TB suspects and increase the speed at which patients start appropriate treatment.

In high income, low incidence settings such as the UK, a high proportion of new cases of active TB are a result of reactivation from LTBI. As such, a major priority for TB control programmes in these settings is to identify individuals with LTBI and to offer prophylactic treatment to those at highest risk of progression to active TB.¹⁶ While IGRAs and TST provide high sensitivity for these individuals, an independent or follow-on test with improved specificity would enable us to reduce the number needed to treat to prevent subsequent cases of progression to active TB.

Therefore, the major clinical unmet needs for improved TB diagnostics in the UK are as follows:

- A triage rule-out test for active TB.
- A rapid, accurate test for identifying ‘highly probable’ TB, where a clinical diagnosis cannot be confirmed by culture.
- A sensitive test to predict those at increased risk of progressing to active TB from exposure and/or positive test for LTBI.

Rationale for VANTDET

Blood-based tests – the rationale, and technological approaches

The low sensitivity of microbiological diagnostic tools for TB reflects the low bacterial load in a high proportion of clinical samples from TB cases. Therefore, host-derived biomarkers such as molecular transcriptomic or proteomic signatures have been proposed as an approach which may have improved sensitivity for TB diagnosis.¹⁷⁻¹⁹ The pursuit of TB biomarkers has been reinvigorated by the developments in high-throughput transcriptomic, proteomic and cellular immune technology platforms and parallel advances in computational analytical methodologies. The past 18 or so years has seen a wealth of publications in this field, particularly in the development of transcriptomic biomarkers for active TB,^{17, 19} where promising test accuracy has been demonstrated in large, multi-site international cohorts of TB cases and controls. In addition, serum proteomic profiling has also led to blood-based biomarkers of TB with promising test accuracies.^{18, 20}

Currently available immune based tests have shown promise of greater sensitivity for detecting active TB when compared to microbiological tests, leading to speculation that development of newer more sophisticated immunological assays might lead to more accurate immune based tests for TB. Several strategies for improving either sensitivity, or specificity for TB have been proposed, including using different/additional antigens for stimulation,²¹ measurement of additional/alternative analytes (to IFN γ),^{22, 23} and/or measurement of cellular phenotype.²⁴ Measurement of additional or alternative analytes has been shown to improve both sensitivity and specificity for active TB in case-control studies. The measurement of phenotype and function, thus far solely using flow cytometry technology, has been shown in multiple studies to improve specificity by discriminating between active TB and LTBI.^{23, 24}

Technologies

In light of the above developments in the field of blood-based diagnostic tests for TB, we sought to validate the following approaches for the diagnostic evaluation of active TB: transcriptomics; proteomics; flow cytometry and chemokine qRT-PCR (see *TABLE 1*).

TABLE 1. Details of Technologies to be assessed in VANTDET.

Blood sample type	Field	Technology to be tested in VANTDET
RNA (whole blood)	Transcriptomics	Illumina Microarray
Serum	Proteomics	SELDI- TOF
PBMC	Cellular immunology	Flow cytometry Improved immune based tests for rule-out of active TB

The need for a real-life clinical cohort of active TB suspects

A large proportion of reported cases of TB are diagnosed without confirmation from standard tests. Furthermore, most studies of new diagnostic approaches fail to include individuals who lack culture confirmation (or those negative with other microbiological tests such as Gene Xpert, smear microscopy etc), despite the fact that those with a negative culture results are the very population in whom a new diagnostic test would be most beneficial. This is why, when validating new diagnostics for active TB, we require a representative spectrum of patients, including those with a culture negative status. We have therefore used the samples collected as part of the IGRA in diagnostic evaluation of Active TB (IDEA) study, an NIHR HTA-funded, prospective, multi-centre UK cohort study of over 1,000 adults presenting with suspected active tuberculosis at NHS out- or in-patient services at 13 participating NHS hospitals in London, Birmingham and Leicester.¹¹ Patients with suspected TB were prospectively recruited into IDEA at the point of initial diagnostic work-up in secondary care in routine clinical practice, before a diagnosis was confirmed. The new diagnostic tests for ruling in active TB were evaluated against the composite reference standard to allow for the inclusion of clinically diagnosed TB patients in the TB disease group. Within VANTDET we also evaluated two candidate approaches for a triage *rule-out test*, which aim to improve on the sensitivity of the current IGRA tests.

Study scope

The scope of the VANTDET project is specifically to address the needs for new diagnostic tests for active TB in individuals for whom there is a clinical suspicion of the condition, i.e the first two clinical unmet needs listed on page 30, but not the third which concerns latent TB. While a triage test for active TB may potentially have a use in other settings, such as in contact clinics or for screening new patients from high incidence areas, this is beyond the scope of this study.

Aim

To validate promising new technologies (namely whole blood transcriptomics, proteomics, flow cytometry and a chemokine qRT-PCR) and existing signatures for detection of active TB in samples obtained from individuals suspected of having active TB.

Research Objectives

Primary objectives:

1. To define the role of previously identified whole blood gene-expression signatures in the diagnostic evaluation of active TB using host transcriptomic microarray;
2. To define the role of previously identified serum proteomic signatures in the diagnostic evaluation of active TB using mass spectrometry (MS), and SELDI-TOF;
3. To define the role of previously established cellular immunological signatures in the diagnostic evaluation of active TB using flow cytometry;
4. To validate our candidate stimulated whole blood chemokine-based RT PCR assay for rapid patient triage by excluding the diagnosis of active TB;

Secondary objectives:

5. To evaluate the use of the previously identified tests in the diagnosis of key sub-populations of active TB, namely: HIV+/-, culture confirmed/highly probable TB, pulmonary/extrapulmonary TB.
6. To define optimal diagnostic algorithms that delineate the role of these new tests in the routine diagnostic assessment of patients with suspected active TB in the NHS;
7. To assess the cost effectiveness of the best performing novel tests validated in this study for the diagnostic work up of TB patients in routine NHS practice.
8. To identify novel signatures of active TB (using the same technologies described in the primary objectives), selected to provide optimal diagnostic performance, i.e beyond those previously proposed

1

2

Chapter 2. Study overview, management and general methodology

This study consists of 4 sub-studies each addressing the main technologies to be assessed in the primary objectives 1-4. The sub-studies are outlined in detail in the chapters 3-6; the secondary objectives (5 -8) are addressed within these same chapters, in the context of each of the technologies being assessed. The health economic assessment is presented for all the best-performing tests in Chapter 7. The reporting of each of the sub-projects adheres to the Standards for Reporting of Diagnostic accuracy studies (STARD guideline), as shown in 0.

The biobank of blood samples that were collected and stored as part of the IDEA study¹¹ were used for the 4 sub-projects. However, the criteria for cohort selection, sample processing, technical methodology and analytical approach differed between these sub-projects. As such, detailed methodology and rationale can be found within the chapter corresponding to each sub-project. This chapter summarises the main features of the overall study cohort (i.e. the IDEA cohort), as well as the overarching methodology and approach for study management.

Study management structure

Overview of study design

The original protocol specified the link between IDEA and VANTDET, as well as the processes and samples for each VANTDET sub-study (*FIGURE 1*). Within the IDEA study, additional blood was taken to allow for long term storage of PAXGene, serum and PBMC samples. These samples, coupled with clinical data, provided the biobank for the VANTDET study.

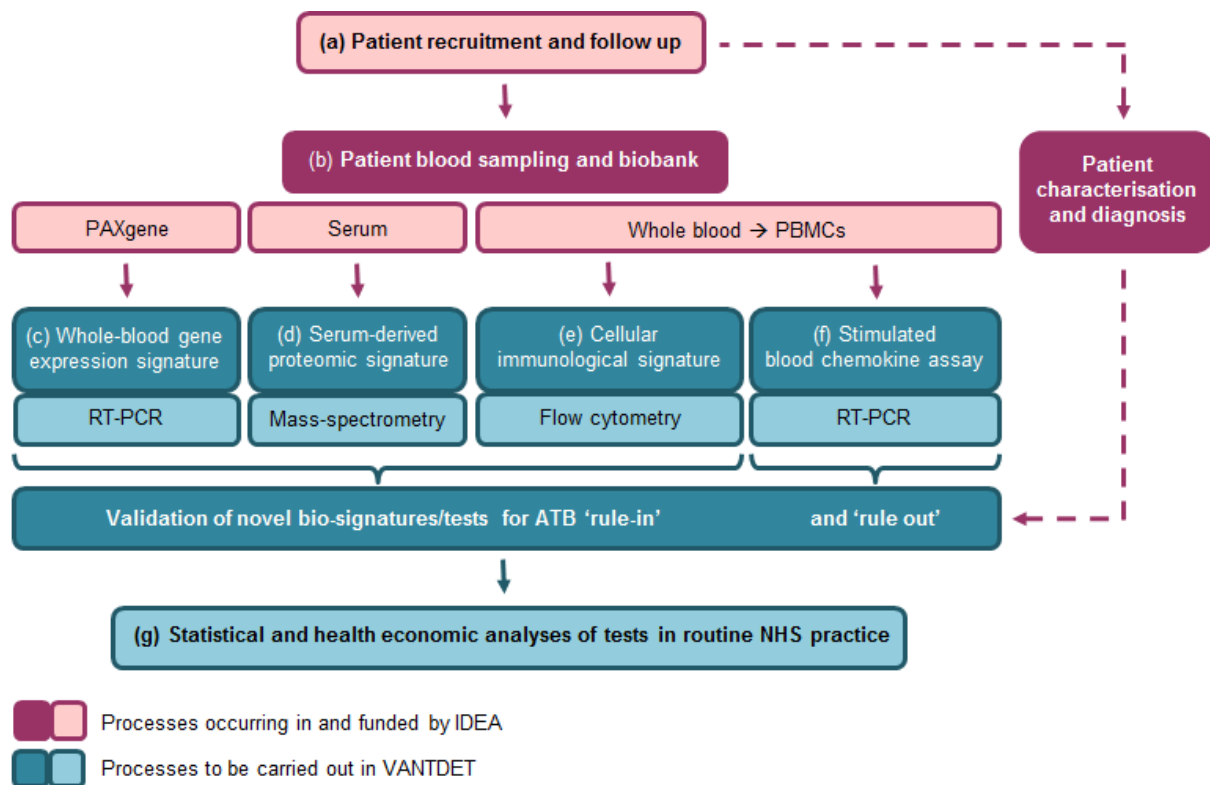


FIGURE 1. Proposed VANTDET study design, and link with the IDEA study.

General management and task allocation

The VANTDET study required a relatively simple management structure and a less rigorous monitoring process because there was no patient recruitment. The Chief Investigator (AL) and interim Chief Investigator (PB led the project from May 2015 to May 2016) provided general oversight of the study. Each of the four sub-projects was managed by a laboratory scientist (3 Research Associates (AH, LH, RP) and a Laboratory Technician (MTW)), who all worked within the TB Research Centre laboratory at St Mary's Campus, Imperial College London. Data processing and analysis support was provided by a bioinformatician (PJ), and laboratory and analytical support was provided by a research MSc student/technician (TM) and other TB Research group members (AB, VM, AF). These individuals comprised the main research team. The SELDI-TOF proteomic validation work was performed by a research associate (SH) and overseen by a Co-Investigator (ML) and the CI (AL). A research associate (NG) and Co-Investigator (PW) in the Department of Infection Disease Epidemiology (also based at St Mary's in Imperial College London) were responsible for the health economic analysis, which was done at the end of the project and involved discussion with the main

research team. Medical statisticians (YT, JD) provided advice on study design, cohort selection, analytical protocols and presentation of data.

Management and oversight meetings

Study Management Group

The Study Management Group (SMG) included the chief investigators, post-doctoral research associates and technicians. The SMG met regularly (at least fortnightly in lab meetings), to report on progress, share/discuss results, methodology, etc.

Wider Study Management Group

The Wider Study Management Group consisted members of SMG and statisticians from the University of Birmingham: Dr Yemisi Takwoingi and Dr John Deeks. The group met three times (22/09/2015, 03/02/2016, 18/05/2016) to discuss the statistical approaches taken for each of the objectives/sub-studies and established the need for Independent Scientific Advisory Board.

Independent Scientific Advisory Board

The Independent Scientific Advisory Board (ISAB) consisted of three independent members: Professor Sanjeev Krishna, Dr Mahdad Noursadeghi, and Dr David Connell. The ISAB and SMG met once (15/07/2016) to discuss overall study design and research plans for each objective.

Issues and problems

The SELDI arrays required to validate the SELDI-TOF signatures were discontinued by the manufacturer. Initially, an alternative supplier was sought, but unfortunately was not found. Hence, this platform will not be available for future diagnostic tests. We are performing analysis using the remaining stock from the manufacture (n=90). We are performing LC-MS to identify serum biomarkers and will develop ELISA tests.

Changes to study protocol and management plan

The study protocol and management plan documents were changed once, in October 2016. The comments below are taken from the progress report which followed, the updates (March 2017).

Protocol

The changes included updated methods for the validation of cellular immunological signatures and candidate rule-out tests TB, based on new literature and results from the IDEA study. Furthermore, there were changes in the sample size for both the cellular immune signature validation and proteomic signature validation.

Cellular immune signature validation:

Additional cellular immune signatures were identified, since conception of the VANTDET project and protocol, which also show promise in their ability to discriminate between active TB and LTBI. These new signatures were tested in house in preliminary experiments, before the laboratory was protocol updated, and optimised so that these additional signatures could be identified within the flow cytometry panel we used on the cohort. These included those which measure the proportion of Mtb-specific CD4⁺ IFN γ ⁺ cells with either an activation phenotype²⁵ or the levels of CD27 as a ratio or in combination with CD45RA expression.^{26, 27}

Candidate 'Rule-out' tests:

In a draft final report, the IDEA study identified that the use of four antigens combined (ESAT-6, CFP-10, EspC and Rv3879c) in the T-SPOT TB assay provided the highest sensitivity for active TB.¹¹ Use of all four of these antigens was therefore included in the assay for validation of a rule-out test for TB.

Study Population and Sample Sizes

Proteomic Signature: For the proteomics experiments samples will be chosen at random from the IDEA cohort (HIV-), the 1st 90 will be compared using SELDI and LC-MS techniques and the sensitivity and specificity compared to the landmark studies. The Levin et al., (unpublished SELDI signature) is currently being adapted into an ELISA based test, if this SELDI signature is validated in the IDEA cohort (n=90, Sensitivity >85%), this ELISA test will be used as validation of the entire IDEA (HIV-/+) cohort to provide CI <5% for ATB versus OD. For LC-MS a total of 200 samples (HIV-/+) will be analysed and this dataset used to explore the diagnostic performance both previously identified and novel protein markers.

Cellular Immunological Signature: Using a nested case control approach, ~82 active TB patients were randomly selected (but enriched for those with HIV infection), to add to the ~82 LTBI patients, to provide a final cohort of 164 individuals. Whilst smaller than the ideal

sample size, this will still be more than sufficient to provide enough power to demonstrate 95% sensitivity and specificity.

Molecular Rule-Out Test: In order for the molecular test to be used to rule out TB, it must have extremely high sensitivity, and therefore must provide a positive result for those individuals who are infected with Mtb but who fail to test positive in the most sensitive of the currently-available tests for Mtb infection (the TSPOT-TB). We have therefore selected an initial cohort of 100 individuals recruited into IDEA, selected at random, but conforming to the following criteria (25 x individuals with active TB and a positive TSPOT-TB; 25 x individuals with active TB and a negative TSPOT-TB result; 25 x individuals with other diseases with a positive TSPOT-TB result; 25 x individuals with a negative TSPOT-TB result). Once this initial cohort has been tested using the molecular rule out test, the results will be analysed to test the performance of the assay in ruling out active TB. If the results are promising (i.e. > 95% sensitivity is demonstrated), we will continue to assay the rest of the available PBMC samples from the IDEA cohort, in order to maximise the confidence of our findings.

Management plan

An updated management plan which reflected the changes in timelines, sample sizes, protocols and experimental approaches was submitted to NIHR/EME in October 2016. The new end of study date of 31 December 2017 was agreed. There were no further changes made to the management plan during the course of the project.

Study cohorts

While the focus of this study project was to evaluate all four novel technologies in the diagnostic evaluation of all TB, each required a slightly different approach to selecting a final study cohort from the full IDEA cohort. The details of the inclusion criteria are presented in detail within each sub-study chapter. The inclusion criteria, as well as the sample size calculations, were agreed by the study oversight groups and the ISAB.

The IDEA cohort and biobank

The IDEA study cohort is described in detail in the IDEA Health Technology Assessment (HTA) report¹¹ and recent manuscript.¹³ In summary, 1,060 adults (aged ≥ 16 years) with suspected active TB were prospectively recruited to 12 NHS hospital sites in London, Slough, Oxford Leicester and Birmingham. Due to a low proportion of HIV-positive patients

being recruited within the main study cohort of active TB suspects, an extended study was performed to increase the number of HIV+ patients. The HIV infection rate in the full cohort was 22.9%, but in the main study it was 16%.

In the main cohort, 845 individuals were included in the final analysis. Of these, 363 (43%) had active TB, of which 261 (72%) were confirmed by culture. This represents the rate of TB in the full cohort of TB suspects recruited in the IDEA study, a population which is reflective of all TB suspects in areas with the high TB rates in England. TB was excluded in 52% of patients, and the remaining 5.1% were clinically indeterminate. For a more detailed description of the full IDEA cohort, including demographic and clinical characteristics, please refer to the publications and reports from this study.¹¹

Reference standard

For the validation of signatures for the detection of active TB, a composite reference standard used, as with the IDEA study: the Dosanjh categorisation, by which patients are divided into 4 possible groups: definite TB; highly probable TB; clinically indeterminate; and non-TB (TABLE 2). Importantly, this allows for inclusion of the full spectrum of TB cases, including those lacking microbiological conformation (Dosanjh category 2).

TABLE 2. Dosanjh categorisation for the diagnosis of active TB suspects, as from.¹⁰. This approach to categorisation of patients was used as the diagnostic reference standard for both IDEA and VANTDET studies. Table adapted from the Annals of Internal Medicine publication: Dosanjh, Davinder P.S.; Hinks, Timothy S.C. et al., Improved Diagnostic Evaluation of Suspected Tuberculosis, Volume No. 148 , Issue No. 5, Pages 325-336. Copyright © [2008] American College of Physicians. All Rights Reserved. Reprinted with modifications, with the permission of American College of Physicians.

Diagnostic category	Criteria
1: Culture confirmed TB	Microbiological culture of <i>M. tuberculosis</i> , AND suggestive clinical and radiological findings.
2: Highly probable TB	Clinical and radiological features highly suggestive of TB unlikely to be caused by other disease, AND a decision to treat made by a clinician, AND appropriate response to therapy, AND histology supportive if available.
3: Clinically indeterminate	Final diagnosis of TB neither highly probable, nor reliably excluded.
4: Active TB excluded: Subclassification	
4A: Inactive TB	History of prior episode OR stable CXR changes, AND TST positive* (if done), AND bacteriologically negative (if done), AND no clinical evidence of active disease.

4B: One or more risk factors for TB exposure [†] , TST positive*	TST positive*, AND bacteriologically negative (if done) AND no clinical evidence of active disease.
4C: One or more risk factors for TB exposure [†] , TST negative	History of TB exposure, AND TST negative (if done).
4D: No risk factors for TB exposure [†] , TST negative	No history of TB exposure, AND TST negative (if done)

CXR: chest radiograph; TB: tuberculosis; TST: tuberculin skin test

* A TST using the Mantoux test with a threshold of ≥ 15 mm considered positive.

[†]Risk factors for TB exposure: recent exposure to active TB patient, born in country of high prevalence or belonging to an ethnic group with a high prevalence of TB (Incidence $>100/100,000$, Rose et al28).

Categorisation of patients was performed by a clinical panel of 4 respiratory medicine and infectious diseases clinicians specialising in TB, who had access to anonymised diagnostic and follow up (up to 12 months post recruitment) clinical details for each patient, including response to treatment, while remaining blind to final IGRA results, as outlined in the IDEA study protocol. It should be noted that not all individuals in Dosanjh 2 category were culture negative; some individuals did not have a culture test done or it failed. This group should therefore be considered as ‘culture unconfirmed’ rather than ‘culture negative’. Given that for active TB to be excluded, a definitive diagnosis of another condition is required. Therefore, in this report, we refer to the ‘active TB excluded’ group (i.e. Dosanjh category 4) as ‘other disease’ (OD).

Patient groups used for main and sub-group analyses

The primary analysis for determining test accuracies and other diagnostic accuracy parameters of previously described tests/signatures was to assess their performance in detecting all TB (i.e. Dosanjh 1 and 2 combined) against all OD within each cohort. Given the uncertainty over the correct final diagnosis of those who were categorised as ‘clinically indeterminate’ (Dosanjh 3) by the clinical panel in the IDEA study, these patients have been excluded from each of the VANTDET sub-studies (as was the case with IDEA).

In each sub-study (with the exception of the candidate rule-out tests), sub-group analyses were also performed, which included determining diagnostic performance for detecting either culture confirmed (Dosanjh 1) or highly probable (Dosanjh 2) TB, as well as for either PTB or EPTB. In these cases, the PTB group included individuals who had pulmonary involvement as well as those who had pulmonary and extra-pulmonary involvement. Those with extra-pulmonary involvement only were classified as EPTB.

Prevalence data

Within this report, we have calculated prevalence values for active TB groups in various parent populations, using the data from the final IDEA study (both the ‘main study’ and ‘HIV cohort’¹¹, without exclusion of the ‘clinically indeterminate’ population (Dosanjh 3)). The prevalence values are given in (TABLE 23 in Appendix 1)(Appendix 1)

Patient and public engagement and involvement

The TB Research Centre where the laboratory team of VANTDET are based, is associated with the Health Protection Research Unit (HPRU) in Respiratory Infections, and both are led by the VANTDET CI, Prof Ajit Lalvani. The HPRU conducts an extensive patient and public involvement and engagement (PPI/E) programme, which has been recognised within Imperial as an exemplar of good practice (see our case studies in the Imperial Patient Experience Research Centre <http://bit.ly/2p9AKei> and <http://bit.ly/2pakraO>). The TB Research Centre team, including the VANTDET researchers AH, MTW, LH, TM and RP, are frequently requested to run public engagement events.

For the IDEA study, Ms Nisha Karnani (a lay member) was appointed as the patient representative for the duration of the study. She was consulted at key points during the study and was invited to SSC meetings and the IDEA study presentation at the end of the study.

HPRU conducts quarterly symposium series I which the VANTDET team has disseminated and discussed project findings to a wide audience, including the HPRU in Respiratory Infections ISAB, patient representatives, collaborators and colleagues. Members of the management board have an excellent track record in active engagement with patients and public.

- Professor Ajit Lalvani was a key member on the panel at the Engagement to End TB event at the University of East London. This was part of the Voices for London Festival which was attended by community members of East London.
- The VANTDET team has participated in several public engagement events to showcase the TB research we are involved in. We also, along with other TB Research Centre members, received specific public engagement funding to develop and evaluate an animation film and latent TB diagnostics and treatment. This film was evaluated with screening and focus groups events for community groups, members of

the public and other institutions involved in public engagement. These include Imperial Festival (May 2017), Bromley-by-Bow Unity Festival (East London, July 2017), Screening of latent TB film at Midaye Somali women's group (White City, December 2017) and Let's Talk about Tuberculosis (White City, December 2017)

The team has also published their research work in various online articles for a wider public audience.

- Tuberculosis in England: How research at Imperial is supporting the national strategy - <http://bit.ly/2oFtBkD>
- Increasing awareness about latent tuberculosis infection - <http://bit.ly/2CPk0g9>
- Multilingual White City residents help Imperial raise awareness of TB research - <http://bit.ly/2CP2YyS>
- Case study: Evaluating TB awareness animation on Imperial PERC PPI Resource Hub <http://bit.ly/2pakrO5>

The HPRU in Respiratory Infections and TB Research Centre have established their first Patient and Public Panel which comprises 7 members of the public directly or indirectly affected by, or interested in, our research areas and both with, and without, previous experience of PPI. The role of the panel is to support the delivery of effective patient and public involvement and engagement approaches and activities within the HPRU. The panel advises on different PPI avenues available for disseminating our findings such as charities/voluntary organisations websites and blogs, and also helps to distribute findings within their informal networks.

Statistical analysis and presentation of data

For statistical analyses, R software or Prism was used as indicated in the methods sections of Chapters 3 to 6. For all technologies and datasets, the outcome of signature performance was presented in either a dot plot, box-whisker plot and/or a ROC-curve, all generated using Prism software (version 7). ROC-curve analysis was used to estimate diagnostic accuracy.

Blinding

All individuals carrying out experiments and determining index test results were blind to clinical data, reference standard results and all other patient specific information. Only study ID was known.

Defining the cut offs

In each sub-study, multiple methods for defining test cut-offs were used, depending on the proposed use of the test. For those tests being assessed as rule in tests, specificity was set at >90%. For those being assessed as rule-out tests, sensitivity was set at >95%. For a balanced approach, either the cut-off with the optimal Youden value was given (for transcriptomic and proteomic tests), while for the cellular immune signatures the cut off where sensitivity=specificity was used, in line with standard practice in these respective fields of study.

Chapter 3. Transcriptomics

Introduction

Several studies have been demonstrated that whole blood gene expression profiles have the potential to be used for diagnostic tests for active TB, in adults and children. In 2010 Berry et al. demonstrated that differential transcriptional patterns in peripheral blood can be harnessed to differentiate TB from other conditions and a whole-blood 86-transcript signature distinguished active TB from other inflammatory and infectious diseases.¹⁷ A major milestone was reached by a study that showed a 44-blood transcript signature was able to differentiate active TB from other diseases (OD) using a cohort from a high burden setting.¹⁹ Anderson used a similar approach to identify a 51 gene signature which could distinguish TB from other disease in a cohort of children, where all patients had suspected TB at recruitment.²⁸ This study has particular strengths in the approach to study design, since highly probable TB cases were also included, and the OD group included a wide variety of differential diagnoses that present in a real life clinical setting. Such a study in adults has yet to be used for validation of transcriptomic signatures for TB diagnosis. Maertzdorf et al. identified a four-gene signature (GBP1, ID3, IFITM3 and P2RY14) that distinguished active TB from healthy controls and latent TB infections with high accuracy.²⁹ Similarly, a combination of four genes (CD177, CLC, HP and IGJ) was identified using support vector machine learning, and was shown to accurately classify active TB (including EPTB) from non-TB febrile patients.³⁰

Most recently, Sweeney et al.³¹ used a meta-analysis approach to derive a TB risk score from three genes (DUSP3, GBP5 and KLF2) which was able to distinguish active TB from OD with high accuracy. The important characteristics and diagnostic classification of TB is summarised in *TABLE 26* for the five studies and the meta-analysis.

The aim of this sub-study is to define the role of previously identified whole blood gene-expression signatures in the diagnostic evaluation of active TB using host transcriptomic microarray.

Methods

Inclusion criteria

The inclusion criteria were: definite diagnosis of either active TB (Dosanjh 1 or 2) or other disease (Dosanjh 4), consent given for genetic analysis, and PAXGene sample available. Those with a clinically indeterminate (Dosanjh 3) diagnosis were excluded. Amongst HIV positive participants in the IDEA study, all active TB cases were selected and a similar number of participants with OD were randomly selected for the microarray assay analysis.

RNA extraction and quality control

For whole blood RNA, 2.5mL of blood were collected into PAXgene® tubes (PreAnalytiX, Hombrechtikon, Switzerland) and were frozen at -80°C as part of the IDEA protocol¹¹. RNA was extracted from PAXgene tube samples using the PAXgene Blood RNA Kit (PreAnalytiX, Hombrechtikon, Switzerland), according to manufacturer's instructions. RNA QC was performed using both Nanodrop and Bioanalyzer instruments.

Whole genome-wide microarray

Illumina gene expression microarray platforms were used in this study. The detailed protocol has been published.³² Briefly, a total of 100ng of RNA in 11ul of RNase free water was used for synthesizing first strand cDNA using a master mix containing 1µl T7 Oligo(dT) primer, 2µl 10X first strand buffer, 4µl dNTP mix, 1µl RNase inhibitor and 1µl ArrayScript. Next, a second strand cDNA was synthesized using 80ul of master mix (63µl nuclease free water, 10µl 10X second strand buffer, 4µl dNTP mix, 2µl DNA polymerase, 1µl Rnase H). After purification, 17.5ul of cDNA was in vitro transcribed to synthesize antisense RNA using IVT master mix (2.5µl T7 10X reaction buffer, 2.5µl T7 enzyme mix, 2.5 biotin-NTP mix). Next, a total of 750ng cRNA was used for array hybridization to the Illumina Human HT12-v4 beadchips according to the manufacturer's protocol. Finally, the beadchip was scanned on the Illumina BeadArray Reader. The gene expression data was extracted from images collected from Illumina BeadArray Reader.

Microarray QC and normalization

Raw probe intensity signal for approximately 47,000 transcripts were acquired from the arrays using Beadstudio software (Illumina Inc.) and normalized using Lumi R package.^{33 34} Firstly, the raw data was log2 transformed using lumi.T function, and then Robust Spline Normalization (RSN) was applied using lumi.N.³⁵ Next, transcripts that were not expressed in

any samples (Illumina detection P value > 0.01) were removed from further analysis. Finally, principle component analysis (PCA) was used to remove outliers.

Statistical analysis

Calculation of signatures scores

For those published signatures where the gene expression levels of selected target genes were combined to give a score, these were calculated as previously described.^{19, 28, 31}

Linear discriminant analysis

Linear discriminant analysis (LDA) is a supervised classification approach. It was used to model the linear combination of candidate transcripts from given previously reported signatures as a function of sample type (i.e. TB or OD). LDA is based on the assumptions of multivariate normal distribution of the explanatory variables (candidate transcripts) and these variables have same covariance matrix for classification categories (sample types) considered.

Testing trained models

For Roe et al., signature³⁰ the original trained support vector machine model was available hence classification performance of this model was assessed, in addition to leave-one-out LDA model, on the full cohort and performance was reported. Support vector machine is a machine learning algorithm that used preselected variables to perform a binary classification of input samples, in our case two classes were Active TB (ATB) and Other Disease (OD).

ROC curve analysis and defining cut-offs

Test values (i.e score, probability or otherwise) for the different signatures in diagnostic evaluation of the key patient groups were assessed for discriminatory ability using ROC analysis, performed using either Prism or R software. Three alternative approaches were used for selecting the cut-off. To prioritise sensitivity, the cut-off providing >95% was given, to prioritise specificity, the cut-off providing >90% specificity was given. To optimise both sensitivity and specificity, the maximum Youden index (i.e. sensitivity + specificity-1) was used to select the cut-off.³⁶

1 ***Calculation of predictive values***

2 To determine the predictive value of tests in given populations where TB prevalence is
3 known (or assumed), we calculated the positive predictive value (PPV) and negative
4 predictive values (NPV).

5

Results

Final cohort

The study flow chart in *FIGURE 2* shows the final cohort selected from the total IDEA cohort. In the IDEA study, 288 did not give consent for a PAXGene sample to be taken/genetic analysis. Of the 799 patients who were eligible for the transcriptomic sub-study, 628 were selected for RNA extraction, had sufficient RNA, were used in microarray analysis, passed the RNA and micorarray QC checks.

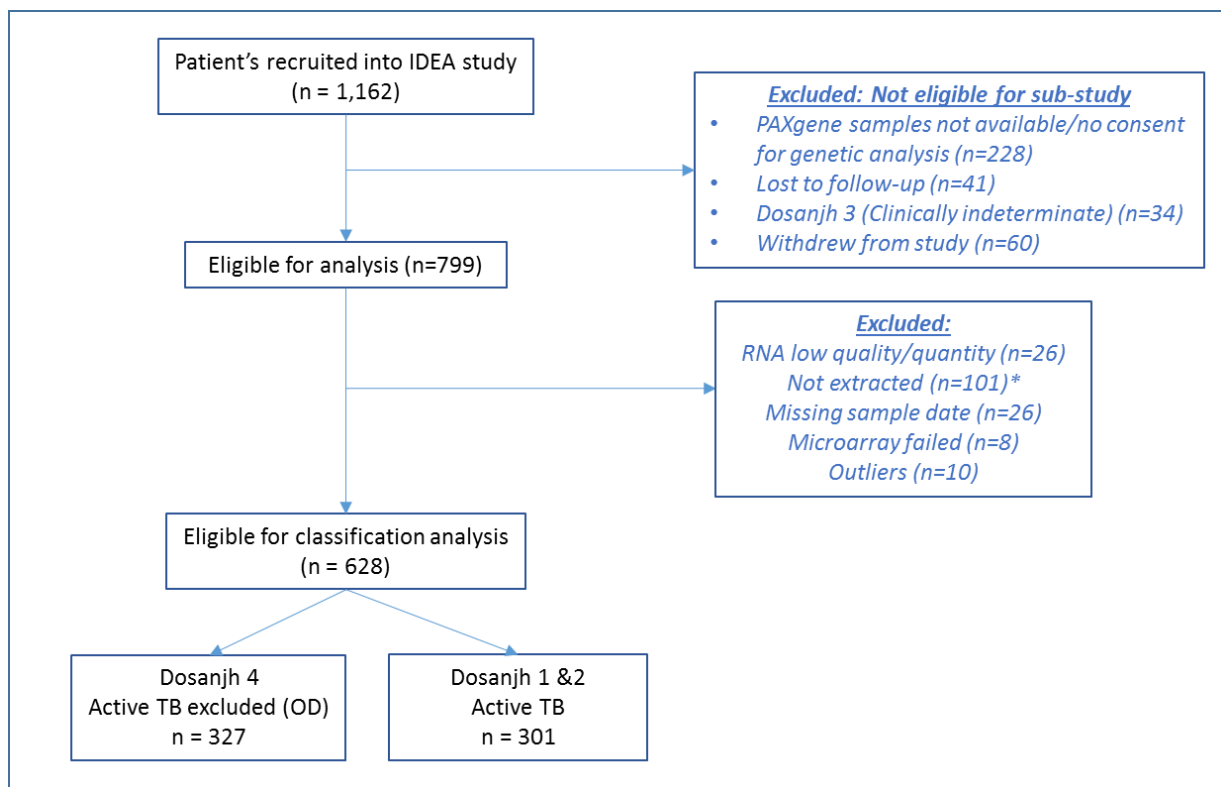


FIGURE 2. Study flow chart for validation of Transcriptomics signatures. From the full cohort of 1,162 patients recruited into IDEA, 628 were selected for RNA extraction. Of note, the rate of active TB amongst the HIV + cohort of patients in IDEA study was low (~14%). Therefore, all the eligible HIV+ active TB cases (n=20) were selected for cohort inclusion, with a similar number of HIV+ OD participants (n=25) randomly selected from the 146 HIV+ participants available.

The main demographic characteristics of the transcriptomics validation cohort are presented in *TABLE 3* below. Of all the TB cases included, 29.6% were culture unconfirmed (highly probable TB/Dosanjh 2), which reflects the proportion of this group in the full IDEA cohort of TB patients (where 28.1% TB cases were highly probable). As in the overall IDEA cohort, significant differences were observed between patients groups, with TB patients being

younger, weighing less and having lower BMI, and more likely to be of Indian-subcontinent ethnicity (includes Indian, Bangladeshi, Pakistani, and Sri Lankan).

TABLE 3. Demographic characteristics of transcriptomic validation cohort (n=628) of patients selected from the IDEA study cohort. The full cohort is divided into the three diagnostic categories as outlined in the Dosanjh reference standard (i.e culture confirmed TB, highly probable TB and active TB excluded (referred to here as ‘Other Diseases’).

Demographic characteristic	Culture-confirmed TB (Dosanjh 1)	Highly probable TB (Dosanjh 2)	Other Diseases (Dosanjh 4)	p.overall*
	N=212	N=89	N=327	
Age (years), median [range]	32.0 [16.0-81.0]	36.0 [18.0-76.0]	43.0 [17.0-87.0]	<0.001
Male, n (%): Male	149 (70.3%)	45 (50.6%)	189 (57.8%)	0.001
Ethnic origin, n (%):				<0.001
Asian	12 (5.66%)	3 (3.37%)	12 (3.67%)	
Black	36 (17.0%)	17 (19.1%)	62 (19.0%)	
Hispanic	1 (0.47%)	0 (0.00%)	1 (0.31%)	
Indian Sub-continent	139 (65.6%)	61 (68.5%)	150 (45.9%)	
Middle Eastern	3 (1.42%)	0 (0.00%)	11 (3.36%)	
Mixed	2 (0.94%)	0 (0.00%)	7 (2.14%)	
Unable/unwilling to respond	0 (0.00%)	0 (0.00%)	1 (0.31%)	
White	19 (8.96%)	8 (8.99%)	83 (25.4%)	
Height (m), median [range]	1.69 [1.37-1.96]	1.67 [1.47-1.98]	1.70 [1.32-1.96]	0.305
Height, n (%): Missing	94 (100%)	33 (100%)	106 (100%)	.
Weight (kg), median [range]	63.1 [34.7;127]	63.0 [40.4;116]	68.7 [41.4-140]	<0.001
Weight, n (%): Missing	6 (100%)	4 (100%)	24 (100%)	.
BMI (kg/m ²), median [range]	22.6 [15.7-48.5]	22.5 [15.6-42.2]	24.4 [14.8-47.2]	0.005
BMI, n (%): Missing	96 (100%)	36 (100%)	114 (100%)	.

*Where the data was categorical, a chi-squared tests was used to compared characteristics between groups; where the data was continuous the groups were compared using a Kruskal Wallis test.

The main clinical features of the transcriptomic signatures validation cohort are presented in TABLE 4 below. There was a higher proportion of HIV+ individuals in the highly probable TB group when compared to the culture-confirmed TB and OD groups. Since the number HIV+ OD cases were higher in the overall IDEA cohort, the HIV+ rate we reported within this sub-cohort does not reflect the parent population. However, all HIV+ TB patients were included. Therefore, the higher rate of HIV infection in the highly probable TB group reflects the comparatively paucibacillary nature of active TB cases in cases of HIV co-infection, resulting in a lower rate of culture confirmed cases compared to HIV- individuals.

TABLE 4. Clinical characteristics of transcriptomics validation cohort, stratified by patient final diagnosis.

Clinical characteristic	Culture-confirmed TB (Dosanjh 1)	Highly probable TB (Dosanjh 2)	OD (Dosanjh 4)	p.overall*
Number	N=212	N=89	N=327	N/A
Clinical setting, n (%):				0.74
Unknown	1 (0.47%)	1 (1.12%)	1 (0.31%)	
Inpatient	68 (32.1%)	31 (34.8%)	103 (31.5%)	
Outpatient	143 (67.5%)	57 (64.0%)	223 (68.2%)	
BCG vaccinated, n (%): Yes	126 (59.4%)	46 (51.7%)	191 (58.4%)	0.439
BCG scar visible, n (%):				0.632
Missing	49 (23.1%)	19 (21.3%)	75 (22.9%)	
No	13 (6.13%)	4 (4.49%)	16 (4.89%)	
Unsure	10 (4.72%)	8 (8.99%)	30 (9.17%)	
Yes	140 (66.0%)	58 (65.2%)	206 (63.0%)	
HIV positive, n (%):	9 (4.2%)	11 (12.4%)	25 (7.6%)	0.04
TB contact, n (%): No	160 (75.5%)	72 (80.9%)	257 (78.6%)	0.527

* Where the data was categorical, a chi-squared tests was used to compared characteristics between groups; where the data was continuous the groups were compared using a Kruskal Wallis test.

The distribution of severity/type of TB and distribution of alternative diagnoses in the OD group are presented in *Appendix 1*.

Diagnostic accuracy analysis

For the validation of published signatures, there was no missing data for any of the genes.

Performance of ‘score’ based signatures in diagnosing active TB

Three of the published signatures presented a score (referred to in the publications as either disease risk score (DRS) or TB score (TBS) as previously described^{19, 28, 31}) which is calculated using the expression levels of the signature genes. We were able to apply the same score based approach to testing these genes signatures within the signatures in our cohort. Cross tabulation of the score value distributions against diagnostic categories according to the reference standard (Dosanjh criteria) are given in *TABLE 27* in *Appendix 4*.

Analysis of performance of score based signatures in full cohort

The distribution of score values in TB and OD cases in the overall cohort, as well as within either the HIV+ or HIV- cohort is shown in *FIGURE 16* in *Appendix 4*; the ROC curves showing diagnostic performance are also shown.

1 The TB group had significantly higher score values than the OD group for all three signatures
2 when assessed within the full transcriptomic cohort, with the difference being lowest for the
3 Anderson DRS. The Kaforou 44 gene signature and the Sweeney 3 gene signature performed
4 comparably, and both better than the Anderson signature in the overall and HIV+ cohorts,
5 with AUC values between 0.82 (0.78-0.85) and 0.84 (0.80-0.86).

6 Interestingly, while the scores were all significantly higher in TB compared to OD within the
7 HIV- population, there were no significant differences between scores in these groups within
8 the HIV+ population (see *FIGURE 16C* in *Appendix 4* and *TABLE 5*). This was reflected in
9 a poor diagnostic performance (AUC 0.50-0.69) for all three signatures in the HIV+
10 population, and the observation that when the HIV+ population was excluded from the
11 overall cohort, the AUC for each signature improved slightly (*TABLE 5*).

12 The full diagnostic performance of the ‘score’ based signatures in full, HIV- and HIV-
13 populations is also presented in *TABLE 5* below, including a range of sensitivity and
14 specificity values selected by varying the cut-off values for each score based test based on
15 pre-defined criteria. The PPV and NPVs are also given.

1
2
3

TABLE 5. Performance of published gene expression score signatures in full cohort, stratified by HIV infection status. Different cut-offs were given, selected to either:
*to provide sensitivity >95%, ^{\$} to provide specificity >90%, ^Y to provide the optimal Youden index.

Signature, ref	HIV	TB (n)	OD (n)	AUC (95% CI)	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV
DRS (Anderson)	-	281	302	0.70 (0.66-0.75)	>86.0* >100.6 ^{\$} >91.4 ^Y	95.0 (91.8-97.3%) 30.3 (24.9-36%) 75.4 (70-80.4%)	23.8 (19.2-29.1%) 90.0 (86.1-93.2%) 53.6 (47.8-59.4%)	0.48 0.70 0.55	0.86 0.63 0.74
	+	20	25	0.50 (0.33-0.68)	<105.6 <84.2 ^{\$} >97.9 ^Y	95 (75.1-99.9%) 5 (0.13-24.9%) 70 (45.7-88.1%)	8 (0.98-26%) 92 (74-99%) 44 (24.4-65.0%)	0.44 0.32 0.49	0.68 0.56 0.66
	-/+	301	327	0.68 (0.64-0.72)	>86.0* >100.8 ^{\$} >91.4 ^Y	95.0 (91.9-97.2%) 28.2 (23.2-33.7%) 75.4 (70.2-80.2%)	22.9 (18.5-27.9%) 90.2 (86.5-93.2%) 52 (46.4-57.5%)	0.48 0.68 0.54	0.86 0.62 0.74
DRS (Kaforou)	-	281	302	0.83 (0.79-0.86)	>101.1* >113.4 ^{\$} >108.6 ^Y	95.0 (91.8-97.3%) 57 (51-62.8%) 74.7 (69.2-79.7%)	32.8 (27.5-38.3%) 90.4 (86.5-93.5%) 78.8 (73.8-83.3%)	0.52 0.82 0.73	0.90 0.74 0.81
	+	20	25	0.69 (0.53-0.85)	>104.3* >124.2 ^{\$} >120.2 ^{Y1} >113.2 ^{Y2}	95 (75.1-99.9%) 35 (15.4-59.2%) 45 (23.0-68.5%) 65 (40.8-84.6)	36 (18-57.5%) 92 (74-99%) 88 (68.8-97.5%) 68 (46.5-85.05%)	0.53 0.77 0.74 0.61	0.91 0.65 0.68 0.72
	-/+	301	327	0.82 (0.78-0.85)	>101.1* >114.7 ^{\$} > 108.6 ^Y	95.0 (91.9-97.2%) 53.2 (47.4-58.9%) 74.4 (69.1-79.3%)	31.8 (26.8-37.2%) 90.2 (86.5-93.2%) 76.8 (71.8-81.2%)	0.51 0.80 0.71	0.89 0.72 0.80
TBS (Sweeney)	-	281	302	0.84 (0.81-0.87)	>-4.05* >-2.60 ^{\$} >-3.2 ^Y	95.0 (91.8-97.3%) 56.2 (50.2-62.1%) 79.4 (74.2-83.9%)	34.4 (29.1-40.1%) 90.0 (86.1-93.2%) 76.8 (71.7-81.5%)	0.52 0.81 0.72	0.90 0.73 0.83
	+	20	25	0.64 (0.47-0.81)	>-4.51* >-1.69 ^{\$} >-2.49 ^Y	95 (75.1-99.9) 15 (3.2-37.9) 55 (31.5-76.9)	0 (0-13.7) 92 (74-99) 80 (59.3-93.1)	0.42 0.59 0.67	0 0.59 0.70
	-/+	301	327	0.83 (0.80-0.86)	>-4.1* >-2.5 ^{\$} >-3.2 ^Y	95.0 (91.9-97.2%) 55.2 (49.3-60.9%) 78.4 (73.3-82.9%)	29.4 (24.5-34.6%) 90.2(86.5-93.2%) 75.5 (70.5-80.1%)	0.50 0.81 0.71	0.89 0.73 0.83

The poor performance of these signatures within the HIV+ population suggests that these signatures are unable to detect TB in this population. However, the variability in the diagnostic accuracy scores demonstrates uncertainty in the estimates and reflects the low numbers of patients.

The performance of score signatures by Dosanjh classification

The effect of culture status within the TB group, as well as the likelihood of LTBI within the OD group, on the ‘score’ signature results within the full cohort can be assessed by comparing performance across the final Dosanjh categories *see FIGURE 3*.

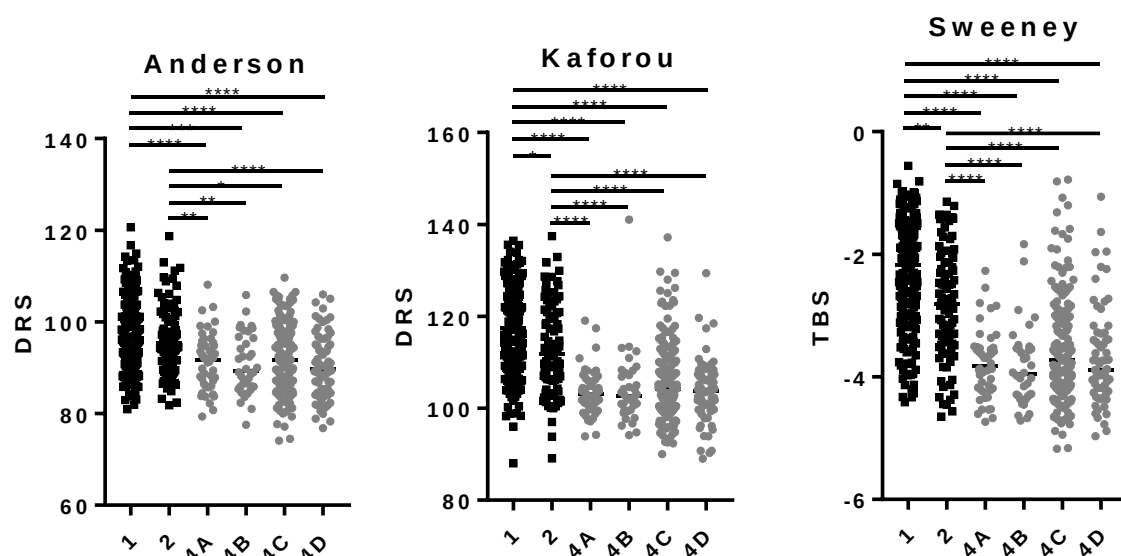


FIGURE 3. The performance of previously published score signatures, stratified by Dosanjh category. Statistical differences between score values between groups were determined using the Kruskal-Wallis with Dunn’s post hoc test for multiple comparisons (* $p < 0.05$; ** $p < 0.005$; * $p < 0.0005$; p**** $p < 0.0001$).**

In all cases, the score signatures were marginally higher (significantly with the Kaforou and Sweeney signatures) in the culture confirmed TB cases (Dosanjh 1) than the highly probable TB cases (Dosanjh 2), and as such, the diagnostic accuracy (AUC values) for culture-confirmed TB was slightly better than for highly probable TB for all score signatures (*Appendix 3*). The maximum AUC for the detection of culture confirmed TB was using the Sweeney TBS where an AUC of 0.86 (0.83-0.89) was achieved (Kaforou DRS gave similar

performance of 0.85 (0.81-0.88)), see *TABLE 28*. For detecting highly probable TB from OD, the highest AUC achieved was 0.75 (0.70-0.81), again with the Sweeney TBS.

No differences were observed in score signatures between the subgroups of OD patients in *Dosanjh 4*, suggesting that neither previous history of TB/inactive TB (not to be confused with diagnosed LTBI), nor stratified risk of LTBI within the OD population affected the outcome of these score signatures. In fact, the highest AUCs were reported when the 4A and 4B groups (highest risk of having LTBI) were classified from all TB cases (*FIGURE 17B*), the opposite direction to the expected. The full diagnostic accuracy data are shown in *TABLE 28* in *Appendix 3*, including additional analyses with the HIV+ excluded from the cohort. As expected given the poor diagnostic performance within the HIV+ population, removing this group of patients resulted in slightly improved diagnostic performance (see *FIGURE 17*).

The performance of score signatures in detecting pulmonary or extrapulmonary TB

The clinical presentation, management and outcome of PTB and EPTB are different, and the latter is typically harder to diagnose due to several factors including difficulty of obtaining biopsy samples. We therefore wished to compare the score values in these groups and determine the diagnostic performance in either of these sub-classifications of TB. As with the previous analyses, the Kaforou and Sweeney score signatures performed similarly classifying either PTB or EPTB, while the Anderson signature had lower AUC values (0). The optimal performance for diagnosing either PTB or EPTB was comparable, with highest AUCs being 0.85 (0.81-0.89) (achieved by both Kaforou DRS and Sweeney TBS) and 0.84 (0.80-0.87, Sweeney TBS) respectively.

In addition, we stratified the OD group by the most prevalent disease types, so that we could assess for any differences within this heterogeneous population of TB suspects with a wide array of ODs. Interestingly, the only condition within OD that consistently did not differ significantly to either type of TB (PTB or EPTB) across all three signatures was Sarcoidosis (*FIGURE 180B,C*), suggesting that this specific differential diagnosis is a major confounding factor for the optimal performance of these score based signatures in a real life clinical setting.

Performance of score signatures in detecting TB in smear negative TB suspects

In those tested for AFB using Smear microscopy within the IDEA population (76% of IDEA cohort), the test provided high specificity (96.7%) and poor sensitivity (23%) for all TB, and thus functions as a reliable cheap and fast rule-in test for (mostly) pulmonary TB (which, when smear-positive indicates high bacterial burden) in the UK setting. This allows clinicians to start treatment for this important infectious sub-group of TB patients very quickly. However, there is currently lacking a rapid and reliable rule-in or rule-out test with which to assess for TB in the ~77% of smear-negative (paucibacillary) TB suspects who present in secondary care, since culture confirmation can take several weeks. We therefore assessed the performance of the score signatures within this smear-negative population of TB suspects, and using the prevalence data from IDEA as the pre-test probability, calculated the PPV and NPV values for these tests. In *TABLE 6* below we present the performance of these scores in the smear negative population of TB suspects.

TABLE 6. Performance of score signatures in the detection of sub-groups of TB cases, within the smear negative population of TB suspects. The diagnostic performance is presented as follows: area under the curve (AUC), sensitivity, specificity, PPV and NPV. *The optimal score cut-off values for providing optimal sensitivity and specificity were calculated using the Youden index

	Anderson	Kaforou	Sweeney
<i>Performance in Smear negative TB (Dosanjh 1 and 2) (n=198 TB vs n=244 OD)</i>			
<i>Prevalence of TB in all smear negative TB suspects= 46% (From overall IDEA population)</i>			
AUC (95% CI)	0.67 (0.61-0.71)	0.79 (0.75-0.83)	0.80 (0.76-0.84)
Cut-off (Youden)	>91.78	>107.5	>-3.241
Sensitivity (95% CI)	72.2 (65.78.3)	77.3 (70.8-82.9)	78.3 (71.9-83.8)
Specificity (95% CI)	50.8 (44.4-57.3)	71.3 (65.2-76.0)	74.2 (68.2-79.55)
PPV /NPV	0.56 / 0.68	0.69 / 0.79	0.72 / 0.80
<i>Performance in Smear- culture + (Dosanjh 1) TB (n=139 TB vs n=244 OD)</i>			
<i>Prevalence of culture + TB = 33%</i>			
AUC (95% CI)	0.66 (0.61-0.72)	0.81 (0.77-0.86)	0.82 (0.78-0.87)
Cut-off	>91.78	>107.5	>-3.233
Sensitivity (95% CI)	72.7 (64.5-79.9)	82.0 (74.6-88.0)	82.7 (75.4-88.6)
Specificity (95% CI)	50.8 (44.4-57.3)	71.3 (65.2-76.9)	74.2 (68.2 – 79.6)
PPV /NPV	0.42 / 0.79	0.58 / 0.89	0.61 / 0.90
<i>Performance in detecting smear – culture negative (Dosanjh 2) (n=59 TB vs n=244 OD)</i>			
<i>Prevalence of culture unconfirmed TB in smear negative TB suspects = 13%</i>			
AUC (95% CI)	0.65 (0.58-0.73)	0.74 (0.67-0.81)	0.75 (0.68–0.82)
Cut-off	>92.41	>109.5	>-3.467

Sensitivity (95% CI)	71.2 (57.9-82.2)	62.7 (49.2-75.0)	74.6 (61.7-85.0)
Specificity (95% CI)	52.9 (46.4-59.3)	76.2 (70.4-81.4)	68.9 (62.6-74.6)
PPV /NPV	0.18 / 0.92	0.28 / 0.93	0.26/ 0.95

These results demonstrate that the performance of the Sweeney and Kaforou transcriptomic signatures within the smear negative population of TB suspects is sufficient to provide a high NPV(89-90%) for culture positive TB in a UK-like setting where the prevalence of culture positive TB in smear negative TB suspects was 33%.

Analysis of Roe signatures using published method

We obtained the SVMML model from Roe et al, and validated the signature using the approach proposed in their publication. The performance of this model for discriminating TB cases from OD in our cohort was very poor with an AUC of 0.59 (0.54 to 0.63). While sensitivity for TB was high using this model (296 of 301 cases correctly classified, sensitivity of 98.3 (96.2-99.3), specificity was very low (10/327 cases correctly classified as fever; specificity of 3.1% (1.7-5.5)).

Analysis of all previously reported signatures using LDA

To allow for a fair and comparable validation of all 6 transcriptomic signatures in our cohort, i.e. including those where the published method of diagnostic assessment was not reproducible,^{17, 29} we applied an LDA approach to assess the accuracy of each signature to each of the pre-defined comparisons we selected for validation analyses. Furthermore, this approach determined the optimal performance of each gene set from the selected signatures within our cohort, without the introduction of learning bias. Thus, the findings using this approach should be reproducible in a new, similarly designed cohort.

Assessment in the full cohort

For the distribution of these signatures by probability as calculated by LDA in the full cohort, and diagnostic performance for detecting all TB cases (including stratification by HIV infection status), see 0.

Interestingly, unlike with the score values where the Anderson DRS underperformed compared to the Kaforou and Sweeney signatures (although no formal statistical comparison

was made), the six signatures performed with similar accuracy when assessed using the LDA approach, with AUCs between 0.81 and 0.84 for detecting all TB cases in the full cohort. The accuracy marginally increased to AUC=0.85 when HIV+ cases were excluded. The performance in the HIV+ population could only be assessed for these signatures using this method (Maertzdorf, Roe, Sweeney), since the other signatures used a larger set of genes, and in all cases the performance was poor and highly variable, with no significant differences between probability scores.

Performance in detecting culture confirmed or highly probable TB

The performance of all six signatures at detecting either culture confirmed (Dosanjh 1) or highly probable (Dosanjh 2) TB was assessed using the LDA approach, see *TABLE 30* in *Appendix 5*. For detecting culture confirmed TB, the performance of all six signatures was comparable, with AUCs between 0.84 and 0.87. However, for detecting highly probable TB, there was a wide range in performance, with the Sweeney signature providing the highest test accuracy (AUC=0.80).

Performance in detecting either PTB or EPTB

When the six signatures were compared using LDA for the detection of either PTB or EPTB, comparable test accuracies were found for all signatures, with detection of PTB being marginally more accurate (maximum AUC=0.86) than for EPTB (maximum AUC=0.85); see *TABLE 32*.

Derivation of novel signatures

To address our secondary objective to derive novel transcriptomic signatures to detection of TB, we plan to split the validation cohort into a training and test set (80/20 % split) and to using a variable selection method to identify our candidate genes within the training set, train them, and finally test their performance on the remaining 20%. We are in the process of doing this analysis, but did not fit within the scope of this report and will therefore be published elsewhere.

Discussion

In this study, we have shown that none of the previously identified transcriptomic signatures for detecting active TB are sufficiently accurate for the detection of all TB in this clinically relevant cohort. Indeed, neither are the signatures sufficiently accurate to detect any of the major sub-groups of TB, even when considering the culture confirmed TB cases alone, which are typically the most straightforward to identify. Their performance was much lower than previously reported, which likely reflects differences in study cohorts, since most of the previous studies did not include the full spectrum of TB cases and their differential diagnoses.

Assessment of the previously proposed signatures in an unbiased and reproducible way, by using either summary scores,^{19, 28, 31} an SVMML model as previously reported,³⁰ or by applying an LDA approach, allowed us to assess performance in key patient groups in our validation cohort. While diagnostic performance amongst culture confirmed TB cases was greater than for highly probable TB cases, neither group could be accurately distinguished from the OD group. For highly probable TB, the Sweeney signature performed the best when compared to other published signatures, although only marginally, and no statistical comparison was made. It is of note that the Anderson TBS underperformed when compared to Kaforou DRS, whilst learning to classify using respective signatures with LDA their performance was comparable. This likely reflects the nature of the derivation cohorts on which the score values were derived, with Anderson deriving the signature from a paediatric cohort²⁸.

However, since there is currently no test for rapidly ruling in the highly probable TB group, it is still possible that a whole-blood transcriptomic signature test may be useful for assessment of TB in those cases where a negative microbiological test result is obtained. The prevalence of highly probable TB within the culture-unconfirmed cohort from IDEA was found to be 17.5%. Thus, we are able to assess the predictive performance of transcriptomic tests for highly probable TB in those with a negative culture test for TB.

The best performing previously identified signature was the Sweeney signature, which gave an AUC 0.80 (0.74-0.85) for detecting highly probable TB cases when assessed using LDA on the full cohort, with a sensitivity of 82.1% and specificity of 73.8% when a balanced cut-off was selected. The scenario in which such a test would be useful is in a TB suspect where the culture (or other microbiological test, e.g. Gene Xpert Ultra) is negative or not available. The prevalence of highly probable TB in the culture unconfirmed TB suspects recruited into

the IDEA study was 17.5%. Using this prevalence value, the Sweeney gene signature for highly probable TB would provide a poor PPV of 0.4, but a very high NPV of 0.95. As such, it could be that such a transcriptomic signature, if developed into a simple and cheap rapid blood test and validated in new cohorts, could provide a sufficiently high NPV to be considered as a useful rule-out test in the microbiologically unconfirmed population. Furthermore, in settings with higher prevalence rates of highly probable TB, the PPV of these tests would be higher.

A rapid test with high diagnostic accuracy would be particularly helpful in the context of patients who are smear negative upon initial assessment for TB, since culture takes a long time to give a definitive result. We present data on the performance of the score signatures (Anderson, Kaforou and Sweeney) in this sub-cohort of patients, both the Sweeney and Kaforou signatures provided reasonably high NPV (89-90%), which, if validated in other studies, may provide clinicians with sufficient reason to withhold treatment (or at least delay it until culture results are available) in smear negative TB who tested negative with these signatures.

The performance of all signatures was very poor within the HIV+ population in this cohort. This suggests that HIV co-infection masks the transcriptomic signal which differentiates TB from OD in the wider population. However, since the number of HIV+ participants in our cohort was low, this study was likely to be underpowered to validate performance within this population. Nevertheless, it is intriguing that others have shown that their identified signatures can perform with comparable test accuracy for detecting TB from OD in HIV+ and HIV- populations.^{19, 31}

We have shown data that suggests that the performance of signatures for detecting TB from OD was compromised by Sarcoidosis as well as HIV co-infection. If Sarcoidosis could be excluded early in the diagnostic work up of active TB, then the performance of transcriptomic signatures would improve and may potentially be of use in specific patient sub-groups.

Our transcriptomic validation sub-study has some limitations. We used microarray to determine the transcriptional profile of the whole blood samples used in this cohort, and this technology has now more or less been replaced by RNA-Seq. Nevertheless, since the previously published signatures have all been generated by microarray, this was the appropriate platform to use. It is generally accepted that clinical diagnostic scores over

1 diagnose TB in children, but to which extent this occurs in adult TB is unknown³⁷. It is
2 possible that the presence of patients without TB in the highly probable group negatively
3 affect the transcriptomic signatures' performance. Thus these signatures could still be
4 clinically useful for paucibacillary (i.e. smear negative and Gene Xpert negative) if these
5 cases are removed from the analysis. In future studies, we aim to use this novel dataset to
6 derive and test new signatures, with a particular focus on identifying signatures with high
7 diagnostic performance for smear negative, culture negative and EPTB cases of TB.

8 The suboptimal performance of the published signatures and the newly derived signatures
9 highlights the unmet need in TB diagnosis which is a rapid, accurate and generalisable test of
10 the full spectrum of TB including culture negative TB. Although a similar cohort has been
11 used to derive signatures of paediatric TB²⁸, our study is the first one to address it in a large
12 clinical adult TB cohort.

Chapter 4. Proteomics

Introduction

Serum protein biomarkers have the potential to be used as rapid, cheap and easy to use tests to rule-in or rule-out tests for active TB disease. The ‘proteomic signatures’ measure the changes in individual proteins found in the serum due to the host response to disease. Many of the changes observed are not unique to one condition, as they reflect overlapping strategies to combat or adapt to disease processes. However, certain proteins when used in combination can provide high level of diagnostic accuracy when applied in the correct clinical situation. Therefore, the cohort used for selection and testing of proteomic signatures needs to be appropriate to the clinical question and condition for which it is being developed.

Proteomic biomarkers – current methodologies

Proteins in sera are present in a dynamic range that spans at least 4 orders of magnitude, and accurate detection of all proteins and their isoforms has not yet been achieved by any technology. In order to overcome this limitation, many proteomic signatures have been generated after immune-depletion of the most abundant proteins. This assumes that depleted proteins are not useful biomarkers and for example will often remove serum albumin and IgG (accounts for >90% of the total protein mass). Depletion allows accurate quantification of lower abundant proteins, but it can alter the abundance of proteins that interact with the proteins being depleted. SELDI-TOF provides a quick, reproducible and cheap profiling technology for intact proteins, but is limited by dynamic range and does not provide identification of the protein signatures. LC-MS is a newer technology and provides both accurate quantification and protein identification, but is slow and costly compared to SELDI and ELISA. ELISA is fast and cheap, and highly suitable for development of a point of care test for TB. It is however limited to the quantification of only a few proteins, where distinct antibody pair have been developed and tested. In ELISA the identity of the signature is required prior to being the investigation and thus it requires the identification of any signature developed by another technology to be known.

To investigate the utility of the serum proteomic signatures to diagnose TB we reviewed the most important and relevant studies carried previously.

Previously published signatures

We selected 12 published studies reporting a proteomic signature for detection of TB cases and summarised the key study characteristics, see *TABLE 7*.

TABLE 7. Summary of publications presenting proteomic signatures for detecting of active TB cases.

First Author, Year, ref	Cohort (in validation if done)	Classification accuracy	Technology	Suitable for validation, why
Agranoff, 2006 ³⁸	TB=179, 86-87% smear positive ORD=170	AUC of 0.96	SELDI-TOF	NO, failed validation
Ratzinger, 2012 ³⁹	?	AUC=54%	SELDI-TOF	NO, low performance
Lui, 2010 ⁴⁰	ATB=87 HC=55 ORD=13	93.55%, 95.46%, and 88.09%	SELDI-TOF	NO, Healthy controls
Zhang, 2012 ⁴¹	TB=129 ORD=69 HC=60	Sensitivity of 96.9% and a specificity of 97.8%.	SELDI-TOF	NO, healthy controls
Sandhu, 2012 ⁴²	TB=151 ORD=110	84% sensitivity, 90% Specificity and an AUC 0.93	SELDI-TOF	YES, no validation done.
Hamilton/Levin (unpublished).	TB=300 ORD=300	AUC= 84.7, CI 0.7-0.9 or AUC=81.9 CI**	SELDI-TOF/ ELISA	YES, validation not done, ELISA signature available
Liu, 2013 ⁴³	TB=180 ORD=121 HC=90	accuracy of 80.1% (75.0% sensitivity, 83.5% specificity	SELDI-TOF	NO, low performance and HC in control group
Song, 2014 ⁴⁴	TB=26 ORD=31	AUC=0.947 (CI 0.8-1)	LC-MS ELISA (AUC)	YES
Xu, 2015 ⁴⁵	PTB=40, Pneumonia n=40 lung cancer n=40 and healthy controls n=40	TB v pneumonia AUC=0.955 v lung cancer AUC=0.954	LC-MS ELISA (AUC)	YES
Li, 2015 ⁴⁶	TB= 125 pneumonia =15 healthy controls=32	AUC=0.904	LC-MS ELISA (AUC)	NO, Healthy controls in validation
Achkar 2015 ²⁰	TB n=28 ORD n=45	HIV- AUC 0.96 and HIV+ 0.95).	LC-MS (AUC)	YES
Chegou, 2016 ⁴⁷	TB n=214 ORD n=487	Sensitivity of 81.3% (95% CI 69.2% to 89.5%) and specificity of 79.5% (95% CI 71.8% to 85.5%)	ELISA	NO
De Groote, 2017 ¹⁸	TB=92 ORD=123	AUC of 0.87 (95% CI, 0.81, 0.91).	SOMA SCAN	NO (technology no available)

In 2014, when the VANDET proposal was funded, the majority of signatures generated were based upon SELDI-TOF analysis. Of note the work by Agranoff et al in 2006 in a large prospective cohort identified a signature with a sensitivity of 93.5% and specificity of 94.9%, (AUC of 0.96), this signature was validated in a cohort of mainly pulmonary TB and gave

1 reduced sensitivity of 88·9% and specificity of 77·2% .³⁸ Ratzinger 2012 also validated this
2 signature externally in a UK cohort which was diverse in the presentation of the site of TB
3 disease including more EPTB. In this report the accuracy reduced to 54% (CI: 47%-61%).³⁹
4 While the latter reported that accuracy of the protein markers could be improved using
5 additional clinical data, the poor performance of the protein signature as a test meant it will
6 not be considered an important test to assess within VANTDET.

7 More recently two published SELDI profiles were identified for ATB compared to healthy or
8 mainly healthy control cohorts making them not useful for clinical validation.^{41, 43} Most
9 recently Sandhu et al., published a novel 54 m/z cluster using SELDI-TOF, that performed
10 with high accuracy (84% sensitivity, 90% Specificity, AUC 0.93) this signature remains
11 unvalidated but was generated using a prospective cohort of mainly pulmonary TB and other
12 non-TB diseases and therefore warrants assessment within VANTDET.⁴²

13 In a larger cohort (n=600) our collaborators (SH, LC, ML) have identified a 4 peak/protein
14 signature with excellent performance in paediatric TB (AUC= 84.7, CI 0.7-0.9 or AUC=81.9
15 CI) depending whether the data was analysed using a disease risk score (DRS) or the LEAPS
16 forward selection methods. Hamilton et al., have also successfully identified all four proteins
17 in the cohort and are currently developing and validating this as an ELISA for point of care
18 testing.

19 LC-MS has become the method of choice for biomarker discovery and we identified 4 studies
20 each identifying a signature for active TB using this methodology. Song et al., identified
21 SERPINA1 (also known as alpha-1-antitrypsin) as a single and highly accurate marker of TB
22 (AUC=0.947 (CI 0.8-1)). They performed no validation and their choice of disease controls is
23 highly suspect, choosing diseased patient that don't exhibit symptoms of active TB.⁴⁴

24 More promisingly, Xu et al., pooled immunodepleted serum and used LC-MS to select detect
25 a 3 protein signature (S100A9, SOD3, and MMP9) which could discriminate active TB from
26 controls. They validated these proteins using ELISA in case control cohort but included
27 diseases that often confound diagnosis of ATB (PTB n=40, Pneumonia n=40, lung cancer
28 n=40 and healthy controls n=40).⁴⁵ This signature could resolve pulmonary TB and
29 pneumonia (AUC=0.955) and discriminate between pulmonary TB and lung cancer
30 (AUC=0.954) with high accuracy, however it was limited in terms of the spectrum of disease
31 sub types included for the ATB group (all culture confirmed PTB) and other none TB disease
32 (limited to two disease types).

The findings from the IDEA study demonstrate that the differential diagnoses for active TB are numerous and varied (over 200 final diagnoses were reported in the other disease group), reflecting the heterogeneous nature of different types of TB disease and the various clinical presentations that are associated with them.¹¹ Achkar 2015 took LC-MS a step closer to a real-life road when they recruited patients from 4 different hospitals including both culture negative and HIV positive patients and a verification cohort of TB cases and OD with TB like symptoms. The signature differed depending on HIV infection status, and gave excellent classification of ATB from other non-TB like disease (HIV- AUC=0.96 and HIV+ AUC=0.95).²⁰ The progression shown by this work demonstrates the potential of a proteomic signature detected by LC-MS in active TB diagnosis.

In the last year a new technology has been utilised to generate a signature for active TB: SOMA-SCAN. This technology relies upon aptamers and combines the benefits of both LC-MS and ELISA into one assay that is potentially customisable into a point of care test.¹⁸ However the commercial availability is limited new and not available to be investigated in VANTDET. In work by De Groote published this year they discover a novel 6 Proteins signature (SYWC (WARS), kallistatin (SERPINA4), C9 (C9), gelsolin (GSN), testican-2 (SPOCK2), and aldolase C (ALDOC) that provides a high classification accuracy AUC of 0.87 (95% CI, 0.81, 0.91) in culture confirmed TB.¹⁸ The cohort tested did not include culture negative TB but did include TB and ORD with similar symptoms and HIV coinfection.

Aim

Our aim was to assess the clinical utility of proteomic signatures for the diagnosis of active TB. We choose five signatures for validation: two identified by SELDI-TOF (Sandhu et al., and Hamilton et al.,) and three by LC-MS (Song, Xu and Achkar). All five were chosen because signatures were developed in a cohort of prospectively recruited patients with OD.

Objectives

1. Assess the utility of the SELDI-TOF 4 protein signature in the IDEA cohort
2. Bridge signatures derived by SELDI-TOF with LC-MS technology
3. Assess the utility of all signatures using LC-MS

Methods

Cohort selection

A cohort of 171 samples were selected randomly from the HIV negative patients of eligible individuals within the full IDEA cohort. The sera was processed in blocks of 9 patients and a pooled serum sample was created by mixing equal volumes patient sera. The internal control was included in each batch processed by LC-MS. SELDI-TOF mass spectrometry was performed in one batch on a total of 90 serum samples, since this was the maximum number achievable given the limited number of chips available (see issues in Chapter 2). The 90 patients used were a direct subset of the 171 individuals eligible for analysis, allowing for a bridging assessment between technologies. During experimentation and analysis samples were excluded if consent was withdrawn, the patient was found to have not been followed up by clinical assessment, or errors occurred in experimental processing. The final cohorts consisted of SELDI-TOF (n=86) and LC-MS (n= 166).

Serum samples

In IDEA, blood for serum collection was collected into a BD serum tube and allowed to clot for 60 min before centrifuging at 1000 g for 10 minutes at room temperature (RT) to remove cell debris and clots. Serum supernatant was aliquoted into sterile 2 ml tubes and kept at -80 until use.

SELDI-TOF

Protein profiling of serum by Surface-enhanced laser desorption ionisation time-of-flight (SELDI-TOF) mass spectrometry was achieved by placing serum samples on anionic (Q10) cationic (CM10) and immobilised metal affinity (IMAC) ProteinChip arrays (Bio-Rad Laboratories). The arrays were primed twice with binding buffers for Q10 (50mM Tris, 0.1 % Triton X-100, pH 7.5 and 9.5), CM10 (50 mM Ammonium acetate, 0.01% Triton X-100 (TX-100), pH 4.0 and 6.0) and IMAC Cu²⁺ (0.1M sodium phosphate, 0.5M NaCl, pH 7.0) and incubated at room temperature (RT) with shaking for 5 min. A 1:10 dilution of serum in binding buffer was then applied to the Q10 and CM10 arrays and a 1:5 dilution on the IMAC IMAC Cu²⁺ array and incubated at RT with shaking for 1 hr. The arrays were washed twice with binding buffer and deionized H₂O and dried at room temperature for 15 min. Saturated sinapinic acid (SPA; 0.7 µl) was applied twice to each spot on the CM10 and IMAC arrays and 50% saturated SPA on the Q10 array, allowing the matrix to air dry between each application. Time-of-flight spectra were generated using a PCS-4000 mass spectrometer

(Bio-Rad). Low range spectra (mass/charge (m/z) ratio 0 - 20,000) were obtained at a laser energy of 2800- 3400 nJ with a focus mass of 6000 and the matrix attenuated to 1000. High range spectra (m/z 0 - 200,000) were obtained at a laser energy of 3800-4200 nJ with a focus mass of 30,000 and the matrix attenuated to 10,000. Ten shots were obtained per position and mass accuracy was calibrated externally using All-in-One Peptide or Protein molecular mass standards (Bio-Rad).

LC-MS

An internal standard was made by pooling equal amounts of serum from the first 100 patients. For each batch, an equal amount of serum from 9 patients and the internal standard were solubilised in 1% sodium deoxycholate, 100 mM Ammonium bicarbonate. 10 µg of protein was reduced with 10 mM DTT 15 minutes at 60 °C followed alkylation with 20 mM Iodoacetamide for 15 minutes at room temperature in the dark. Trypsin (Promega) was added at a 1:50 (enzyme:protein) ratio and digestion carried out at 37°C overnight. Peptide digests were purified using the C₁₈ S_{Top} And Go Extraction (STAGE) tips and eluted peptides were dried and labelled with 9 labels from the TMT10plex Mass Tag Labeling Kit as described in instruction with minor modifications. Peptides were dissolved in 25 µl of 100 mM TEAB and 10 µL of each label in acetonitrile and incubated for 60 min at room temperature before it was quenched with 2.5 µL of 0.5 M Hydroxylamine and combined. Samples diluted to a final acetonitrile concentration of 3 % acidified to 0.1% (v/v) trifluoroacetic acid purified again by the C₁₈ STAGE tips. Each fraction was dried completely and dissolved in 2% (w/v) acetonitrile, 0.1% (v/v) formic acid prior to LC-MS.

Mass spectrometry analysis was performed as described in Reuschl et al., 2017⁴⁸ except the resolution of MS2 scans was 35K, MS2 charge targets were limited to 1E⁵ and isolation window set to 1.5 m/z..

Data processing

SELDI-TOF spectra were analysed using ProteinChip Data manager (Bio-Rad version 4.1.0) and normalized using total ion current. Peak clusters were auto-detected using a peak threshold of 20%, a mass window of 0.3% and the raw data converted for subsequent analysis using R software.

Raw LC-MS spectra were identified and quantified using Maxquant 1.5.15 using a 1% peptide and protein FDR as described previously by PMID: 19029910 using the uniprot SwissProt database downloaded on 06/06/14 with additional common contaminant proteins

and TMT-10 plex quantitation mode enabled.. Reporter ion intensities for MS/MS scans were filtered to ensure <75% precursor isolation purity, summed and assigned to proteins based upon unique matches and parsimony as described previously. For every patient each protein was quantitated using the sum of all peptide intensities. This number was divided by the same measure for the pooled internal control to give a block normalised ratio (accounting for batch processing). This value was pre-processed by log² transformation, checked for normality and batch effects.

Statistical analysis

For assessment of performance of the 4 protein signature evaluated using the SELDI platform, the 4 proteins were identified using their mass values. The disease risk score (DRS) method^{19, 49} was used to calculate a score from the abundance data of the 4 proteins, as described previously.

To identify differentially expressed proteins, ANOVA and Welch test were calculated between all disease groups, P values were adjusted for the effect of multiple hypothesis testing using the FDR (<0.1) method.

The performance of published signatures using the depleted LC-MS dataset was assessed using either abundance data (for the Song signature containing one protein⁴⁴) or Linear Discriminant Analysis (LDA) for signatures incorporating multiple proteins, on full cohort.

To assess for diagnostic performance, either the abundance, DRS or probability values generated for each signature (depending on whether raw abundance, DRS or LDA approach was taken for analysis, respectively) were used to generate ROC-curves and determine diagnostic accuracy data (sensitivity, specificity etc).

Results

A cohort of 171 samples were selected randomly from the HIV negative patients of eligible individuals within the full IDEA cohort. SELDI-TOF mass spectrometry was performed on a total of 90 serum samples. This was the maximum number achievable given the limited number of SELDI-TOF chips available (see Chapter 2, 'Issues') . LC-MS analysis was performed on the full cohort of 171 samples, the 90 samples used in SELDI-TOF were a direct subset of the 171 individuals eligible for analysis, allowing for a bridging assessment between the two technologies *FIGURE 4*. During experimentation and analysis samples were

1 excluded if consent was withdrawn, the patient was found to have not been followed up by
2 clinical assessment, or errors occurred in experimental processing. The final cohorts with
3 complete SELDI-TOF and LC-MS data consisted n=86 for SELDI-TOF n=86 and n=166 for
4 LC-MS *FIGURE 4*.

5 We also performed an additional LC-MS experiment on a larger number of sample (n=279)
6 using crude sera, including from HIV positive patients. This experiment was sent to an
7 external contractor fingerprints (<http://proteomics.lifesci.dundee.ac.uk/>). This dataset was
8 restricted to 218 proteins and we were unable to consistently identify and quantitate the
9 protein signatures found in the studies reviewed (see *TABLE 36* in *Appendix 6*), apart from
10 SERPINA1. This dataset was therefore not utilised in biomarker validation results presented
11 below.

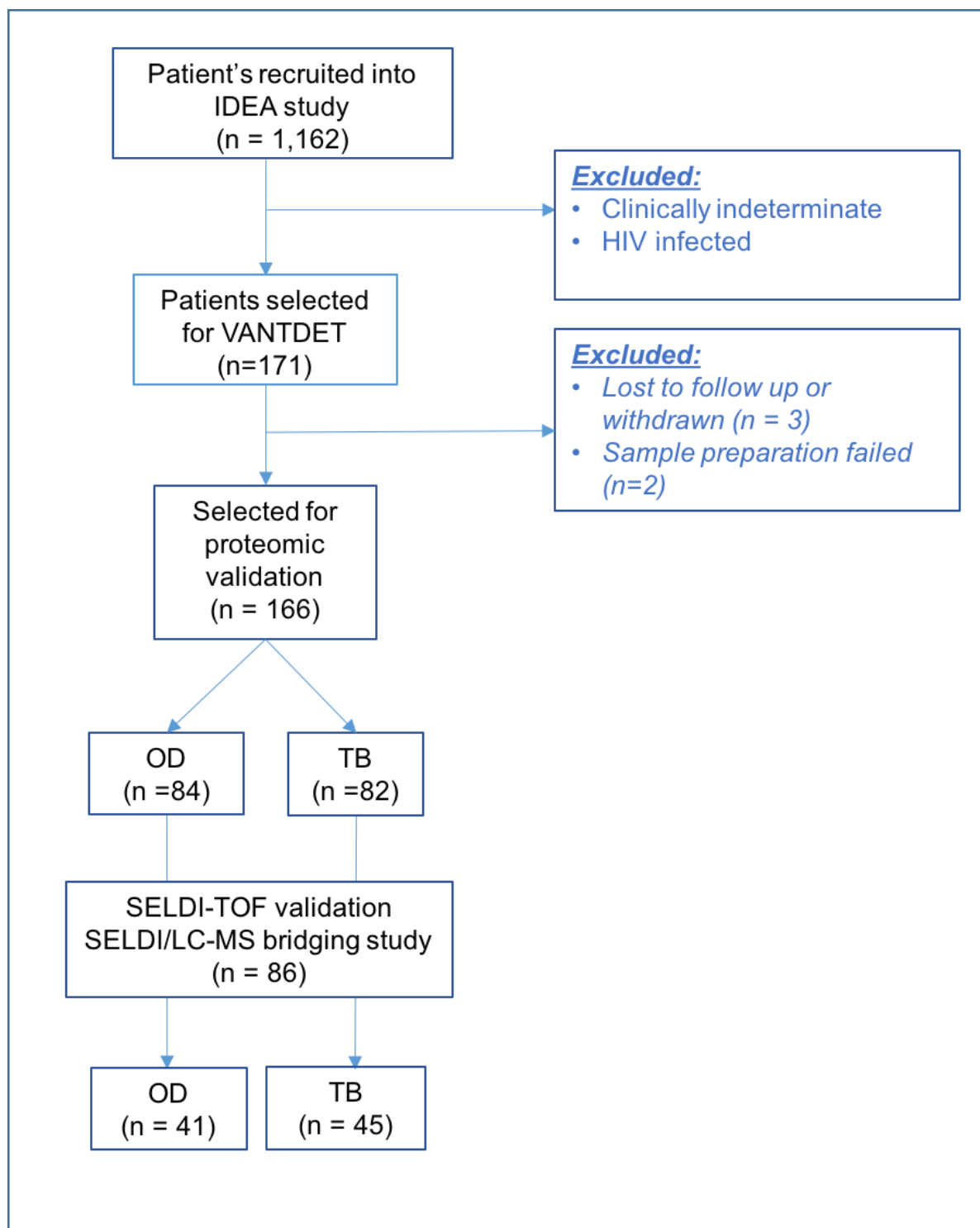


FIGURE 4. Flow diagram for the proteomics validation sub-study.

Cohort characteristics

The demographic characteristics for the proteomic validation cohorts are given in TABLE 8 (LC-MS) and TABLE 9 (SELDI-TOF). Statistical analysis of the patient demographics in the LC-MS cohort (n=166) identified differences ($p < 0.05$) between the two major diagnostic groups under investigation; with patient age, weight and BMI all lower in patients with a final

diagnosis of active TB (Dosanjh 1 and 2) compared to those classified with OD (Dosanjh 4). A similar trend was observed for these characteristics in the SELDI-TOF subset with statically significant difference in age and weight, see *TABLE 9*. No HIV infected individuals were included, because at the time of the initial experimentation there was not a validated procedure for ensuring samples were not infectious for steps required for proteomic sample processing. It was considered that the validity of a test could be adequately assessed in a HIV negative cohort.

TABLE 8. Demographic and clinical characteristics of the LC-MS proteomic signature validation cohort.

	Culture-confirmed TB (Dosanjh 1)	Highly probable TB (Dosanjh 2)	Other Disease (Dosanjh 4)	p.overall*
	N=60	N=22	N=84	
Clinical setting, n (%):				0.59
Inpatient	22 (36.7%)	7 (31.8%)	24 (28.6%)	
Outpatient	38 (63.3%)	15 (68.2%)	60 (71.4%)	
Age (years), median [range]	30.5 [17.0;72.0]	31.0 [18.0;54.0]	47.0 [21.0;85.0]	<0.001
Male, n (%): Male	37 (61.7%)	10 (45.5%)	47 (56.0%)	0.416
Ethnic origin, n (%):				0.598
Asian	3 (5.00%)	0 (0.00%)	4 (4.76%)	
Black	9 (15.0%)	3 (13.6%)	15 (17.9%)	
Indian Sub-continent	41 (68.3%)	17 (77.3%)	43 (51.2%)	
Middle Eastern	1 (1.67%)	0 (0.00%)	1 (1.19%)	
Mixed	1 (1.67%)	0 (0.00%)	3 (3.57%)	
Unable/unwilling to respond	0 (0.00%)	0 (0.00%)	1 (1.19%)	
White	5 (8.33%)	2 (9.09%)	17 (20.2%)	
Height (m), median [range]	1.69 [1.50;1.88]	1.63 [1.48;1.85]	1.70 [1.51;1.93]	0.475
Height, n (%): Missing	26 (43.3%)	7 (31.8%)	31 (36.9%)	0.578
Weight (kg), median [range]	60.0 [38.5;127]	61.4 [44.6;77.0]	68.2 [41.5;124]	0.001
Weight, n (%): Missing	2 (3.33%)	2 (9.09%)	5 (5.95%)	0.508
BMI (kg/m ²), median [range]	21.6 [15.7;40.1]	21.0 [16.0;34.6]	24.5 [17.3;36.7]	0.039
BMI, n (%): Missing	26 (43.3%)	9 (40.9%)	33 (39.3%)	0.888
BCG vaccinated, n (%): Yes	39 (65.0%)	9 (40.9%)	54 (64.3%)	0.104
BCG scar visible, n (%):				0.125
Missing	15 (25.0%)	2 (9.09%)	11 (13.1%)	
No	2 (3.33%)	1 (4.55%)	5 (5.95%)	
Unsure	3 (5.00%)	4 (18.2%)	15 (17.9%)	
Yes	40 (66.7%)	15 (68.2%)	53 (63.1%)	
HIV negative, n (%)	0 (0%)	0 (100%)	0 (0%)	.
TB contact, n (%): No	45 (75.0%)	19 (86.4%)	68 (81.0%)	0.491

* Where the data was categorical, a chi-squared tests was used to compared characteristics between groups; where the data was continuous the groups were compared using a Kruskal Wallis test.

TABLE 9. Demographic and clinical characteristics of the n=86 participants used for validation of the 4 protein SELDI-TOF signature.

	Culture-confirmed TB (Dosanjh 1)	Highly probable TB (Dosanjh 2)	Other Disease (Dosanjh 4)	p.overall*
	N=33	N=12	N=41	
Clinical setting, n (%):				0.454
Inpatient	14 (42.4%)	5 (41.7%)	12 (29.3%)	
Outpatient	19 (57.6%)	7 (58.3%)	29 (70.7%)	
Age (years), median [range]	29.0 [17.0;72.0]	29.5 [18.0;54.0]	46.0 [22.0;85.0]	<0.001
Male, n (%): Male	20 (60.6%)	6 (50.0%)	21 (51.2%)	0.68
Ethnic origin, n (%):				0.389
Asian	3 (9.09%)	0 (0.00%)	2 (4.88%)	
Black	6 (18.2%)	1 (8.33%)	9 (22.0%)	
Indian Sub-continent	21 (63.6%)	10 (83.3%)	18 (43.9%)	
Mixed	0 (0.00%)	0 (0.00%)	1 (2.44%)	
Unable/unwilling to respond	0 (0.00%)	0 (0.00%)	1 (2.44%)	
White	3 (9.09%)	1 (8.33%)	10 (24.4%)	
Height (m), median [range]	1.70 [1.50;1.88]	1.63 [1.48;1.85]	1.68 [1.52;1.85]	0.806
Height, n (%): Missing	12 (36.4%)	5 (41.7%)	16 (39.0%)	0.955
Weight (kg), median [range]	62.0 [38.5;127]	58.0 [44.6;77.0]	69.5 [41.5;96.0]	0.04
Weight, n (%): Missing	2 (6.06%)	0 (0.00%)	3 (7.32%)	1
BMI (kg/m ²), median [range]	21.4 [15.7;40.1]	20.8 [16.0;34.6]	24.6 [17.3;32.6]	0.147
BMI, n (%): Missing	12 (36.4%)	5 (41.7%)	17 (41.5%)	0.912
BCG vaccinated, n (%): Yes	19 (57.6%)	5 (41.7%)	22 (53.7%)	0.639
BCG scar visible, n (%):				0.322
Missing	10 (30.3%)	0 (0.00%)	8 (19.5%)	
No	1 (3.03%)	1 (8.33%)	2 (4.88%)	
Unsure	2 (6.06%)	1 (8.33%)	5 (12.2%)	
Yes	20 (60.6%)	10 (83.3%)	26 (63.4%)	
HIV positive, n (%): No	33 (100%)	12 (100%)	41 (100%)	.
TB contact, n (%): No	24 (72.7%)	10 (83.3%)	34 (82.9%)	0.55

Where the data was categorical, a chi-squared tests was used to compared characteristics between groups; where the data was continuous the groups were compared using a Kruskal Wallis test.

Performance of previously published signatures using SELDI-TOF

We selected two SELDI-TOF signatures for validation using this technology in the SELDI-TOF VANTDET cohort: those described by Sandhu et al.⁴², and Hamilton and Levin et al (unpublished, see 0). To allow inter-study comparisons to be made all SELDI-TOF profiles were acquired for the surfaces utilised in these studies.

The Sandhu et al., signature consisted of two set of m/z SELDI peaks, one set (98) derived from crude sera on the CM10 and the 2nd from 54 m/z peaks from fractionated sera. We acquired SELDI-TOF data on CM10 surface, at pH 4 and 6 in line with these signatures but found that only find 2/54 and 2/98 peaks that matched between the profiles of Sandhu and our spectra and based on this we were unable to validate the Sandhu et al., signatures for diagnosis of TB in our cohort.

The signature proposed by Levin et al., is based on the signature of 4-6 m/z peaks that corresponds to 4 unique proteins, and working closely with Levin et al., we could detect the same peaks in our SELDI-TOF profiles (TABLE 330). With these data we assessed the classification accuracy in the VANTDET cohort (n=86) using the DRS methodology they propose (see methods). This resulted in an AUC=0.66 (0.54-0.77) in the full cohort when comparing all TB (Dosanjh 1,2, n=45) to OD (Dosanjh 4, n=41). The AUC improved if only culture confirmed TB (Dosanjh 1, n=33) alone was included in the TB disease group, providing an AUC=0.74 (0.63-0.86), while no classification was observed for detection of highly probable TB (Dosanjh 2, n=12) from OD patients (AUC=0.58 (0.4-0.76); see FIGURE 5.

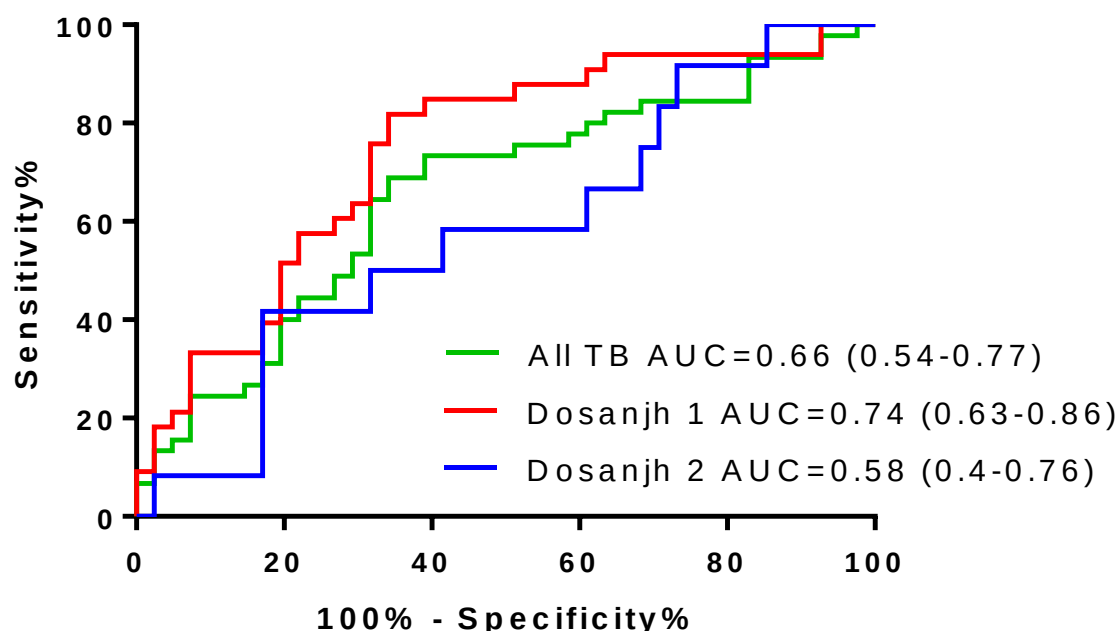


FIGURE 5. Performance of 4 protein signature (Hamilton and Levin) on SELDI-TOF, for the classification of TB in the VANTDET SELDI TOF sub-study cohort (n=86). The DRS for the 4

target proteins was calculated, and assessed to diagnostic performance using ROC-curve analysis. The performance for classifying all TB (Dosanjh 1 and 2 combined, n=45, green line), culture confirmed TB (Dosanjh 1, n=33, red line) and highly probable TB (Dosanjh 2, n=12, blue line) against all OD patients (n=41) is displayed.

Bridging study

By analysing the same 86 samples by LC-MS and SELDI-TOF our study was designed to bridge these two technologies. This was done to answer the question ‘can signatures identified in SELDI-TOF be reproduced by LC-MS?’. With a positive answer enabling the transfer of signature testing to LC-MS for larger cohorts, limited in SELDI-TOF due to the discontinuation of consumables. To do this we took the 3/4 protein identifications provided from the Hamilton and Levin SELDI-TOF signature (patent pending), extracted the quantitative data for these proteins in the LC-MS dataset, and analysed them using correlation (see 0). Correlation between methods was low only P1 achieved $P < 0.05$ (FIGURE 20A). We also compared the proteins for significance between the different categories in diagnosis of ATB and found no proteins significant in the SELDI-TOF data and one protein P2 significant when measured by LC-MS.

Performance of previously published signatures using LC-MS

Next, we asked if any of the key published and unpublished signatures (i.e the signatures identified by Hamilton and Levin et al (unpublished), Song et al,⁴⁴ Xu et al,⁴⁵ Achkar et al²⁰ and De Groote et al,¹⁸ termed ‘Levin’, ‘Song’, ‘Xu’, ‘Achkar’ and ‘De Groote’ respectively from hereon) could provide accurate disease classification for active TB using the LC-MS platform. To do this we utilised the quantitative LC-MS dataset generated on immunodepleted sera for n=166 individuals; extracting specific quantitative data for the protein signatures within each signature. The dataset consisted of spectra collected from immunodepleted sera digested with trypsin and labelled by 10-plex TMT labelling (see methods). Samples were fractionated by strong anion exchange chromatography into 6 fractions and each fraction analysed by LC-MS.

Using the software Maxquant we identified 8773 peptide identifications mapping to 720 proteins each with a quantitative value in at least one patient, and 380 proteins with a quantitative value for all patients. Imputation was required for 18% of the proteins used for

1 the discovery of novel signatures (42/228). Overall the imputation rate for other diseases
2 samples (n = 84) was 0.3% of all of the reported protein abundances. Similarly, for TB
3 samples (n = 82) the proportion of imputed values was 0.1%. The proteins for which
4 imputation was required, the number of missing values in TB and OD samples and the values
5 used for imputation are available in supplementary files 1-6.

6 ***Detection of key proteins from published signatures***

7 All of the proteins from 4 of the key publications (Song, Xu, Achkar, and Levin) were
8 detected in our depleted LC-MS dataset, and could therefore be validated within our cohort of
9 166 patients. Unfortunately the De Groote signature was unable to be validated within this
10 dataset, since only 4 of the 6 proteins could be detected, and 2 of these had >50% missing
11 values. For those signatures we could validate, we first assessed the abundance of the proteins
12 in the key patient groups. The average of the relative abundance (log2 ratio) and statistical
13 analysis of the proteins is shown in *TABLE 35* in *Appendix 6* . for the key patient groups.
14 When comparing all TB cases to OD the majority of proteins (10/18) were found to be
15 significantly different between these groups, this number reduced to 5/18 if only highly
16 probable TB (Dosanjh 2) cases were considered (P08294 from the Xu signature, Levin 1 and
17 2 from the Levin signature, and P02654 and P49908 from the Achkar signature) (see *TABLE*
18 *35*).

Validation of diagnostic performance of serum protein signatures using LC-MS

To assess whether these protein signatures could be used to classify all TB cases from OD we used LDA (see *FIGURE 6*) to test the performance of the signatures to separates diagnostic groups, with the exception of the Song signature, where only one protein (P01009 or SERPINA1) is used⁴⁴ and the raw abundance data can be used for validation purposes (see *FIGURE 210*).

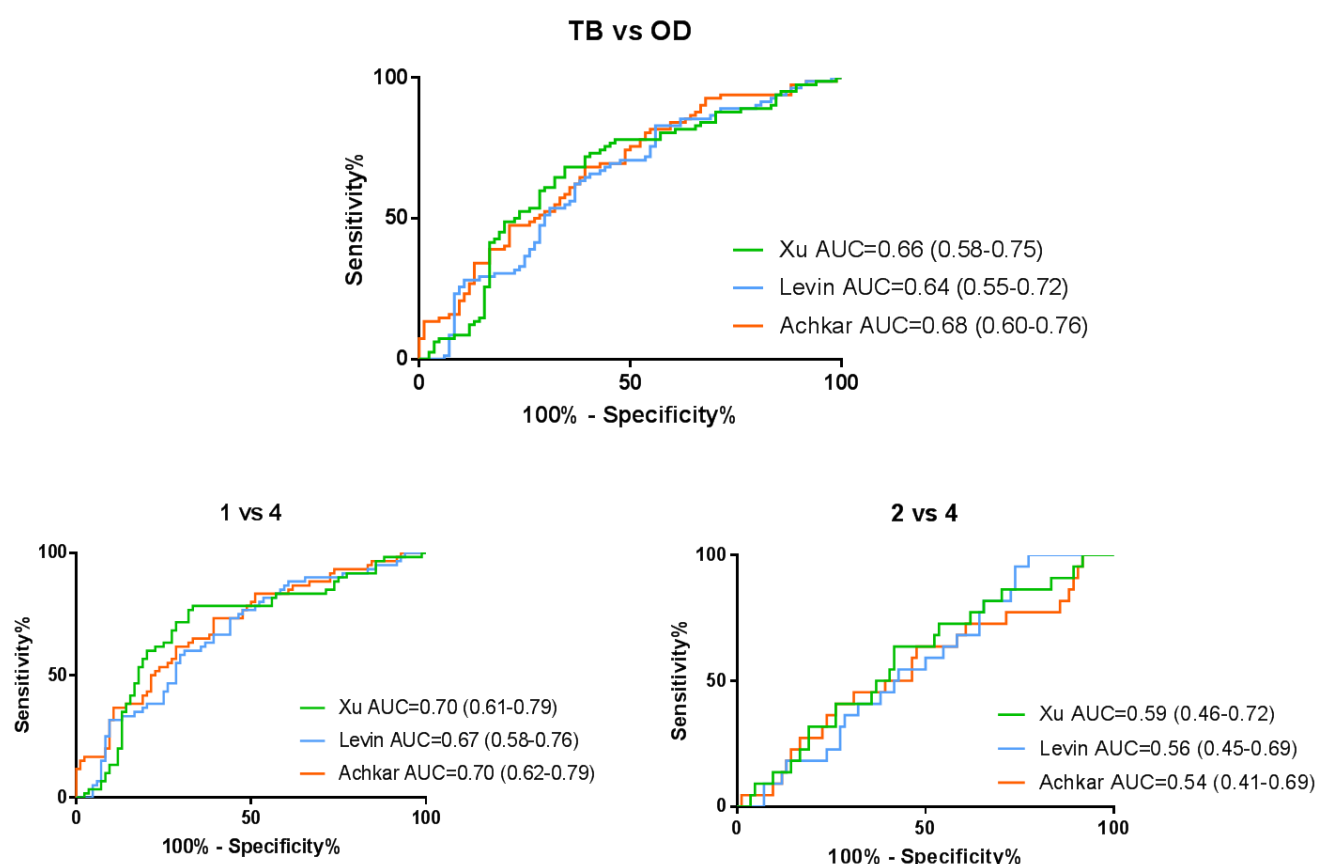


FIGURE 6. The diagnostic performance of the Xu, Levin and Achkar protein signatures of, using LDA, on the LC-MS depleted dataset (n=166). The diagnostic performance is presented for either: discrimination of TB vs OD, discrimination of Dosajh1 vs OD (Dosanjh 4) and discrimination of Dosanjh 2 vs OD (Dosanjh 4).

All four signatures performed similarly, providing an AUC of 0.62-0.67 for detection of all TB cases within the cohort. The AUC values, as well as sensitivity and specificity score can be found in *Appendix 6 (TABLE 37)* and *Appendix 7 (TABLE 38)*. We further assessed their performance for classification of either culture confirmed or highly probable ATB (i.e Dosanjh 1 or Dosanjh 2 respectively) against all OD For culture confirmed TB, the best

1 classification accuracy was achieved with the Xu and Achkar signatures, providing AUCs of
2 0.70 (TABLE 39). All of the signatures gave a poor performance for detecting highly probable
3 TB, with no significant discriminatory ability (AUC=0.53-59). The performance of all
4 signatures for diagnosing PTB was greater than that for EPTB where no discriminatory
5 ability was achieved by any of the published signatures (TABLE 37, TABLE 39 in Appendix 6
6 and Appendix 7). Thus, the sub-group in which all signatures performed optimally in
7 VANTDET was the culture confirmed TB group.

8 ***Removal of Sarcoidosis from other disease***

9 It is known from our own studies, and those by others, that the serum proteomic profile of
10 patients with Sarcoidosis are very similar to those with active TB. We therefore carried out
11 LDA analysis on the detection of all TB, as well as stratifying by Dosanjh category, when
12 patients with Sarcoidosis were removed from the OD cohort. The results from these analyses
13 are shown in Appendix 8 (TABLE 40, TABLE 41). For the detection of all TB, the
14 performance of all of the published signatures improved slightly. However, when the
15 performance in key sub-groups was assessed, the performances of the signatures changed
16 very little, or even reduced in diagnostic accuracy.

Discussion

The focus of this sub-study of VANTDET was to test the diagnostic accuracy of previously identified protein signatures that occur in the serum of patients with active TB. To do this we reviewed and tested previously published proteomic signatures using the two major technological platforms, SELDI-TOF and LC-MS.

Our results indicate that diagnosis of TB cases in UK hospitals is not possible using currently identified proteomic signatures in SELDI-TOF or LC-MS technology. We used ROC-curve analysis to indicate diagnostic accuracy for the binary classification of TB or OD. Applying this, none of the signatures tested met the go/no go threshold ($AUC > 0.85$) outlined in the management plan for this proteomics sub-study of VANTDET. The best performing signature was described by Achkar et al., and resulted in an $AUC = 0.68$ when measured by LC-MS and analysed by LDA. Using SELDI-TOF the Levin signature performed with an $AUC = 0.66$ when analysed by calculating the DRS. The classification accuracy was reduced compared to that reported by our collaborators in their cohort of TB and OD patients ($AUC = 0.85$). The other SELDI-TOF signature developed by Sandhu was not measurable due to poor inter-laboratory reproducibility. All signatures improved their classification accuracy if only culture confirmed ATB cases were considered for diagnosis. Herein the Levin SELDI-TOF signature gave an $AUC = 0.74$ which was the optimal diagnostic performance we identified using any proteomic technique; the Xu and Achkar gave a test accuracy performance of $AUC = 0.70$ using LC-MS as assessed using LDA.

We utilised the Dosanjh categorisation of active TB which as our references standard, in which TB diagnosis is based on both gold standard microbiological culture (1) and clinical diagnosis of culture negative TB (2). Approximately 28% of all TB cases recruited were culture negative, and this was driven by the relatively high prevalence of EPTB and other forms of paucibacillary TB cases amongst UK cohort. Here, the development of new diagnostics is vital as standard methods such as culture or PCR perform poorly. Our data shows that the host response signatures found in the serum proteome for TB also perform poorly in culture unconfirmed TB. The markers tested here perform better in culture confirmed (mainly pulmonary) TB potentially due to the robust acute phase response in this type of disease that tracks with increased symptoms and more extensive immunopathology or necrotic disease. Many of the previous signatures for TB were developed in cohorts with high TB prevalence, where severity of disease at presentation is likely greater, and/or a pre-

1 selection of culture confirmed and/or pulmonary TB groups. The study reported by Achkar et
 2 al²⁰ was the only signature developed from patients in different hospitals, including both
 3 culture negative TB and OD with TB like symptoms, very similar to the IDEA cohort utilised
 4 in VANTDET. IDEA also recruited prospectively resulting in a cohort of diverse OD
 5 diagnoses (100+) all classified as one group: Dosanjh 4. The high diversity amongst this
 6 group, and similarity of symptoms reported by patients from all groups reflects the real
 7 clinical problem, but it hampers any biomarker that is driven by a generalised mechanism in
 8 the host response to disease. For example, diseases consisting of both neutrophil driven
 9 necrotic disease, and/or macrophage recruitment are likely to lead to the same set of proteins
 10 (e.g. S100A8/A9, MMP9) altering in the serum as found in active TB. In developing their
 11 signatures both Xu et al⁴⁵ and Song et al⁴⁴ limit selection of control groups to specific
 12 diseases, and in Song et al, controls often didn't have similar symptoms. This artificial
 13 selection leads to the development of tests that work well between diseases that are
 14 specifically different in the host response orchestrated, but also leads to the selection of
 15 markers that only reflect the difference between healthy and sick people. In utilising the IDEa
 16 cohort, VANTDET improved on this by creating a design that tested the utility of such
 17 markers in complex environment where symptomatic people present on a case by case basis.

18 Our study was limited by technological problems in the platforms utilised, and the cross-
 19 platform compatibility available for protein detection and quantitation. The discontinuation of
 20 SELDI-TOF limited the numbers of samples that could be analysed, and the poor inter-
 21 laboratory reproducibility limited us to validating only the 4 protein signature developed in-
 22 house (developed by Hamilton and Levin et al (unpublished)). The transfer of these
 23 biomarkers to new methods LC-MS and ELSIA is troublesome. Firstly, the identification of
 24 the proteins was not made until late 2017 and remains unpublished. Secondly, markers
 25 detected by SELDI-TOF do not always correlate accurately to antigenic regions raised *in vivo*
 26 immune responses making the development of ELISA difficult. LC-MS detects tryptic
 27 peptides for proteins and therefore can also not full recapitulate what is observed in SELDI-
 28 TOF which may be large fragments of proteins. Our goal was to analyse 200 samples from
 29 the IDEA cohort, but this was limited to 90 by SELDI-TOF and 171 by LC-MS and was due
 30 to low levels of consumables left in the UK and limited access to LC-MS equipment
 31 internally. Furthermore, the larger dataset of crude sera resulted in poor protein coverage and
 32 an inability to detect the published signatures. To combat the limited power of our initial
 33 investigations we analysed only HIV negative samples. HIV status is known to affect

1 proteomic profile in the serum of people with TB but as our cohort consisted of limited
2 numbers of HIV positive (15 %) co-infected patients we decided to test signatures in HIV
3 negative patients initially, and expand into HIV positive individuals if we achieved an
4 AUC>0.85. This AUC was never achieved using our approach by LC-MS or SELDI-TOF.

5 LC-MS and SELDI-TOF failed to achieve the minimum requirements set in VANTDET for a
6 clinically useful signature. Problems that exist technologically for example proteomic depth,
7 sample throughput and cross platform development of an easy to use point of care test were
8 apparent in our study. Recently, a new technology termed SOMA-SCAN offers an attractive
9 method able to overcome these limitations. De Groote et al¹⁸ used this technology to discover
10 and test a novel 8 protein signature for TB in the serum of people with pulmonary TB
11 compared to other related non-TB diseases. This study was multi-centre and prospective, but
12 was limited to culture confirmed pulmonary TB, where it gave an AUC of 0.87 (0.81-0.91).
13 Testing of this signature in VANTDET using SOMA-SCAN was not possible due to the
14 limited accessibility to this technology in the public domain at the time of project conception.
15 Furthermore, the proteins comprising this signature were not all detected in either of our LC-
16 MS datasets.

17 To address our secondary objective of using these proteomic datasets to derive novel
18 signatures which may offer improved accuracy for diagnosing either all TB cases, or specific
19 subgroups of TB, we will be applying variable selection methods to identify the proteins
20 providing the best classification accuracy (in both crude and depleted sera samples). We
21 intend to train a prediction model on 60% of the cohort, and test the accuracy of the model on
22 the remaining 40% (test set). This work was not within the scope of this report, but will be
23 the logical next step in utilising this novel proteomic dataset to derive a clinically useful
24 proteomic signature with potential to be used in the diagnostic work-up of TB.

25 Our study provided two important conclusions. The previously reported proteomic signatures
26 for detecting TB were not found to be clinically useful in this cohort, likely due to the
27 differences in the complexity of TB/OD patients in the respective cohorts. Therefore, our
28 data from the IDEA cohort, which does reflect real life clinical practice, provides a spring
29 board for future studies, and an opportunity to derive new signatures that are robust and can
30 pave the way for future validation in similar cohorts.

Chapter 5. Cellular immune signatures

Introduction

The IGRA tests measure the presence of circulating memory T lymphocytes producing IFN γ in response to Mtb RD-1 antigens. In addition to being the gold standard test for LTBI, they are able to detect between 55-87% of all active TB cases depending on the population and test used.¹¹ However, since IGRAs cannot discriminate between these two states of infection, they lack the specificity to be used as a rule-in test for active TB. To improve on the specificity of such an immune based test, various groups have explored the Mtb-specific T cell response with greater complexity than simply measuring production of IFN γ , and have found that by using flow cytometry to measure phenotypic and functional markers of Mtb-specific T cells, cellular immune signatures have the potential to be used as immune based tests to allow for discrimination between active and latent infection.

The first group to identify a signature which differed between disease states sufficiently enough to allow for discrimination between active TB and LTBI was Harari *et al*, who in 2011 showed that the functionality of CD4⁺ T cells could discriminate between active and latent TB.²³ Specifically, the proportion of Mtb-specific T cells with a polyfunctional profile (IFN γ ⁺IL-2⁺TNF- α ⁺) was greater in those with latent infection, while the proportion secreting TNF α -only (IFN γ ⁻IL-2⁻TNF- α ⁺) was greater in active TB and could discriminate between the two groups with high accuracy in a training and test cohort O ²³.

Work from our laboratory exploited the potential diagnostic value of TNF α -only-secreting Mtb-specific T cells whilst simultaneously improving diagnostic accuracy by combining measurement of T cell memory/differentiation phenotype (based on cell-surface markers) in parallel. Indeed, Pollock *et al* demonstrated that measurement of the proportion of TNF α -only secreting PPD-specific CD3⁺CD4⁺ cells that were CD45RA⁻CCR7⁻CD127⁻ (i.e. highly-differentiated late-effector T cells) was found to be highly discriminatory between active TB and LTBI (O).²⁴ The signature had 100% sensitivity and 93% specificity in a test cohort of 34 patients (n=13 active TB, n=21 LTBI), of whom 50% were HIV-positive.²⁴ Since then, additional cellular immune signatures have been identified which also show promise in their ability to discriminate between active TB and LTBI, including one which combined the proportion of TNF α -only population (among all cytokine responders as in²³), with the presence of a detectable CD8⁺IFN γ ⁺ T cell response.⁵⁰

1
3 However, in recent years the research focus has returned to the CD4⁺ IFN γ -producing Mtb-
4 specific population, and several groups have reported using the memory marker CD27 to
5 differentiate TB from LTBI patients. Different approaches to incorporating this marker into a
6 cellular immune signature have been taken: i.e. by measuring the levels of CD27 as a ratio in
7 relation to the parent population,²⁶ or measuring the presence/absence of expression in
8 combination with the memory marker CD45RA.²⁷ The discrimination between groups was
9 high in these studies, with a specificity of above 90% for active TB for the CD27⁺CD45RA⁻
10 signature.²⁷ More recently, the incorporation of activation markers such as HLA-DR in
11 cellular immune signatures has been shown in several studies to provide improved
12 discrimination between these two disease states,²⁵ a finding which has been validated by
13 external groups, including those with/without co-infection with HIV.^{51, 52} For reference
14 throughout this chapter, the signatures are referred to using either the numerical value, or
15 short description, as follows: 1) %TNF α -only;²³ 2) %T_{EFF};²⁴ 3) SCORE;⁵⁰ 4) CD27 MFI
16 Ratio;²⁶ 5) %CD27^{+/+}CD45RA⁻;²⁷ 6) %HLA-DR.²⁵

17 Flow cytometry is already established in many NHS diagnostic laboratories throughout the
18 UK and is suitable for routine diagnostic use.⁵³ These signatures therefore warrant
19 prospective validation in a large-scale independent cohort in routine practice. As yet, none of
20 the studies have tested the performance of such signatures in key patient groups which reflect
21 those to present in a clinical setting, such as those with OD and LTBI, and TB patients with
22 hard to diagnose forms of the disease (i.e. culture unconfirmed and/or EPTB). Indeed, none
23 of the previous studies have been conducted where patients were recruited as part of routine
24 clinical practice.

26 **Aim**

27 By combining the markers required to measure all of the signatures which have shown
28 promise in to one panel of antibodies, we sought to validate the performance of these
29 signatures on a cohort of individuals with suspected TB recruited as part of routine clinical
30 practice (i.e using the IDEA cohort).

Objectives

1. To validate the performance of the six (1-6 described above) previously reported cellular immune signatures in diagnosing active TB
2. To perform sub-study analyses where specific clinical groups are included/excluded in the classification analysis (i.e culture confirmed/highly probably TB, PTB/EPTB etc).

Methods

Cohort selection

Since the cellular immune signatures to be validated are dependent on a detectable T cell response to Mtb antigens, and have thus been proposed as a test to allow for discrimination between active and latent TB, only those with a positive IGRA can be assessed. We therefore identified those individuals eligible for the final IDEA analysis, who had sufficient PBMCs available in the biobank and had a positive IGRA test (with a positive QFT-GIT and/or T-SPOT being sufficient to be classified as IGRA positive). To balance the two groups, we carried out a nested control approach and randomly selected patients from the ATB group to generate our final cohort, with the exception that all eligible HIV-infected samples were included in the final cohort. Using this approach, all of the individuals selected from the Dosanjh 4 category had a positive IGRA result, and can therefore be classified as having both ‘other disease’ and LTBI; we therefore refer to this group in the chapter as ‘OD/LTBI’.

Experimental design

Blinding of patient status

The patients were selected for this study by the data manager (AB) and the statistician (YT), using the criteria detailed above. The laboratory researchers (research associates and technicians: AH, MTW, TM) involved in the sampling handling, experiments and analysis, only had access to the patient IDs and locations. The demographic and clinical data related to each patient was not revealed to the researchers until the final analysis was complete and locked (i.e. the cellular immunology data could not be changed).

Batching

Each experiment took place over 2 days. Between 5-9 samples were defrosted in each experiment, and between 1-3 vials of PBMCs were required depending on recovery. The

1 details of viability, cell counting, plate layouts, antibody quantities and timings etc were
2 entered on to a standards experimental log sheet. Records of antibody volumes/concentrations
3 used, antigen batches and other details were also taken and stored for each experiment.

4 ***Cell culture and antigen stimulations***

5 Thawed PBMCs were initially resuspended in RPMI-1640 (company) supplemented with
6 10% heat-inactivated foetal bovine serum (FBS) (referred to as R-10 from hereon). The cells
7 were then washed, resuspended in RMPI-1640 and counted using Trypan Blue and an
8 automated cell counter (Cell Countess; Invitrogen). PBMCs were then resuspended in R10 to
9 give a final concentration of 15×10^6 cells/ml. The cells were distributed into 3-6 wells of a
10 48-well plate (in a volume of 100-250 μ l) and rested for precisely 6 hours (37°C, 5%, CO₂).
11 After resting, the cells were stimulated with PMA (phorbol 12- myristate 13-
12 acetate)/ionomycin positive control (PMA, final concentration of 5ng/ml and 500ng/ml for
13 Ionomycin), tuberculin purified protein derivative (PPD) RT50 (AJ vaccines/ Statens Serum
14 Institut) (16.7 μ g/mL final concentration) and a combination of four RD-1-related peptides
15 shown previously to have a powerful discriminatory power; include the 6kDa early secretory
16 antigenic target (ESAT-6), culture filtrate antigen (CFP-10), Rv3615c and Rv3879c) (ECRR
17 (final concentration of 5 μ g/ml per peptide)). Unstimulated cells were used as a negative
18 control. After stimulation the cells were incubated for a further 2 hours and 2 μ l Monensin /
19 Brefeldin A mix (both at 1X conetration according to Manufacturer's instructions;
20 Biolegend) is added and the plate is incubated for a further 16 hours (total stimulation
21 time=18 hours).

23 ***Multi-colour staining***

24 Following stimulation, cells were incubated on ice for 10 minutes followed by scraping and
25 washing with phosphate-buffered saline (PBS). Cells were then stained with a dead cell
26 marker (LIVE/DEAD Fixable Near InfraRed (NiR) (Life Cell Technologies)) at 1 x in PBS,
27 for 10 minutes on ice protected from light. After washing in FACS buffer (PBS containing
28 2mM EDTA, 0.5% BSA), all cells were placed in Block buffer (10% human serum in FACS
29 buffer) containing a mixture of pre-titrated fluorochrome-conjugated antibodies for detection
30 of surface markers (see *TABLE 43*). After washing with FACS buffer (PBS, 2mM EDTA,
31 0.5% BSA), the cells were fixed and permeabilized using Cytofix/CytopermTM buffer (BD
32 Biosciences fixation/permeabilization kit) for 20 minutes on ice protected from light. The
33 cells were washed with BD Perm/WashTM Buffer (BD Biosciences) from the same kit and

1 then stained with pre-titrated fluorochrome-conjugated antibodies diluted in Perm/Wash for
2 the detection of intracellular cytokines (table 1). The cells were washed twice with FACS
3 buffer before acquiring samples on a BD LSRFortessaTM cell analyser, containing 5 lasers
4 (UV 355nm, Violet 405 nm, Blue 488 nm, Yellow Green 561nm, Red 633 nm). Antibody
5 concentrations were optimised with during titration experiments using the *MFI staining index*
6 calculation approach.

8 ***Inter-experimental controls***

9 Various controls were included in each experiment, including the use 8 peak rainbow beads
10 (Biolegend) which allowed for fine tuning the voltage settings to allow for comparable
11 detection of fluorescence peaks across experiments. Single-stained controls for each
12 antibody/viability dye cells or mouse/rat beads (selected to allow to detection of maximum
13 possible brightness for each marker) were used to allow for calculation on the compensation
14 matrix. Finally, PBMCs from a PPD-responsive anonymous leucocyte cone donor was used
15 as a fully stained control in each experiment, as well as for fluorescence minus one (FMO)
16 controls for each target antibody. All the controls ensured the standardization and calibration
17 of the Fortessa instrument across experiments.

18 ***Acquisition on flow cytometer***

19 The samples and controls were acquired straight after the last experimental step using BD
20 Fortessa, coupled with a computer and FACS Diva v8.01 software. A maximum of 2×10^6
21 events (set at live singlet lymphocytes) were collected for each condition of each sample.
22 FCS files were exported later that day to be further analysed using FlowJo version 10.4.2
23 (Treestar/ BD).

24 ***Use of PPD-responsive control***

25 Leukocyte cones obtained from an anonymous donor were purchased from NHS Blood and
26 Transplant. Upon receipt, cells were washed in RPMI, counted and stored in cryovials
27 containing approx. 20 million live cells per vial. Cells from these samples were screened for
28 responsiveness to PPD using flow cytometry. One PPD cytokine responder was used as an
29 internal responsive control in each experiment.

Analysis and gating

The data was analysed using FlowJo version 10. Before analysis was applied to each experiment, the compensation matrix (derived from the FACS Diva software and compensation controls) was checked and adjusted accordingly so that over/under compensation did not affect gating. For images showing the gating strategies used for the full analysis of all signatures, see *Appendix 9 (FIGURE 22, FIGURE 23, FIGURE 24)*.

Selection of viable T cells

Gates were set on lymphocytes followed by single cells using forward and side scatter properties. Live lymphocytes were selected using Live-Dead NiR vs Forward Scatter (FSC-A). Samples were checked against time in case an abnormality occurred whilst acquiring. CD3⁺ T-cells were selected within the live lymphocytes population gated against the CD4⁺ T-cells; this gate allowed for detection of highly activated CD3⁺CD4⁺ T-cells on which surface expression of surface CD3⁺ can be downregulated (ref). To visualise these gates see *FIGURE 22*.

Selection of cytokine specific T cells

Within the CD3 population, CD4⁺ and CD8⁺ T-cell populations were further selected. The IFN γ ⁺ gate (after removal of artefacts as described in *Appendix 9*) was selected on either CD4⁺ vs IFN γ or CD8 vs IFN γ depending on the parent T cell population of interest. IL-2 and TNF α were plotted against each other and gated on this plot for either CD4⁺ or CD8⁺.

Phenotypic gates (i.e for CD45RA/CCR7, CCR7/CD127, CD127⁺/⁻ CD27/CD45RA, and HLA-DR⁺) were first gated on the CD4⁺ population, and were then transferred to cytokine specific populations as required for each published signature (see *FIGURE 23 and FIGURE 24*).

Determining the inclusion criteria for antigen responsiveness

For the purpose of this study, we defined a set of inclusion criteria to ensure that only those samples with a definite response to Mtb antigens, and with sufficient cells to allow for accurate calculation of phenotype, were included. Thus, to identify samples with a sufficient antigen-specific cytokine response (whether CD4⁺IFN γ ⁺ or CD4⁺TNF α -only), only those with at least twice the levels of cytokine above background were included (i.e. % of cells > 2

x unstimulated). In addition, if phenotypic analysis of cytokine-specific cells was required for signature calculation, only those with 10 or more analysable events were included. These criteria are similar, although not identical, to those used in the publications where the signatures were first proposed.

Statistical analysis

All statistics were performed using Prism software (GraphPad Software). Differences in signature measurements between groups was determined using either a Mann-Whitney or a Kruskal Wallis test (depending on the number of groups). Diagnostic accuracy for discriminating between groups was calculated using ROC-curve analysis.

Results

Cohort selection

The nested case control approach was used to select the final cohort for validation of cellular immune signatures, resulting in a cohort of 184 participants samples being defrosted for the stimulation and flow cytometry staining (*FIGURE 7*), with equal numbers of TB and OD/LTBI participants (n=92).

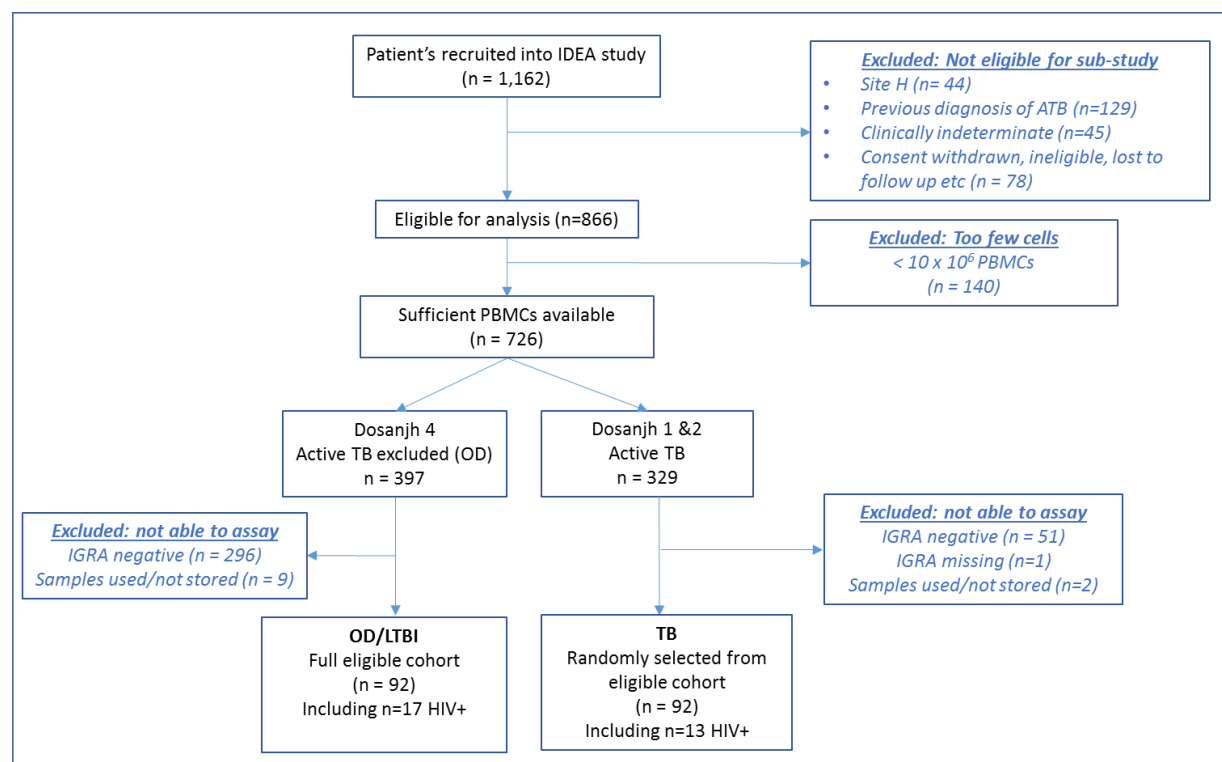


FIGURE 7. Study flow chart of patients included in the validation of cellular immune signatures sub-study. A nested case control approach was used to selected the patients from the full cohort of IGRA+ participants who were eligible for analysis in the IDEA study, resulting in a total of n=92 per group (with the TB patients being randomly selected and all the HIV+ IGRA + patients being selected). N.B. This cohort was larger than originally anticipated (i.e. n=164), since additional IGRA+ patients were identified since the October 2016 protocol was written.

Cohort characteristics

The main demographic characteristics of the final analysed cohort are present in TABLE 10 below. As with the other VANTDET sub-study and main IDEA study cohorts, there was a significant difference in age between the groups, with the OD group being older than the TB groups. Unlike in the IDEA study and transcriptomic sub-study cohort, we do not observe a significant difference in ethnicity in this sub-study cohort. This is likely due to the smaller cohort size. There were no significant differences in the key clinical characteristics (TABLE 11).

TABLE 10. Demographic characteristics of final cohort for validation of cellular immune signatures sub-study (n=184). The groups were compared for each characteristic, and the p value is reported in the right column.

Characteristic		Dosanjh 1: Culture Confirmed TB	Dosanjh 2: Highly Probable TB	Dosanjh 4: OD/LTBI	P value
Total	N	60	32	92	
Age	years, median (range)	33 (16-72)	36.5 (21-76)	40.5 (17-80)	0.0027
Gender	Female, n (%)	18 (30%)	13 (40.6%)	34 (37%)	0.5365
Ethnic Origin					0.2136
	Asian, n (%)	3 (5%)	3 (9.4)	7 (7.6%)	
	Black, n (%)	18 (30%)	6 (18.8)	29 (31.5%)	
	Hispanic, n (%)	1 (1.7%)	1 (3.1)	1 (1.1%)	
	Indian Sub-Continent, n (%)	32 (53.3%)	20 (62.5)	34 (37%)	
	Middle Eastern, n (%)	0 (0%)	0 (0%)	5 (5.4%)	
	Mixed, n (%)	0 (0%)	1 (3.1)	1 (1.1%)	
	White, n (%)	6 (10%)	1 (3.1)	14 (15.2%)	
	Unknown, n (%)	0 (0%)	0 (0%)	1 (1.1%)	

TABLE 11. Clinical characteristics of the validation of cellular immune signature sub-study cohort (n=184). The groups were compared for each characteristic, using either a Mann-Whitney U test or chi-squared test and the p value is reported.

Characteristics		Dosanjh 1: Culture Confirmed TB	Dosanjh 2: Highly Probable TB	Dosanjh 4: OD/LTBI	P value
Total	N	60	32	92	
Height (m)	Median (range) Missing, n (%)	1.68 (1.5-1.89) 24 (40%)	1.71 (1.5-1.8) 13 (40.6%)	1.7 (1.5-1.96) 26 (28.3%)	0.823
Weight (kg)	Median (range) Missing, n %	60.2 (41.3-97) 0 (0%)	64 (44.7-116) 1 (3.1%)	70.5 (42-132) 4 (4.3%)	
BMI	Median (range) Missing, n %	21.7 (15.7-32) 24 (40%)	23.7 (16.1-42.2) 14 (43.8%)	24.3 (14.9-47.2) 27 (29.3%)	
BCG positive	N, %	46 (76.7%)	25 (78.1%)	64 (69.6%)	0.500
HIV positive	N, %	6 (10%)	6 (18.75%)	13 (14.1%)	0.495

The distribution of disease in the TB groups is presented in *TABLE 44*, while the list of main diagnoses in the OD group is presented in *TABLE 45* (both in *Appendix 10*). Individuals with PTB and EPTB were presented in both the culture confirmed (Dosanjh 1) and highly probable (Dosanjh 2) TB groups.

Recovery and detection of antigen specific responses

Of the 184 samples where PBMCs were defrosted, 181 (98%) were used for the full 2-day experiment (3 were not plated or used for the full experiment due to recovery of no/too few live cells (2 were TB; one was OD/LTBI)). Of the 181 samples which were assayed (i.e. stimulated, stained with full antibody panel and analysed), all were stimulated with PPD, all had a negative control (unstimulated), 179 (99%) were stimulated with PMA/ionomycin and 107 (59%) were stimulated with the RD-1 antigens (ECRR). The Mtb antigen condition was only included in each assay when there were sufficient cells to have four conditions ($>7 \times 10^6$). The PMA/ionomycin stimulation condition was excluded in 2 cases due to low numbers of cells. For details of the recovery rate of PBMC from storage, viability before plating and at acquisition, viability after stimulation and overall cytokine responses to all antigens, see *Appendix 11 (TABLE 46-49, FIGURE 25)*. In summary, all of those samples stimulated with PMA/ionomycin, while responses to PPD and Mtb antigens was less frequent (with 87% and 93% of those stimulated responding with a detectable IFN γ response, respectively). Those individuals who did not meet the criteria for antigen responsiveness were excluded from diagnostic accuracy analysis.

Diagnostic performance of signatures in full cohort

Here we present the diagnostic performance of six previously identified cellular immune signatures for the detecting of TB from OD/LTBI in the IGRA+ population of TB suspects.

Distribution of signature values in TB and OD/LTBI groups

The distribution of signature measurements in the TB and OD/LTBI groups after stimulation with PPD are displayed in *FIGURE 8A*, whilst the values after RD-1 antigen stimulation are displayed in (*Appendix 14*). There was significant and strong correlation in 4 of the 6 signatures scores between PPD and RD-1 stimulations in those individuals who had both stimulations (see *FIGURE 29*- only shown for %HLA-DR).

Performance characteristics

The performance of each of the signatures for discriminating all TB from those with OD/LTBI is presented in *FIGURE 8* and *TABLE 12* for the PPD stimulation, stratified by HIV infection status. The signature with the highest test accuracy was the %HLA-DR (of CD4+ IFN γ + cells) signature first proposed by Adekambi *et al*, which gave AUCs of 0.90 (0.84-0.95), 0.91 (0.76-1.1) and 0.89 (0.84-0.94) in the HIV-, HIV+ and combined populations respectively. As has been previously reported, it is necessary to evaluate performance of this signature in the context of HIV infection (i.e. with/without and not combining HIV+/-) since HIV infection increases the proportion of overall HLA-DR+ CD4+ T cells, and as such, the cut-off for this signature is higher in the HIV+ population.

The next best performing signatures in the HIV- and HIV-/+ combined sub-groups were, in order of decreasing accuracy, the %T_{EFF} (of CD4+ TNF α -only cells) proposed by Pollock *et al* (AUC= 0.86 in HIV-; 95% CI 0.79-0.92), %CD27-CD45RA- proposed by Petruccioli *et al* (AUC=0.82 in HIV-; 95% CI 0.75-0.89) and CD27 MFI Ratio proposed by Portevin *et al* (AUC=0.78; 95% CI 0.70-0.86) all of which showed highly significant discrimination between TB and OD/LTBI above the level of AUC= 0.5 ($p < 0.0001$). However, none of these signatures performed well in the HIV+ population, with AUCs ≤ 0.58 in all cases (*TABLE 12*). No statistical analysis to compare diagnostic accuracy between signatures was performed.

The two signatures which measured only proportions of functional Mtb-specific subsets, and not phenotype of Mtb-specific cells, the %TNF α -only (of CD4+ Cytokine+) signature proposed by Harari *et al* and the SCORE signature proposed by Rozot *et al*, did not provide significant discrimination between groups in the HIV- or HIV+/- groups, with AUCs ≤ 0.57

- 1 which were not significant ($p > 0.05$). Due to the small numbers of responders in the HIV-
- 2 cohort, this population was excluded from subsequent sub-cohort analyses.

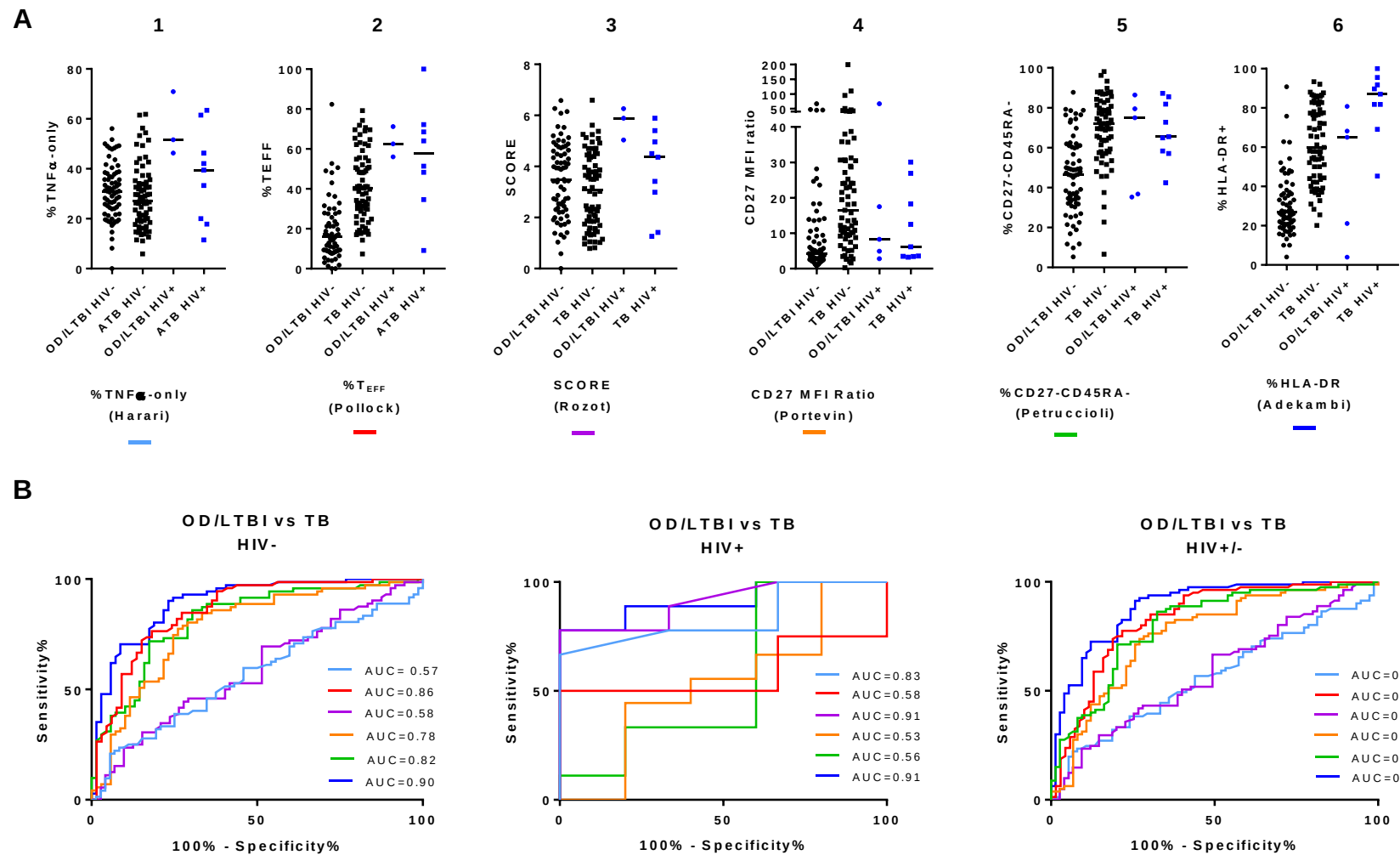


FIGURE 8. Performance of published signatures in full cohort, stratified by HIV infection status. (A) Dot plots score the signature values after PPD stimulation in TB suspects, stratified by diagnosis (TB or OD) and HIV infection status (+ or -). (B) ROC curves demonstrating diagnostic performance of

signatures in the detection of active TB cases within the combined HIV +/- population, HIV+ alone and HIV- alone. The signatures are as follows: 1) Harari²³ (%TNF α -only); 2) Pollock²⁴ (%T_{EFF}); 3) Rozot⁵⁰ (SCORE); 4) Portevin²⁶ (CD27 MFI Ratio); 5) Petruccioli²⁷ (%CD27-CD45RA-); 6) Adekambi²⁵ (%HLA-DR).

TABLE 12. Performance of published cellular immune signature in diagnosing all TB, stratified by HIV infection status. Two approaches to deriving the cut-off were taken: either the point at which sensitivity=specificity*, or setting specificity to >90% ^{\$}. For calculation of PPV and NPV, the following prevalence values were used: for HIV- and +/- population, the prevalence of TB in the T-SPOT+ population was 76.7% TB; For the HIV+ population, the prevalence of TB in the T-SPOT+ population was 43.6%.

Signature, ref	HIV	TB (n)	OD/ LTBI (n)	AUC (95 CI)	Cut-off	Sensitivity (95 CI)	Specificity (95 CI)	PPV	NPV
1 %TNF α -only ²³	-	72	72	0.57 (0.47-0.66)	<29.3* <18.7 ^{\$}	54.2 (42-66) 23.6 (14.4-35.1)	54.2 (42-66) 90.3 (81-96)	0.80 0.89	0.26 0.26
	+	9	3	0.83 (0.59-1.07)	<49* <44.2 ^{\$}	77.8 (40-97.2) 66.7 (29.9-92.5)	66.7 (9.4-99.2) 100 (29.4-100)	0.64 1.0	0.80 0.80
	-/+	81	75	0.56 (0.47-0.65)	<29.8* <18.7 ^{\$}	55.6 (44.1-66.6) 23.46 (14.8-34.2)	56 (44.1-67.5) 90.7 (81.7-96.2)	0.81 0.89	0.28 0.26
2 %T _{EFF} ²⁴	-	72	66	0.86 (0.79-0.92)	>27.6* >37.2 ^{\$}	76.4 (64.9-85.6) 56.9 (44.73-68.6)	77.3 (65.3-86.7) 90.9 (81.3-96.6)	0.92 0.95	0.50 0.39
	+	8	3	0.58 (0.25-0.92)	<59.3* <53.7 ^{\$}	50 (15.7-84.3) 50 (15.7-84.3)	66.7 (9.4-99.2) 100 (29.4-100)	0.54 1.0	0.63 0.72
	-/+	80	69	0.83 (0.76-0.90)	>28.1* >49.4 ^{\$}	77.5 (66.8-86.1) 36.3 (25.8-47.8)	76.8 (65.1-86.1) 91.3 (82.0-96.7)	0.92 0.93	0.51 0.30
3 SCORE ⁵⁰	-	72	72	0.58 (0.49-0.68)	<3.43* <1.7 ^{\$}	52.8 (40.7-64.7) 23.6 (14.4-35.1)	52.8 (40.7-64.7) 90.3 (81-96)	0.79 0.89	0.25 0.26
	+	9	3	0.91 (0.73-1.1)	<5.2* <5.0 ^{\$}	77.8 (40-97.2) 77.8 (40-97.2)	66.7 (9.4-99.2) 100 (29.2-100)	0.64 1.0	0.85 0.85
	-/+	81	75	0.58 (0.49-0.67)	<3.5* <1.7 ^{\$}	51.9 (40.5-63.1) 23.5 (14.8-34.2)	52 (40.2-63.7) 90.7 (81.7-96.2)	0.78 0.89	0.25 0.26
4 CD27 MFI Ratio ²⁶	-	71	69	0.78 (0.70-0.86)	>9.5* >25 ^{\$}	74.7 (62.9-84.2) 32.4 (21.8-44.6)	75.4 (63.5-85) 91.3 (82-96.7)	0.91 0.92	0.47 0.29
	+	9	5	0.53 (0.19-0.88)	<7.2* <4.3 ^{\$}	55.6 (21.2-86.3) 44.4 (13.7-78.9)	60 (14.7-94.7) 80 (28.4-99.5)	0.52 0.63	0.64 0.65
	-/+	80	74	0.76 (0.68-0.84)	>9.3* >25.0 ^{\$}	72.5 (61.4-81.9) 31.3 (21.4-42.6)	73 (61.4-82.7) 90.5 (81.5-96.1)	0.90 0.92	0.45 0.29

5 % CD27- CD45RA ⁻²⁷	-	71	69	0.82 (0.75-0.89)	>60.1 [*] >76.5 ^{\$}	73.2 (61.4-83.1) 39.4 (28-51.8)	72.5 (60.4-82.5) 91.3 (82-96.7)	0.90 0.94	0.45 0.31
	+	9	5	0.56 (0.19-0.92)	>69.1 [*] >86.8 ^{\$}	44.4 (13.7-78.8) 11.1 (0.28-48.3)	40 (5.3-85.3) 100 (47.8-100)	0.36 1.0	0.48 0.59
	-/+	80	74	0.80 (0.73-0.87)	>60.7 [*] >76.6 ^{\$}	72.5 (61.4-81.9) 38.8 (28.1-50.3)	73 (61.4-82.7) 90.5 (81.5-96.1)	0.90 0.93	0.45 0.31
6 %HLA-DR ²⁵	-	71	69	0.90 (0.84-0.95)	>41.55 [*] >49.6 ^{\$}	80.3 (69.1-88.8) 70.4 (58.4-80.7)	81.2 (70-89.6) 91.3 (82-96.7)	0.93 0.96	0.56 0.48
	+	9	5	0.91 (0.76-1.1)	>75 [*] >81.3 ^{\$}	77.8 (40-97.2) 77.8 (40-97.2)	80 (28.4-99.5) 100 (47.8-100)	0.75 1.0	0.82 0.85
	-/+	80	74	0.89 (0.84-0.94)	>42.2 [*] >54.3 ^{\$}	80.3 (69.9-88.3) 64.2 (52.8-74.6)	79.2 (68.5-87.6) 90.9 (82.2-96.3)	0.93 0.96	0.55 0.44

1

2

Diagnostic performance in key sub-groups of TB

Next, we sought to evaluate the performance of the cellular immune signatures in detecting important subsets of TB. Firstly, we assessed the diagnostic accuracy for identifying either the culture confirmed TB cases (Dosanjh 1) or culture-unconfirmed TB (Dosanjh 2). We were also able to evaluate the influence of the Dosanjh 4 sub-categories (B, C and D, which correspond to decreasing risk/likelihood of having LTBI) on the signature performance. The signature values in each of the Dosanjh categories are presented in 0.

As with the full cohort analysis, the signatures measuring functionality only (cytokine responses), i.e. the TNF α -only (1) signature and SCORE (2) signatures, showed no differences between Dosanjh categories in the HIV- cohort. However, the four signatures measuring phenotype (2, 4, 5, 6) showed significantly higher values in both Dosanjh 1 and 2 TB groups when compared to the OD/LTBI Dosanjh 4 groups. Interestingly, no differences were observed in the signature values between the Dosanjh 4 sub-categories (only 4B and 4C could be used for statistical comparisons since there were too few participants in the 4D group).

1

2 TABLE 13. The diagnostic accuracy of the discriminatory signatures was optimal for the culture confirmed TB group (with the %HLA-DR signature achieving highest
3 AUC of 0.92 (0.88-0.97)) although the performance for the culture unconfirmed group (Dosanjh 2) was still reasonable for this hard-to diagnose group of patients; see
4 TABLE 13. To evaluate whether the Dosanjh 4 subcategory influences the diagnostic accuracy of the signatures, we evaluated the performance of detecting all TB using
5 only the Dosanjh 4B or 4C patients in the TB-excluded (OD/LTBI) group. Interestingly, the performance of the cellular immune signatures was comparable, suggesting
6 that likelihood/risk of LTBI in the OD/LTBI group does not influence test accuracy of cellular immune signatures; see TABLE 51 in Appendix 12 . Performance of
7 cellular immune signatures in detecting TB, stratified by Dosanjh criteria. Cut-offs were selected using two approached: * Balanced sensitivity and specificity. \$ setting
8 specificity at > 90%. For calculating PPV and NPV, the prevalence of Dosanjh 1 was 57.1%, and the prevalence of Dosanjh 2 was 19.6%. The signatures are as follows:
9 1) Harari²³ (%TNF α -only); 2) Pollock²⁴ (%T_{EFF}); 3) Rozot⁵⁰ (SCORE); 4) Portevin²⁶ (CD27 MFI Ratio); 5) Petruccioli²⁷ (%CD27-CD45RA-; 6) Adekambi²⁵ (%HLA-
10 DR).

Comparison	Sig	TB (n)	OD/LTBI (n)	AUC (95% CI)	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV
Dosanjh 1 vs Dosanjh 4	1	47	72	0.58 (0.47-0.68)	<28.9* <18.7\$	55.3 (40.1-69.8) 23.4 (12.3-38.0)	55.6 (43.4-67.3) 90.3 (81.0-96.0)	0.62 0.76	0.48 0.47
	2	48	66	0.88 (0.82-0.94)	>29.3* >42.2\$	81.3 (67.4-91.1) 47.9 (33.3-62.8)	81.8 (70.4-90.2) 90.9 (81.3-96.6)	0.86 0.88	0.77 0.57
	3	47	72	0.58 (0.48-0.69)	<3.46* <1.70\$	51.1 (36.1-65.9) 23.4 (12.3-38.0)	51.4 (39.3-63.4) 90.3 (81.0-96.0)	0.58 0.76	0.44 0.47
	4	46	69	0.81 (0.73-0.89)	>10.4* >25.0\$	73.9 (58.9-85.7) 39.1 (25.1-54.6)	75.4 (63.5-85.0) 91.3 (82.0-96.7)	0.80 0.86	0.68 0.53
	5	46	69	0.84 (0.77-0.92)	>61.3* >76.5\$	78.3 (63.6-89.1) 47.8 (32.9-63.1)	78.3 (66.7-87.3) 91.3 (82.0-96.7)	0.83 0.88	0.73 0.57
	6	46	69	0.92 (0.88-0.97)	>44.2* >49.6\$	82.6 (68.6-92.2) 78.3 (63.6-89.1)	82.6 (71.6-90.7) 91.3 (82.0-96.7)	0.86 0.92	0.78 0.76
Dosanjh 2 vs Dosanjh 4	1	25	72	0.55 (0.41-0.69)	<30.3* <18.7\$	52.0 (31.3-72.2) 24.0 (09.4-45.1)	52.8 (40.7-64.7) 90.3 (81.0-96.0)	0.21 0.38	0.82 0.83
	2	24	66	0.81 (0.71-0.90)	>23.0* >37.2\$	70.8 (48.9-87.4) 45.8 (25.6-67.2)	72.7 (60.4-83.0) 90.9 (81.3-96.6)	0.39 0.55	0.91 0.87
	3	25	72	0.58 (0.45-0.72)	<3.3* <1.6\$	56.0 (34.9-75.6) 24.0 (09.4-45.1)	56.9 (44.7-68.6) 90.3 (81.0-96.0)	0.24 0.38	0.84 0.83
	4	25	69	0.73 (0.62-0.84)	>7.5* >25.1\$	68.0 (46.5-85.1) 20.0 (06.8-40.7)	68.1 (55.8-78.8) 91.3 (82.0-96.7)	0.34 0.36	0.90 0.82

Comparison	Sig	TB (n)	OD/LTBI (n)	AUC (95% CI)	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV
	5	25	69	0.77 (0.67-0.87)	>57.4* >76.5 ^{\$}	72.0 (50.6-87.9) 24.0 (9.4-45.1)	71.0 (58.8-81.3) 91.3 (82.0-86.7)	0.37 0.40	0.91 0.83
	6	25	69	0.84 (0.76-0.93)	>38.1* >49.6 ^{\$}	76.0 (54.9-90.6) 56.0 (34.9-75.6)	76.8 (65.1-86.1) 91.3 (82.0-96.7)	0.44 0.61	0.93 0.89

The performance of signatures in the smear negative IGRA positive cohort of TB suspects

Smear microscopy results are available on the same day that the patient presents to secondary care and IGRA results are usually available by the next day. The flow-cytometry signature aims to distinguish active from latent TB in those who are IGRA-positive where it unclear whether the Mtb infection is latent or active. In this sub-analysis, we therefore assessed the performance of the two best flow signatures (the %HLA-DR and %T_{EFF}) within the IGRA-positive Smear-negative population within this cohort; see *TABLE 14* below.

TABLE 14. The performance of the %HLA-DR and %T_{EFF} signatures in detecting sub-groups of TB patients in the smear negative patients, within the IGRA positive cohort used in the validation of flow cytometric signatures sub-study. *For this analysis, the cut-offs for the signatures were determined by finding the value where sensitivity = specificity.

	%HLA-DR ²⁵	%T _{EFF} ²⁴
<i>Performance in detecting Smear negative TB (Dosanjh 1 and 2)</i>		
<i>Prevalence of TB in all smear- T-SPOT+ TB suspects= 79% (From overall IDEA population)</i>		
n (TB) / n (OD)	49 / 45	50 / 43
AUC (95% CI)	0.91 (0.84-0.97)	0.90 (0.83-0.96)
Cut-off*	>41.2	>24.6
Sensitivity	81.6 (68.0-91.2)	80 (66.3-90.0)
Specificity	80 (65.4-90.4)	81.4 (66.6-91.6)
PPV /NPV	0.94 /0.54	0.94 / 0.52
<i>Performance in detecting Smear- culture + (Dosanjh 1) TB</i>		
<i>Prevalence of culture + TB in all smear negative T-SPOT+ TB suspects = 59%</i>		
n (TB) / n (OD)	29 / 45	31 / 43
AUC (95% CI)	0.91 (0.83-0.98)	0.92 (0.86-0.98)
Cut-off*	>43.1	>28.2
Sensitivity	82.8 (64.2-94.2)	83.9 (66.3-94.6)
Specificity	82.2 (68.0-92)	86.1 (72.1-94.7)
PPV /NPV	0.87 / 0.73	0.90 / 0.78
<i>Performance in detecting smear – culture negative (Dosanjh 2)</i>		
<i>Prevalence of culture unconfirmed TB in smear negative TB suspects = 20%</i>		
N (TB) /n (OD)	20 / 45	19 /43
AUC (95% CI)	0.90 (0.82-0.98)	0.86 (0.76-0.96)
Cut-off*	>39.1	>22.5
Sensitivity	80 (56.3-94.3)	79 (54.4-94.0)
Specificity	77.8 (62.9-88.8)	79.1 (64.0-90.0)
PPV /NPV	0.47 / 0.94	0.494

The pre-test probability of a TB-suspect having TB after presenting with both a negative smear and a positive IGRA test is 79% (N.B. 59% and 20% for culture positive and culture negative TB sub-groups respectively), which is high, but not high enough to be confident that commencing TB-treatment is the correct treatment course. The findings from this sub analysis show that by using either the %HLA-DR or %T_{EFF} signature tests, a high PPV for all TB (0.94) and culture confirmed TB alone (0.87) can be achieved.

The HIV- cohort was further stratified by disease type, where the TB group was split into either PTB or EPTB, and the OD/LTBI group was split into groups based on most frequent final diagnoses; see (*FIGURE 27, TABLE 52*). Again, only the signatures measuring phenotype showed differences between groups. Whilst no differences between other diagnoses in the OD/LTBI were observed, both PTB and EPTB showed elevated signature levels compared to the OD/LTBI groups.

Diagnostic accuracy was highest for detection of PTB cases from the overall OD/LTBI group (max AUC of 0.93 (0.87-0.98)) compared to EPTB where performance was reduced (maximum AUC of 0.88 (0.82-0.94)), as achieved using the %HLA-DR signature which achieved the highest test accuracy. The diagnostic performance characteristic, including AUCs, sensitivity and specificity, for detecting either PTB or EPTB are presented in *0*.

Discussion

Here we present the first validation of previously identified cellular immune signatures for the detection of active TB cases in a cohort of TB suspects recruited as part of routine clinical practice. The immunological nature of these assays, with the signatures only being measurable when a detectable Mtb-specific CD4+ response can be characterised, confines the eligible patient population to those who respond to Mtb antigens (i.e. the IGRA positive individuals), and as such the challenge for these assays is to discriminate between active TB and OD/LTBI. We found that of the six signatures validated in this cohort, only those which measured incorporated phenotypic markers in the flow cytometry signature (i.e. %T_{EFF}, CD27 expression of %HLA-DR²⁴⁻²⁶) showed discriminatory ability, while the signatures measuring functionality did not differentiate between TB and OD/LTBI.

1 The best performing signature was found to be the %HLA-DR signature, which has recently
2 been validated in several cohorts and shown to have high test accuracy for differentiating TB
3 from healthy LTBI.^{25, 51, 52} This is the first study to demonstrate it's diagnostic performance in
4 a cohort where the LTBI group also had OD, and where all patients were recruited with
5 suspected TB. The performance we present here (AUC=0.90) is not as high as previously
6 found (where AUCs were 0.92-1.0). However, when only culture confirmed TB amongst
7 HIV- patients was considered, the diagnostic accuracy of this signatures was comparable to
8 that reported by Riou et al (also AUC= 0.92).⁵² Therefore, the overall lower performance we
9 present here for detection of all TB using previously described cellular immune signatures,
10 likely reflects the inclusion of paucibacillary and hard to diagnose TB patients, particularly
11 those with culture unconfirmed TB.

12 The %T_{EFF} signature also performed with reasonably high accuracy for detection of all TB,
13 providing an AUC of 0.86 for detecting all TB in the HIV- population. Unlike in the original
14 paper, where performance was high in a mixed cohort of HIV+/- patients, in this study, the
15 signature could not discriminate between TB and OD/LTBI in the HIV+ population.

16 Both the %HLA-DR and %T_{EFF} signatures could be useful in the setting of IGRA positive
17 TB suspects, since in this scenario they are able to provide a high PPV for both all TB groups
18 (0.95 and 0.95 respectively) and for culture confirmed TB (0.92 and 0.88 respectively).
19 Therefore, upon initial screening of an IGRA positive suspect, a positive result in these tests
20 could provide clinicians with sufficient evidence (i.e. ~90% probability) to start treatment
21 immediately, without any need to wait for culture confirmation. This is the testing algorithm
22 in which these assays could be useful. In future studies, we wish to determine these
23 signatures in the smear negative group of TB patients. Importantly, we also assessed the
24 performance of these two signatures in the smear negative sub-group within the cohort. These
25 signatures were found to provide very high PPVs for all TB, and culture confirmed TB,
26 within the T-SPOT+ smear negative population of TB suspects. These patients usually
27 require several more investigations to inform treatment decision, or else treatment will be
28 started without confidence in the correct diagnosis having been made. For culture-confirmed
29 TB, where the prevalence within the smear- T-SPOT + population was 59%, the PPVs of
30 0.87 and 0.88, raised the pre-test probability of TB by 28-29% (i.e. from 59% pre-test
31 probability to 87-88% post-test probability). The increase in the pre-test probability for all TB
32 cases in this specific population was 13%, from 79% to a PPV of 94%, demonstrating that a
33 positive test result with either of these assays would provide a rapid rule-in test facilitating an

1 accelerated decision to treat. A positive result in this setting would thus provide strong
2 support for a diagnosis of TB and a decision to treat.

3 We found in this study that the signatures measuring polyfunctional profiles (1, 3)^{23, 50} don't
4 appear to be able to discriminate between TB and OD/LTBI groups. Several research groups
5 have measured similar profiles, and it appears that these signatures are not reproducible
6 across different studies. Day et al found that there were difference between LTBI vs smear
7 negative and smear positive TB, but not with TNF α -only CD4+ responses, but
8 IFN γ ⁺TNF α ⁺IL-2⁻ and IL-2 responses.⁵⁴

9 Incorporating measurement of CD27 showed reasonable test accuracy in our cohort, but only
10 in the HIV- population, and to a lesser extent than the %HLA-DR and %T_{EFF} signatures
11 (although statistical analyses to compare tests was not performed).

12 There are some limitations. Some of the PBMC stimulations did not results in detectable
13 cytokine responses, meaning that they were excluded from the diagnostic performance
14 analysis. Therefore the performance may not be fully reflected in the analysis we present, and
15 in practice these test may result in a number of failed or indeterminate results.

16 This is the largest validation of cellular immune signatures in a cohort of TB and LTBI
17 patients, and the first to test their performance in a cohort of TB suspects. We have
18 demonstrated that cellular immune signatures can provide high test accuracy and may be
19 useful for a rapid blood based test for TB in the IGRA positive population.

20

Chapter 6. Candidate rule-out tests

Introduction

IGRAs are the recommended immunological assays to detect Mtb infection. There are two commercially available IGRAs, both of which measure IFN- γ production in response to RD-1 antigens: the T-SPOT.TB which uses the T-SPOT platform, and the QFT-GIT which uses ELISA. IGRAs are unable to discriminate between active disease and LTBI, and as such, are unsuitable as tests to rule-in active TB (due to poor specificity). Their sensitivity for detecting active TB cases is reasonably high, leading to the possibility that these tests could be used as a *rule-out* test for active TB. Ideally, such a test would provide over 95% sensitivity for all TB cases for it to be clinically useful. The IDEA study demonstrated that the T-SPOT-TB tests provided optimal accuracy (as compared to QFT-GIT), but that neither test in their current formats, provides sufficient sensitivity (between 67.3-82.3% for detecting all TB¹¹) for this important clinical utility.^{10, 11} In this sub-study we explored two candidate approaches to improve on the sensitivity of immunological tests for active TB.

Additional Mtb antigens

A strategy for improving diagnostic sensitivity of immune based tests is to include additional Mtb antigens. Use of Mtb-specific antigens for stimulation ensures that, although different secreted immune mediators are being measured than in IGRA (ie chemokines instead of IFN γ), diagnostic specificity is not compromised. Various studies have taken this approach in an attempt to stratify different types of TB infection and/or improve sensitivity^{21, 55}. Indeed, the new updated version of the Quantiferon test (Quantiferon TB Gold-Plus) contains an additional antigen tube containing peptides which are proposed to stimulate Mtb CD8⁺ T cells and drive improved test sensitivity for TB. However, very few studies have been performed to validate this new test, and one recent study found it provided comparable sensitivity and specificity for active TB patients than the QFT-GIT.⁵⁶

The recently completed IDEA study identified that four Mtb RD-1-related antigens were able to provide good sensitivity for active TB when used in a next-generation T-SPOT TB assay: ESAT-6, CFP-10, Rv3615c (Esx-1 substrate protein C (EspC)) and Rv3798c.¹¹ Use of all four of these antigens in a stimulation assay together, will therefore likely maximise detection of Mtb infection. Indeed, in the next generation T-SPOT assay, combinations of these four

antigens could provide an improved performance, with a sensitivity of 89.9% for all TB and 94.4% for culture confirmed TB.¹¹

Detection of additional or alternative analytes

A further powerful and pragmatic approach to enhance test sensitivity is to include measurement of alternative or additional immune mediators. We and others have proposed to improve on the sensitivity provided by detection of IFN γ by measuring immune mediators whose secretion is promoted by IFN γ , thereby providing an amplified signal.⁵⁷⁻⁵⁹ Two such chemokines are CXCL9 (also known as MIG) and CXCL10 (also known as IP-10), both of which are produced by monocytes, neutrophils and other cells in response to IFN γ production by antigen specific T cells, and have been found to provide increased sensitivity for TB cases in some studies.^{59, 60} Others have proposed to detect alternative cytokines such as IL-2, TNF α or Th2 cytokines,^{59, 61} although the evidence for these cytokines providing improved sensitivity for TB (over that of IFN γ) is limited.

Alternative assay platforms

We and others have previously developed a highly sensitive whole blood molecular test with high potential to reliably rule out the diagnosis of TB based on measurement of IFN γ -dependent chemokine release in response to Mtb-specific antigens.²² The assay involves overnight stimulation of blood with ESAT-6 and CFP-10, and subsequent measurement of CXCL9 and CXCL10 by qRT-PCR. To further improve sensitivity, we incorporated a third highly immunodominant and Mtb-specific antigen, EspC (Rv3615c), an immunogenic RD-dependent antigen²¹. In preliminary studies using a case control approach, the assay was found to be 100% sensitive for active TB (n=39) when combining pairs of antigens, regardless of the pairing (i.e. any 2 of ESAT-6, CFP-10 and EspC (also known as Rv3615c)) (Connell *et al. Thorax. 2012* (BTS Abstract)). Specificity was also high (94%), with only 1 of 16 IGRA-negative BCG-vaccinated healthy controls scoring positive in the assay. These compelling data suggest potential for a stimulated chemokine qRT-PCR test with sufficiently high diagnostic sensitivity to be used to rule out TB infection (and therefore active TB disease) at first presentation. The promising results with this assay to date now merit validation in a large-scale prospective cohort study of patients undergoing routine diagnostic work up for suspected active TB.

As an alternative to the ELISA platform for detecting proteins as used in the current QFT-GIT, we further propose that the detection of target analytes from plasma of stimulated blood

could be improved on this platform using Meso-Scale Discovery (MSD) technology, which provides a broader dynamic range and greater sensitivity for low abundant proteins than ELISA (<https://www.mesoscale.com/>).

Rationale for candidate rule-out assays

Assay platforms

Given the promising performance of using qRT-PCR detection of Mtb antigen specific CXCL9 and CXCL10 production in stimulated PBMC samples in our unpublished case-control cohort, coupled with data from Kasproicz et al²² where performance appeared to be greater than T-SPOT, we decided to use this assay approach for the detection of our target genes. An additional IFN γ -related gene was selected for inclusion in the candidate molecular rule out test. BATF2, which was significantly over-expressed in whole blood collected from TB patients (Hoang et al, unpublished data), encodes a transcription factor and is expressed in macrophages stimulated by IFN γ .⁶²

Additionally, within the IDEA biobank we were also able to use the surplus supernatants from the QFT-GIT tests performed to evaluate whether the use of selected analytes. Therefore, as an additional sub-study to further test the proof of concept that detection of alternative analytes can improve upon test accuracy when compared to IFN γ , we chose to measure target chemokines in the QFT-GIT supernatants using MesoScale Discovery (MSD).

Aim

The aim of this sub-study was to evaluate the diagnostic accuracy of two candidate immune based tests to measure previously identified chemokine and cytokines for their use in ruling out active TB cases

Primary objective:

- To determine whether the detection of molecules downstream of IFN γ , namely CXCL9, CXCL10 and BATF2, by qRT-PCR from RD-1 stimulated PBMCs can provide a greater sensitivity for active TB than currently available tests (i.e a molecular rule-out test).

Secondary objectives:

- To determine whether detection of CXCL10 or IFN γ in the GFT-GIT supernatants by MSD can provide improved diagnostic accuracy for detecting TB cases than the ELISA platform.
- To determine the performance of the above tests for discriminating between active TB vs OD in the IGRA + population of TB suspects.

Methods

Cohort selection

Since the existing data on this candidate molecular rule-out test is limited when compared to the other tests, we decided (at the SMG meetings and ISAB meetings) that for this technology we would take a two staged approach to validation. Firstly, we planned to assess the performance of the test in an artificial cohort, selected from within the IDEA cohort, to act as a proof of principle that the candidate tests can provide greater sensitivity for active TB than the current immunological tests: the T-SPOT and QFT-GIT. Thus we designed a GO/NO GO cohort for the first stage of validation. If high sensitivity for ATB was achieved in the cohort (i.e. greater than T-SPOT), then the test would be rolled out on the full cohort of eligible patients for whom a PBMC sample was available (N.B that the cellular immune test was prioritised within VANTDET, but for most participants more than one PBMC vial was available so inclusion in both sub-studies was achievable in many cases).

For overall eligibility, the same inclusion criteria was used for assessment of the T-SPOT and QFN tests within the IDEA study (i.e. diagnosis of either TB or OD, no history/evidence of previous TB), with the exception that individuals from site H were included in this cohort, and only those with an available PBMC sample were included. HIV+ individuals were included.

GO/NO GO cohort

For the GO/NO GO cohort, 25 individuals with available PBMC samples were picked randomly from each of the four different groups; active TB T-SPOT+, active TB T-SPOT-, OD T-SPOT+, and OD T-SPOT-. Previous studies exploring novel immune tests for active TB lack the active TB IGRA negative group, which we have enriched for in this cohort.

Should any of the analytes we detect in these candidate assays provide >95% sensitivity for

active TB, with an acceptable level of specificity (>65%), then the GO/NO GO criteria from the first study would have been met, and the assays would have been performed on the full IDEA cohort, where PBMCs were available. For detection of CXCL10 and IFN γ by MSD using QFT-GIT supernatants, we used a similar cohort of TB and OD patients.

Full cohort

In the event that 95% sensitivity and >65% specificity for all active TB cases is achieved (i.e GO criteria met), the assay will be performed on all available eligible individuals with available PBMCs.

Experimental set-up for molecular rule-out test

All 100 samples were placed into blocks of 5 or 6 samples for each experiment, selected randomly.

Stimulation

After thawing the frozen PBMCs and counting the with Countess II FL Automated Cell Counter (Thermo Fisher) the samples were spun down and re-suspended in R10, and plated at 1×10^6 PBMCs per well in a 48 well plate (CELLSTARTM); at least 3 wells were required for each sample. Cells were immediately stimulated with one of the following antigen formulations (200 μ l total volume): pooled 15-mer overlapping peptides from the Mtb antigens ESAT-6, CFP-10, Rv3615c, and Rv3798c (ECRR) in R-10 (final peptide concentration of 10 μ g/ml); IFN γ (10 μ g/ml) in R-10, to act as the positive control; R-10 containing DMSO (final concentration 1.48% to match that of ECRR) to act as a negative control. Stimulated PBMCs were incubated at 37⁰C and 5% CO₂ for 16 hours.

RNA extraction and quality control

After 16 hours stimulation the cells were lysed with RLT buffer with 1% β -Mercapethanol. Total RNA from the stimulated cells were extracted using Qiagen RNeasy Micro kit as per manufacturer's instructions. Total RNA was eluted in 14 μ l of RNase free water and stored in -80⁰C until further used. Quantification of concentration and quality of RNA was measured using NanoDrop 8000 (Thermo Fisher) and Agilent 2100 Bioanalyzer (Agilent Technologies) respectively.

Two step RT-qPCR

Extracted mRNA was reverse transcribed to cDNA using First Strand cDNA synthesis (Thermo Fisher) PCR was performed using TC-512 Thermal Cycler (TECHNE) with RNase

Inhibitor and Random Hexamers. The amplification and detection was performed using 7500 Fast Real-Time PCR System (Applied Biosciences) using the default settings suitable for Quantitation-Comparative Ct (threshold cycle) to perform the analysis $2^{-\Delta\Delta C_t}$. Probes and primers (Life Technologies), Assay ID; for the internal reference HPRT1 (Hs02800695_m1) and target genes CXCL10 (Hs01124251_g1), CXCL9 (Hs00171065_m1), IFN- γ (Hs00989291_m1) and BATF2 (Hs00912737_m1) were used. Relative fold change was calculated as the ratio between the mean Ct values of the target gene and reference gene (GAPDH) in each stimulated sample in relation to the unstimulated sample.

Detection of candidate biomarkers from Quantiferon supernatants

During the IDEA study, the excess supernatants from the QFT-GIT test were harvested and stored at -80°C . As an additional project to test the performance of promising candidate immune mediators, IFN γ , IL-2, CXCL10, MCP-1, MCP-3, MIP-1B, and TNF α . Using the MSD U-Plex platform to detect all these chemokine and cytokine as per manufacturer's manual with some variation with longer stimulation time, plates were read using MESO QuickPlex SQ 120 instrument. Concentration was calculated by normalising the stimulated supernatant from the nil sample.

Statistical Analysis

All statistics were performed using Prism software (GraphPad Software). Differences between groups (in normalised fold change gene expression or analyte concentration) was determined using a Kruskal Wallis test with Dunn's post hoc test for multiple comparisons. Diagnostic accuracy for discriminating between groups was calculated using ROC-curve analysis.

Results

Cohort characteristics

The details of the cohort eligible for analysis, and the stratification between the full eligible cohort and GO/NO GO cohort is given in *FIGURE 9* below.

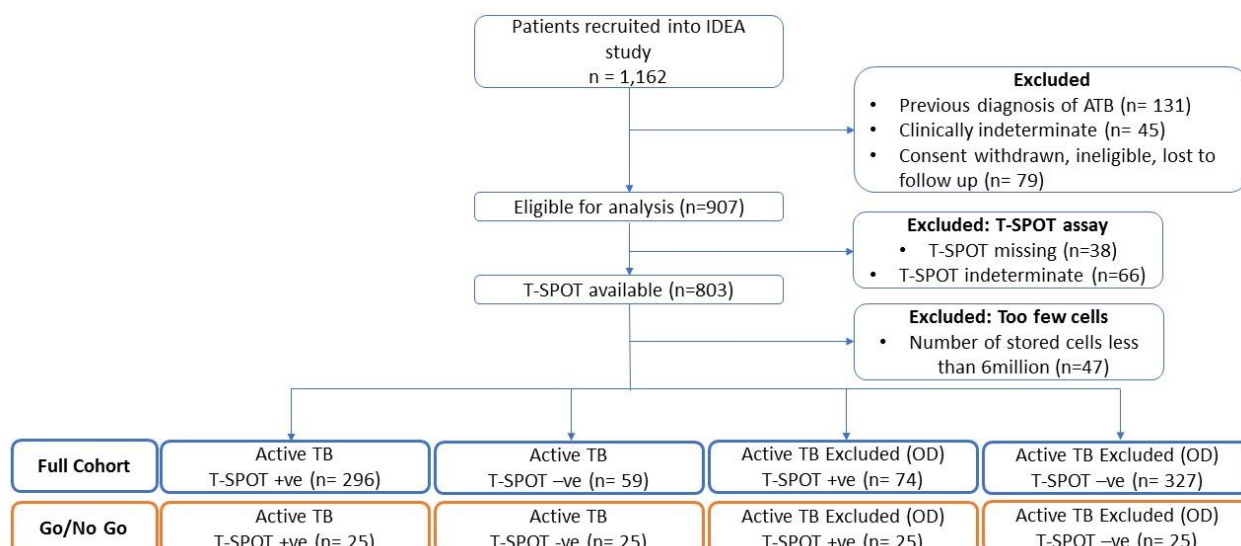


FIGURE 9. Study flow diagram for the validation of the candidate rule-out tests.

From the 100 individuals selected to be part of the GO/NO GO cohort, 3 were not found in the biobank, and an additional 3 samples were found subsequently to be ineligible (having a previous history of TB, leaving a total of 94 in the full cohort of assay individuals.

The demographic and clinical characteristics of the final cohort are shown in TABLE 15 below. Further details on the severity and type of TB disease, or OD, within the full cohort is given in Appendix 15 (TABLE 55, TABLE 1).

TABLE 15. Demographic and clinical characteristics for GO/NO GO cohort for assessment of candidate molecular rule-out test for active TB. Characteristics were compared across groups using either a Kruskal-Wallis test or a chi-squared test; p values are reported.

	Active TB		Other Disease		P value
	T-Spot + (n=24)	T-Spot – (n=22)	T-Spot + (n=22)	T-Spot – (n=26)	
Median Age (Range)	33 (19-57)	48.5 (27-81)	43.95 (19-74)	49.19(21-84)	0.0011
Female Gender, n (%)	12 (50)	7 (31.8)	8 (36)	10 (38)	
Ethnic Origin, n (%)					N/A
Asian	2 (8.3)	2 (9.1)	1 (4.5)	2 (7.7)	
Black	4 (16.6)	0 (0)	3 (13.6)	7 (26.9)	
Indian Subcontinent	17 (70.8)	13 (59.1)	10 (45.5)	9 (34.6)	
Mixed	0 (0)	2 (9.1)	0 (0)	0 (0)	
White	1 (4.2)	5 (22.7)	8 (36.4)	8 (30.1)	
HIV positive, n (%)	1 (4.2)	2 (9.1)	6 (27.3)	1 (3.8)	0.032
Tuberculin Skin Test					<0.0001
Positive, n (%)	5 (20.8)	9 (40.9)	0 (0)	1 (3.8)	
Negative, n (%)	0 (0)	0 (0)	1 (4.5)	8 (30.1)	
Not Tested, n (%)	19 (79.2)	13 (59.1)	21 (95.5)	19 (73.1)	

BCG					
Yes, n (%)	18 (75)	17 (77.3)	19 (86.4)	22 (84.6)	
No, n (%)	6 (25)	5 (22.7)	3 (13.6)	4 (15.4)	
Height (M), Median (Range)	1.69 (1.52-1.85)	1.69 (1.61-1.8)	1.65 (1.55-1.88)	1.68 (1.52-1.86)	0.3981
Height Missing, n (%)	4	10 (45.5)	2	7	
Weight (kg), Median (Range)	61 (42-123.5)	61.5 (47.4-96)	64.5 (45-82.6)	72.3 (48.3-140)	0.2815
Weight Missing, n (%)	1	0	0	4	
BMI, Median (Range)	23.1 (17.6-36.9)	22.6 (18.9-31.1)	22.4	24.4 (15.9-43.2)	0.6108
BMI Missing, n (%)	4	10	2	8	

Validation of molecular rule out tests

To measure gene expression of all analytes (CXCL10, CXCL9, BATF2 and IFN γ) using qPCR, a minimum of 50ng of total RNA concentration was needed. Of the 94 samples that were assayed, 13 samples yielded low amounts of RNA of around ~40ng, therefore not all analytes could be measured; in these cases, BATF2 expression was not measured.

Gene expression of target analytes in response to the positive control

Since all of the target genes (CXCL9, CXCL10, BATF2) are known to be up-regulated in response to IFN γ , this was used as the positive control in the assay. Previously, PHA was used in our laboratory for these assays (as is used in T-SPOT and QFT-GIT). However, preliminary optimisation experiments within VANTDET revealed that the levels of the target genes were not reliably increased using PHA. This is likely due to the reduced viability of PBMCs following PHA stimulation, which we confirmed by flow cytometry (data not shown). By plotting the normalised relative expression of target genes in response to ECRR or IFN γ , we were able to select a cut-off for an acceptable response in the positive control condition. The cut-offs selected were as follows: >100 for CXCL10, >100 for CXCL9 and >10 for BATF2, as shown in *FIGURE 31* in *Appendix 16*.

Gene expression in response to Mtb antigens

The results of the PCR detection of target genes in response to the antigen conditions (normalised to unstimulated, are shown in *TABLE 16* below.

TABLE 16. qPCR detection of CXCL10, CXCL9, and BATF2 in response to Mtb antigens or positive control in the GO/NO GO cohort of patients selected for validation of candidate molecular rule-out tests.

Active TB		Other Disease	
T-SPOT +	T-SPOT -	T-SPOT +	T-SPOT+
N= 24	N =22	N =22	N=26

CXCL10	Total for analysis, n (%)	19 (75)	19 (87)	12 (55)	21 (81)
	Total failed, n (%)	5 (25)	3 (13)	10 (45)	5 (19)
	<i>PCR indeterminate</i>	0	1	0	1
	<i>HPRT1 fail</i>	3	2	6	5
	<i>Low positive control</i>	2	1	4	0
CXCL9	Fold change ECRR/ US mean (range)	23604.50 (116.68-140596.57)	9362.95 (8.11-46681.24)	115353 (1.45-56513.26)	1408.53 (1.5-18053.68)
	IFN- γ / US n, mean (range)	50483.2 (224.9-325447.0)	126019.7 (829.6-432419.1)	104350.1(156.8-384828.4)	55180.4 (361.6-381780.3)
	Total for analysis, n (%)	18 (76)	18 (83)	13 (59)	21 (81)
	Total failed, n (%)	6 (24)	4 (17)	9(40.9)	5 (19)
	<i>PCR indeterminate</i>	0	1	1	1
BATF2	<i>HPRT1 fail</i>	3	2	6	5
	<i>Low positive control</i>	3	2	2	0
	Fold change ECRR/ US mean (range)	33280.75 (350.84-120140.97)	15049.99 (0.99-74914.48)	1030.38 (0.19-5146.35)	788.49 (0.23-7952.44)
	IFN- γ / US mean (range)	97456.2 (301.1-605633.7)	245973.5 (1025.6-804235.3)	171387.5 (220.2-792079.3)	123803.4 (1033.6-795414.8)
	Total for analysis, n (%)	19 (80)	14 (66.7)	10 (45.5)	18 (69.2)
BATF2	Total failed, n (%)	5 (20)	8 (33.3)	12(54.5)	8 (30.8)
	<i>Low RNA concentration</i>	1	4	4	4
	<i>PCR indeterminate</i>	0	1	1	0
	<i>HPRT1 fail</i>	3	2	6	4
	<i>Low positive control</i>	1	1	1	0
BATF2	Fold change ECRR/ US mean (range)	33280.75 (350.84-120140.97)	15049.99 (0.99-74914.48)	42919.01 (0.19-545582.52)	788.49 (0.23-7952.44)
	IFN- γ / US mean (range)	97456.2 (301.1-605633.7)	245973.5 (1025.6-804235.3)	171387.5 (220.2-792079.3)	123803.4 (1033.6-795414.8)

1

2 Of note, a high proportion (15-27%) of the assays resulted in ‘fail’ result, due to a number of
3 factors. The most prevalent reason for assay failure was unreliable detection of the house
4 keeping gene HPRT1. Individuals with HIV co-infection resulted in many of the assay
5 failures, particularly for detection of Mtb specific CXCL10. We also, for comparison,
6 measured the levels of IFN γ in response to the antigens using the same approach. The
7 findings are shown in *TABLE 57 in Appendix 16* . The gene expression data for individuals
8 within each group who passed test inclusion criteria for passing the assay are presented in
9 *FIGURE 30 in Appendix 16* .

The T-SPOT+ TB group displayed the highest Mtb-specific expression of all target genes when compared to other groups, with the exception of CXCL10, where levels were also very high in the T-SPOT+ OD (i.e OD/LTBI) group.

Diagnostic performance for detection of active TB

The diagnostic performance of the target genes for detecting all TB cases vs all OD are presented in *FIGURE 10*. While obtaining >95% sensitivity for all active TB is achievable when cut-off for fold/change over unstimulated is set at around 10 for CXCL9 and CXCL10, this was coupled with very low specificity (~5%). For BATF2 and IFN γ , the cut-off corresponding to >95% sensitivity was below 1, suggesting that it is not achievable to use these analytes to detect this proportion of all TB cases.

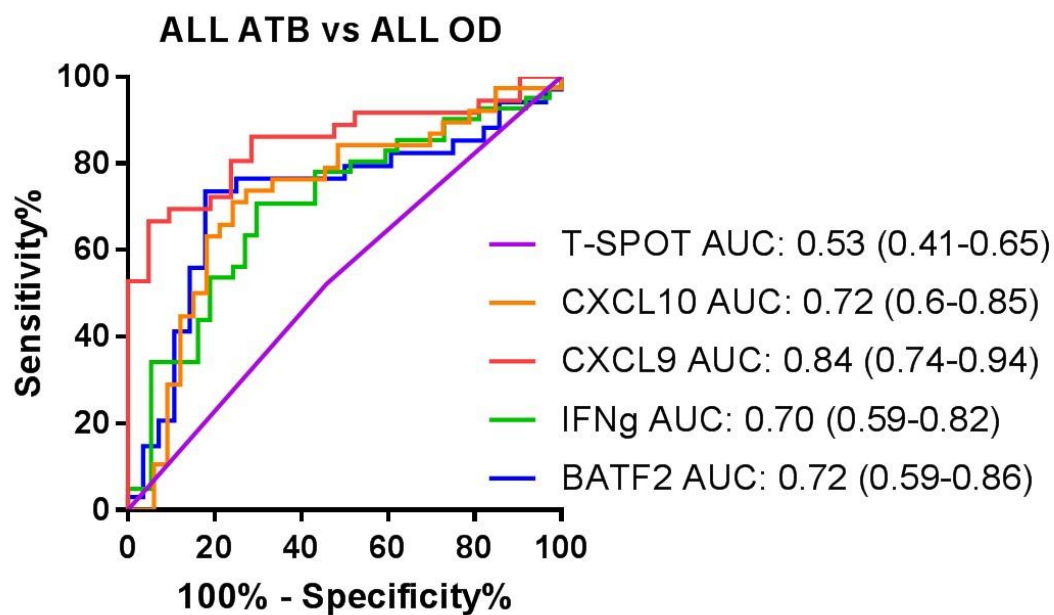


FIGURE 10. Diagnostic performance of candidate qRT-PCR detection of CXCL10, CXCL9, BATF2, or IFN γ in the candidate molecular rule-out as assessed in GO/NO GO cohort. The classification of all TB from all OD (either T-SPOT +/- for both) using these read-outs is shown (n=2—48 per group depending on the test). The performance of the T-SPOT assay in this cohort is also shown.

When compared to the T-SPOT test within the same cohort, which showed no diagnostic accuracy in this setting due to the artificial way the participants were selected (AUC of 0.53), all of the analytes showed improved diagnostic accuracy. This suggests that this method of qRT-PCR measurement of cytokines may provide improved sensitivity than the T-SPOT

platform, since even IFN γ detection by qRT-PCR performed better than the T-SPOT in detecting active TB patients. However, these ROC-curve analyses do not include those where the qRT-PCR assay failed.

The best performing gene for detection of all active TB was CXCL9, which provided an AUC of 0.84 (0.74-0.94) for detecting all TB cases in this cohort. However, none of the new assays were able to provide 95% sensitivity with an acceptable corresponding specificity (i.e. >65%), see *Appendix 16* , *TABLE 58*. In fact, in all cases the specificity for either TB group, or indeed LTBI, was unexpectedly low, with a wide range of gene expression values in all groups.

Since the proportion of T-SPOT+ OD patients within this artificial cohort was inflated, the likely true specificity of these tests is likely to be greater in a cohort that better reflects the levels of T-SPOT positivity in the TB and OD groups. To test this, we assessed the diagnostic performance in detecting all TB cases (T-SPOT+/-) when compared to T-SPOT – OD cases (see *FIGURE 32 in Appendix 16*) of the tests are reduced. In this setting, the performance of target genes was more comparable to the T-SPOT test (AUC = 0.76), with CXCL9 again providing the best test accuracy (AUC = 0.86).

We also compared the performance of target genes when only the T-SPOT+ TB and T-SPOT- OD were considered. In this setting, none of the test provided 100 accuracy and therefor performed sub-optimally to the T-SPOT (*FIGURE 33 in Appendix 16*). The full performance data for these calculations, including sensitivity and specificity values are shown in *Appendix 16* .

Discriminating active TB from LTBI/OD

To address our secondary objective of assessing the performance of our target genes in discriminating between active TB and LTBI in our cohort, we assessed their performance using ROC-analysis comparing TB –T-SPOT + and OD T-SPOT + only; see *FIGURE 34*. In this setting, the T-SPOT provides no discrimination (since results used are either ‘positive’ or ‘negative’ rather than numerical). All of the target genes, including IFN γ , provided some discrimination between action TB and OD/LTBI in this sub-cohort of patients, with CXCL9 providing the highest AUC of 0.97 (0.91-1.0).

Performance of CXCL10 and IFN γ in QFT-GIT supernatants using MSD

Of the n=95 IDEA participants selected for QFT GIT supernatant experiments using the MSD platform, n=75 samples overlapped between the candidate molecular rule-out test cohort, and two of the analytes measured (CXCL10 and IFN γ) could therefore be compared across the two approaches, data shown in *0*. The diagnostic performance of both IFN γ and CXCL10 for detecting TB cases was similar using both platforms. The detection of difference between patient groups based on either T-SPOT assay (for qRT-PCR) or QFT-GIT for MSD assays, was good, but greater discrimination was seen between QFT-GIT +/- groups in the MSD assay platform, likely reflecting the fact that the analytes were measured from the same stimulated samples (whereas a new stimulation was performed for the molecular rule out assay).

Discussion

In this sub-study we have evaluated two candidate approaches for improving the sensitivity of immune based tests for active TB: the molecular rule-out test for measuring chemokines from stimulated PBMC using qRT-PCR, and the measurement of IFN γ and CXCL10 in QFT-GIT supernatants using MSD. Unlike other studies which used a case-control approach and where patient were often pre-selected for immune assay responsiveness, in this study we have designed a cohort where we deliberately enriched for patients where current immune based tests fail: i.e those in the TB group who are IGRA negative, and those in the OD group who are IGRA positive. This artificial cohort selection approach was used to provide a GO/NO GO point for the validation of these candidate tests.

We found that CXCL9, CXCL10, and BATF-2 gene expression was upregulated in response to RD-1 antigens (ECRR) in TB patients, including in T-SPOT- TB patients. We confirmed previous findings that the levels of CXCL9 and CXCL10 produced by stimulated blood cells are detected at amplified levels compared to those of IFN γ .^{22, 58} However, these genes were also upregulated in a large proportion of OD patients in our cohort, which compromised diagnostic performance. This phenomenon was also observed by others, where IGRA-controls also displayed production of CXCL10 in response to Mtb antigens,^{22, 58} suggesting potential non-specific production of this chemokine in some tests. While reasonably high diagnostic accuracy was found (AUC= 0.84 for CXCL9) for all TB in our GO/NO GO cohort, this did not correspond to a test which could provide >95% sensitivity and 65% specificity at a given threshold. Therefore, this approach is unlikely to be clinically useful as a rule out test for all TB. Of all the analytes measured in the molecular rule-out assay approach, CXCL9 provided best diagnostic accuracy for detection of all TB cases from all OD, and for discrimination of TB from OD/LTBI in T-SPOT + patients. To our knowledge, this is the first study to test the performance of these chemokines in discriminating between Tb and LTBI in a cohort where the LTBI patients had OD and were suspected of active TB at the point of recruitment. It appears that CXCL9 may be a useful marker for differentiating between these groups. However, it should be noted that statistical analyses to compare diagnostic accuracy (i.e. AUC, sensitivity or specificity values) between analytes was not performed.

We reported a high proportion of assay failures for the candidate molecular rule out test assay, which were likely due to poor recovery of some PBMC samples from cryopreservation

1 and into culture. Using either fresh blood or fresh isolated PBMC for this candidate assay
2 would have been optimal, but we did not have this material available within VANTDET.
3 Kasprowicz *et al* have already demonstrated that detection of Mtb-specific CXCL9 and
4 CXCL10 using a similar qRT-PCR approach and as little as 50 µl fresh blood per condition is
5 achievable.²² Their findings now warrant validation in a cohort of TB suspects, similar to the
6 IDEA cohort, but with using fresh blood samples. We anticipate that fresh samples would
7 result in fewer assays failures and a better performance for detecting all TB cases.

8 Using supernatants stored from the QFT-GIT tests conducted as part of the IDEA study, we
9 were able to assess the use of the sensitive MSD platform for detection of target analytes and
10 compare the performance of CXCL10 and IFN γ between the two candidate platforms. This
11 way, we could compare the performance (although somewhat indirectly) in detection of these
12 analytes by using either qRT-PCR or ELISA. We found comparable performance for
13 detection of the analytes, and very similar diagnostic accuracy for detecting TB. However,
14 neither approach was able to provide the required diagnostic accuracy to be used as a rule-out
15 test for all TB, and furthermore, not all IGRA+ individuals were detected using the new
16 assays. Nevertheless, we were able to demonstrate that MSD is able to detect analytes such
17 as CXCL10 in plasma including a wide range of concentrations. While this is not required for
18 a simple positive/negative immune test, a test for discriminating between disease states,
19 where differences in quantity of analytes may provide discriminating, may benefit from using
20 an MSD platform for quantitation. We plan to investigate this in our future experiments using
21 MSD.

22 Both the candidate rule-out test experimental approaches we have evaluated are complex,
23 with multiple steps required (which can introduce significant variability), and requiring
24 technical expertise as well as expensive equipment to perform. This is not in keeping with the
25 WHO target product profile for a triage test for TB, which stipulates that such a triage test
26 should take between 5-30 minutes to get a result, cost ~\$1-2 and be achievable using a single
27 device which requires no calibration.¹⁵ However, we propose that in high income settings,
28 where TB incidence is generally low, an immune based test providing >95% sensitivity for
29 all active TB would be clinically useful if available at low cost. Further work is needed to
30 determine at which cost such an assay would be cost-effective.

Chapter 7. Health Economic Analysis

Introduction

We present methods for assessment of the cost-effectiveness of several novel tests as rule-out and rule-in tests for active TB. That is, in the case of a rule-out test we consider using novel tests as an initial test, with a negative result indicating that a patient does not have TB, thus accelerating diagnosis of the actual cause of disease in such patients. In the case of a rule-in test, we consider using novel tests as an initial test, with a positive result indicating that a patient does have TB, averting the need for other tests for TB. The use of several tests was compared against current practice as determined by analysis of patient records. We considered which diagnostic tests were performed, their costs, and the time taken between decision points involving each test. The time taken to diagnose, rule-in or rule-out TB is a key consideration. Our report adheres to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement.⁶³

Methods

Decision tree model

We developed a decision tree model to calculate the incremental costs and incremental quality-adjusted life-years (QALYs) of changing from current practice to using a novel test as an initial rule-in or rule-out test. Current practice was determined by analysis of patient records. Adding a rule-out test to the diagnostic pathway introduces additional delays in diagnosis of active TB in those patients who have the disease, as it introduces an additional step in the pathway. Patients who were not initially diagnosed with active TB have a follow up consultation after approximately two months; those who had a false negative rule-out test result, i.e. they had TB incorrectly ruled out, can have TB identified at this point. The final diagnostic outcomes were the four categories described in Dosanjh et al¹⁰ herein referred to as “Dosanjh categories”.

The health economic analysis was undertaken from an NHS perspective. No discounting was required since the diagnostic process occurs over a relatively short time period. The time horizon was up to final diagnosis of TB which was assumed to be under year.

The model contained two levels of uncertainty:

1 1. Individual level uncertainty – patient records revealed variation in the number and
2 type of tests used for TB diagnosis, and time to diagnosis; and
3 2. Parameter uncertainty – uncertainty in the costs of tests and procedures, and the
4 sensitivity and specificity of the novel tests.
5 Uncertainty distributions for each are estimated either for Uniform or Gamma distributions
6 depending on the form of the data available.
7
8 A (balanced) bootstrap sample of TB and non-TB patients was created for each simulation of
9 the decision tree, which retained the TB and non-TB subsample sizes. Individual patient costs
10 and time to diagnosis were jointly sampled to preserve the dependency between them. We
11 sampled across the model parameter distributions to obtain a set of realisations . This was
12 then repeated in a Monte Carlo (MC) framework to obtain a sample of 10,000 model runs
13 which encapsulated the first order (patient variability) and second order (parameter)
14 uncertainty in the model outputs. *TABLE 17* summarises the key model parameters.
15 The model was implemented in the statistical programming language R (R Core Team
16 (2016). R: A language and environment for statistical computing. R Foundation for Statistical
17 Computing, Vienna, Austria. <https://www.R-project.org/>). The key model parameters are
18 presented in *TABLE 17*.
19

20 *TABLE 17.* Key model parameters

Parameter	Symbol	Main model value
Follow-up time for those not diagnosed with TB (days)	$T_{followup}$	Direct estimate from clinical dataset
Cohort ATB prevalence	n_+/n	Direct estimate from clinical dataset, also stratified by HIV status
Active TB QALY loss	q	Gamma(5.427, 0.0154) ⁶⁴
Hepatotoxicity due to treatment QALY loss	q_h	Gamma(65.753, 7e-5) ⁶⁴
Nausea due to treatment QALY loss	q_n	Gamma(109.67, 3e-6) ⁶⁴
Current time to diagnosis by TB status	$\bar{t}_+, \bar{t}_-$	Direct estimate from clinical dataset
Current combined cost of diagnosis by TB status	$\bar{c}_+, \bar{c}_-$	Direct estimate from clinical dataset

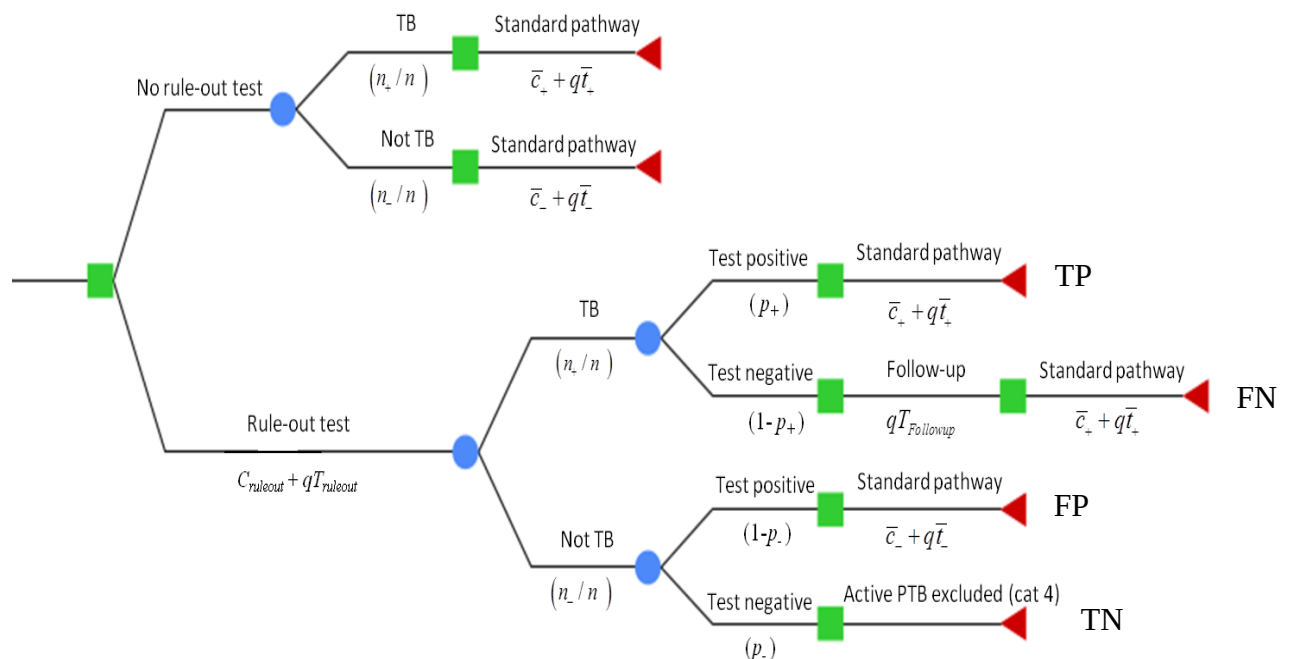


FIGURE 11. Decision tree comparing current practice (“No rule-out test”) with a diagnostic pathway incorporating an initial rule-out test (“Rule-out test”). The “Standard pathway” branch represents the range of variation observed in the patient cohort. Probabilities are shown below branches following a circular chance node and costs are below branches following a square decision node.

Estimation of costs and health impact used in the model

The costs and distributions used in the probabilistic sensitivity analyses are summarised in TABLE 64- TABLE 66 (in Appendix 17) using 2014/2015 prices. Where necessary, costs were inflated from previous years using the Hospital and Community Health Service (HCHS) pay and price index.⁶⁵

Where uncertainty bounds were not available in the recent sources used, uncertainty ranges were informed by previous studies. Uniform distributions were used where upper and lower limits were available and Gamma distributions were used where the standard error of the average cost was available.

When modifying costs the uncertainty around the central value was transformed maintaining the proportions from the original value rather than an absolute difference. For example, a

lower bound of half and an upper bound of twice the point estimate values was used (see⁶⁶, Pareek et al¹⁶, NICE guidance NG33⁹). Skewed distributions were represented using a Gamma distribution.

As the end-point of the diagnosis is diagnosis of TB or ruling-out TB, treatment costs after final diagnosis are out of scope. However, where a patient was started on TB treatment and then a lack of response to that treatment informed a decision that the patient did not in fact have TB, the cost of this treatment was included, since it is part of the cost of ruling out TB for those patients.

The treatment costs are given in *TABLE 67*. Following the NICE guidelines for active TB management, it was assumed that such patients are on treatment until their two-month follow up appointment when they are reassessed for response to treatment. The regimen in this period is a daily treatment with rifampicin, isoniazid, pyrazinamide and ethambutol. From the NICE British National Formulary, adult dosages are fixed except for ethambutol, which is determined by patient weight. The mean weight at time of first presentation of 67.98 kg was used in the model.

Selection of tests to evaluate in health economic analysis

In VANTDET, we have evaluated several signatures using each new technology, but we only wished to evaluate the health economics in the most promising signatures. Since none of the proteomic signatures performed with high enough test accuracy to be clinically useful, these signatures were not used in these analyses. For the transcriptomic signature, the test accuracy data for the Sweeney 4 gene signatures, calculated using the TBS method, was used. For the flow cytometry data, the test accuracy for the %HLA-DR signature in the HIV- cohort was used. For the molecular rule-out test, the performance of the CXCL9 RT-PCR in Mtb antigen stimulated PBMC was used. *TABLE 18* gives the test sensitivity and specificity values used for the probabilistic sensitivity analyses.

1 TABLE 18. Diagnostic test performance and distributions for different tests.

Patient strata	Test	Rule-in		Rule-out	
		Sensitivity	Specificity	Sensitivity	Specificity
All	Transcriptomic	Unif[0.865,0.932]	Unif[0.493,0.609]	Unif[0.245,0.346]	Unif[0.919,0.972]
All	Molecular	-	-	Unif[0.85,1.0]	Unif[0.057,0.44]
HIV negative	Cellular immune flow cytometry	Unif[0.548,0.771]	Unif[0.836,0.971]	-	-

2

3 The QALY loss due to TB treatment are given in *TABLE 17*. The sum total was used in the
4 analysis,

$$q_{Tx} = q_h + q_n$$

5 Cost-effectiveness outcomes

6

7 The current ('status-quo') cost and QALY loss are the weighted sums of the TB and non-TB
8 pathways observed in the clinical data,

9

$$C_{status-quo} = (n_-/n)\bar{c}_- + (n_+/n)\bar{c}_+$$

$$E_{status-quo} = (n_-/n)q\bar{e}_- + (n_+/n)q\bar{e}_+$$

10

11

12 Results

13 For decision tree diagrams, please see *Appendix 17*. The results of the cost effectiveness
14 analysis for the Sweeney transcriptomic signatures used as a rule out test are given in
15 *FIGURE 12* and *TABLE 19*, and used as a rule in test in *FIGURE 13* and *TABLE 20*.

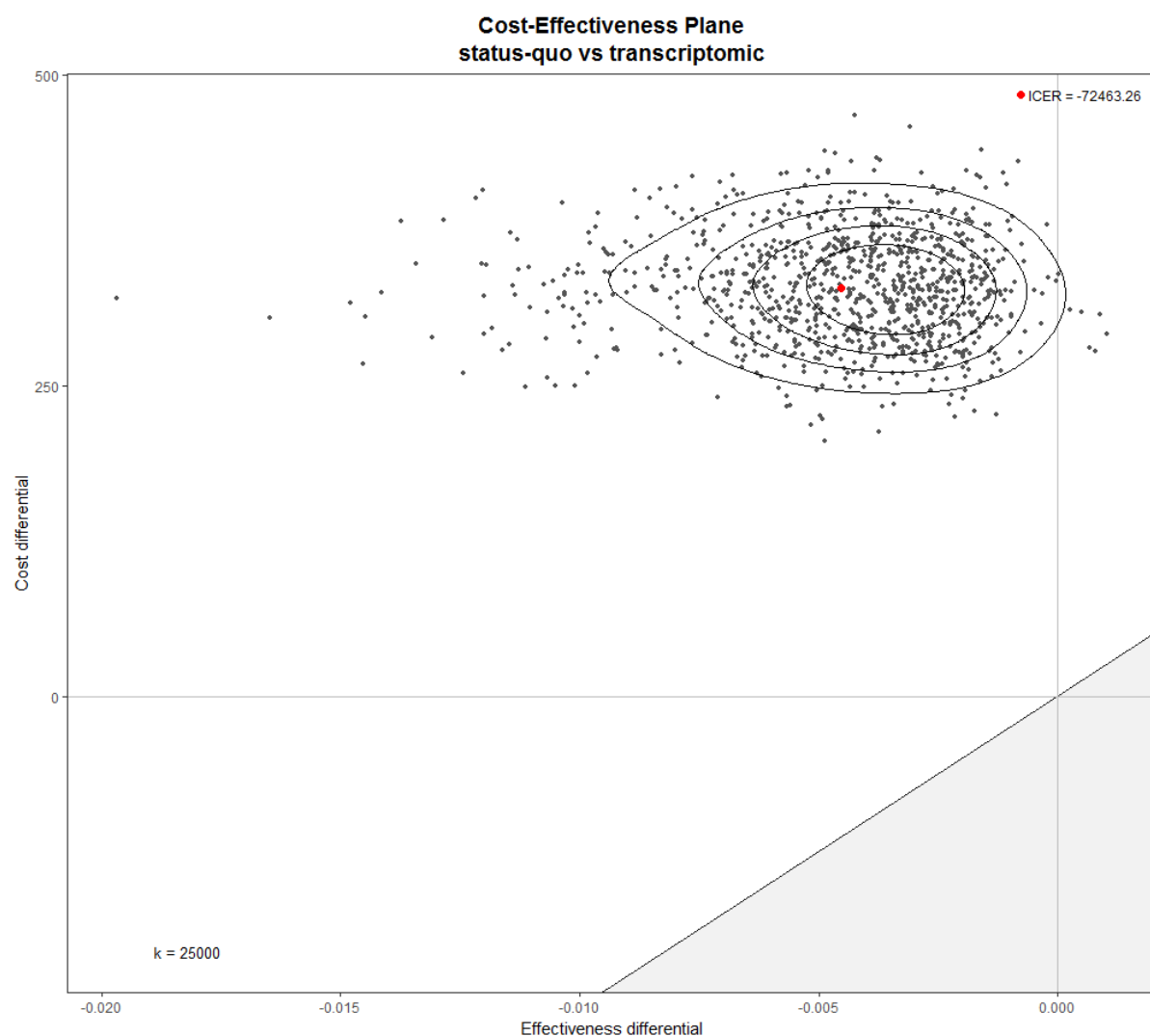


FIGURE 12. Cost-effectiveness planes for status-quo against transcriptomic rule-out test scenario, with contour lines.

TABLE 19. Cost-effectiveness statistics for transcriptomic rule-out scenarios

Scenario	Incremental QALY gain	Incremental cost incurred (£)	Probability cost-effective	
			WTP = £20000/QALY	WTP = £30000/QALY
HIV negative	-0.0045	329	0	0

1 **Rule-in**

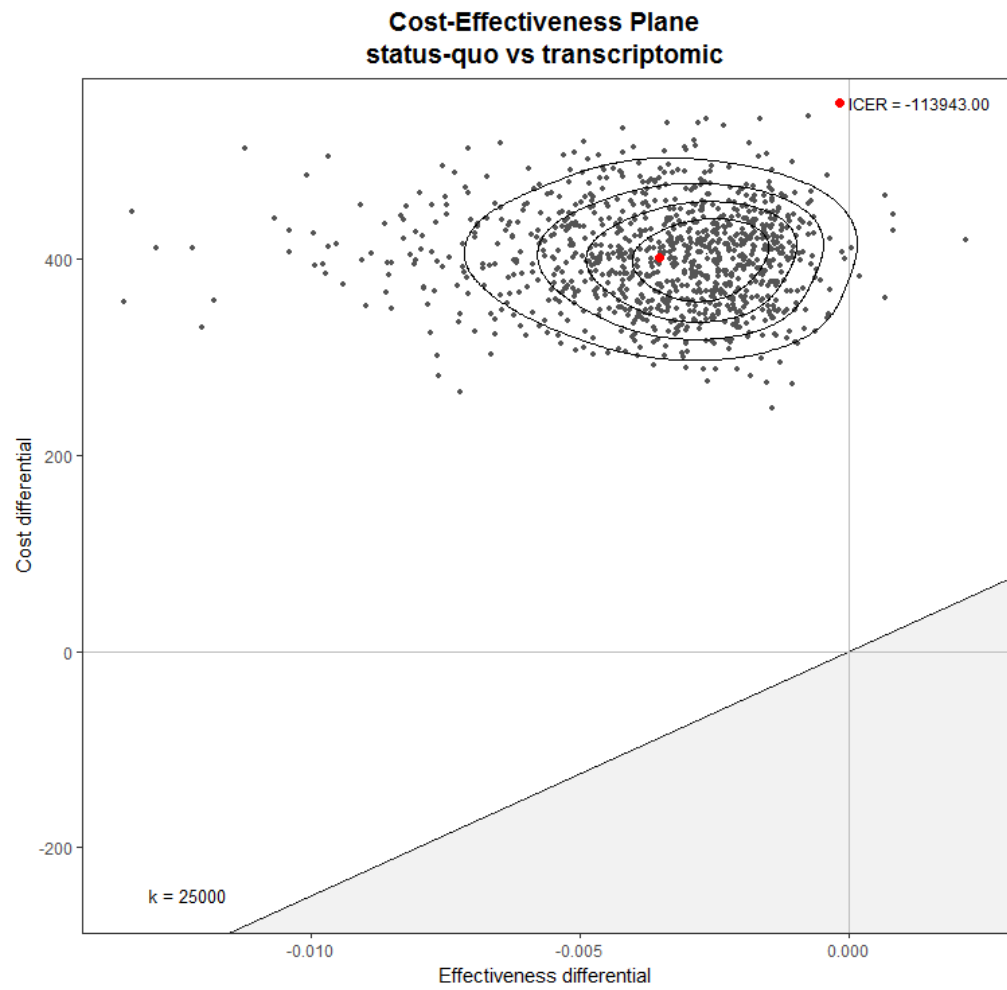


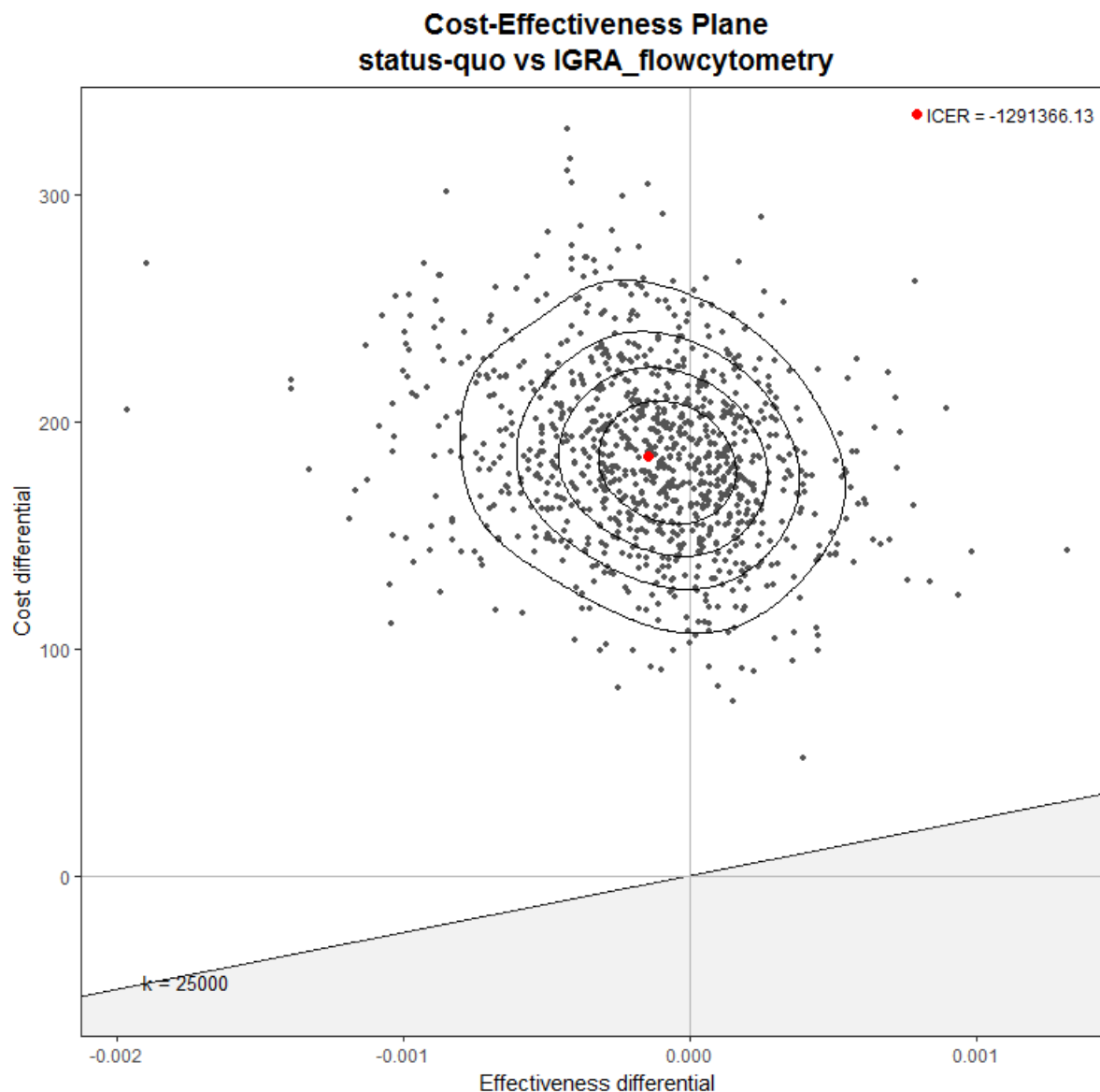
FIGURE 13. **Cost-effectiveness planes for status-quo against transcriptomic rule-in test scenario, with contour lines (right).**

TABLE 20. Cost-effectiveness statistics for transcriptomic rule-in scenarios.

Scenario	Incremental QALY gain	Incremental cost incurred (£)	Probability cost-effective	
			WTP = £20000/QALY	WTP = £30000/QALY
Total	-0.0035	402	0	0

1 Flow cytometry rule-in

2 The results for the cost effectiveness analysis for the %HLA-DR flow cymometry signatures
3 to be used as a rule in test are presented in *FIGURE 14* and *TABLE 21*.



4
5 **FIGURE 14. Cost-effectiveness planes for status-quo against flow cytometry %HLA-DR rule-in test**
6 **scenario, with contour lines.**

7 **TABLE 21. Cost-effectiveness statistics for flow cytometry rule-in scenario.**

Scenario	Incremental QALY gain	Incremental cost incurred (£)	Probability cost-effective	
			WTP = £20000/QAL Y	WTP = £30000/QAL Y
Total	-0.00014	185	0	0

8

Molecular rule-out test

The results for the cost effectiveness analysis for the molecular rule out test are presented in *FIGURE 15* and *TABLE 22*.

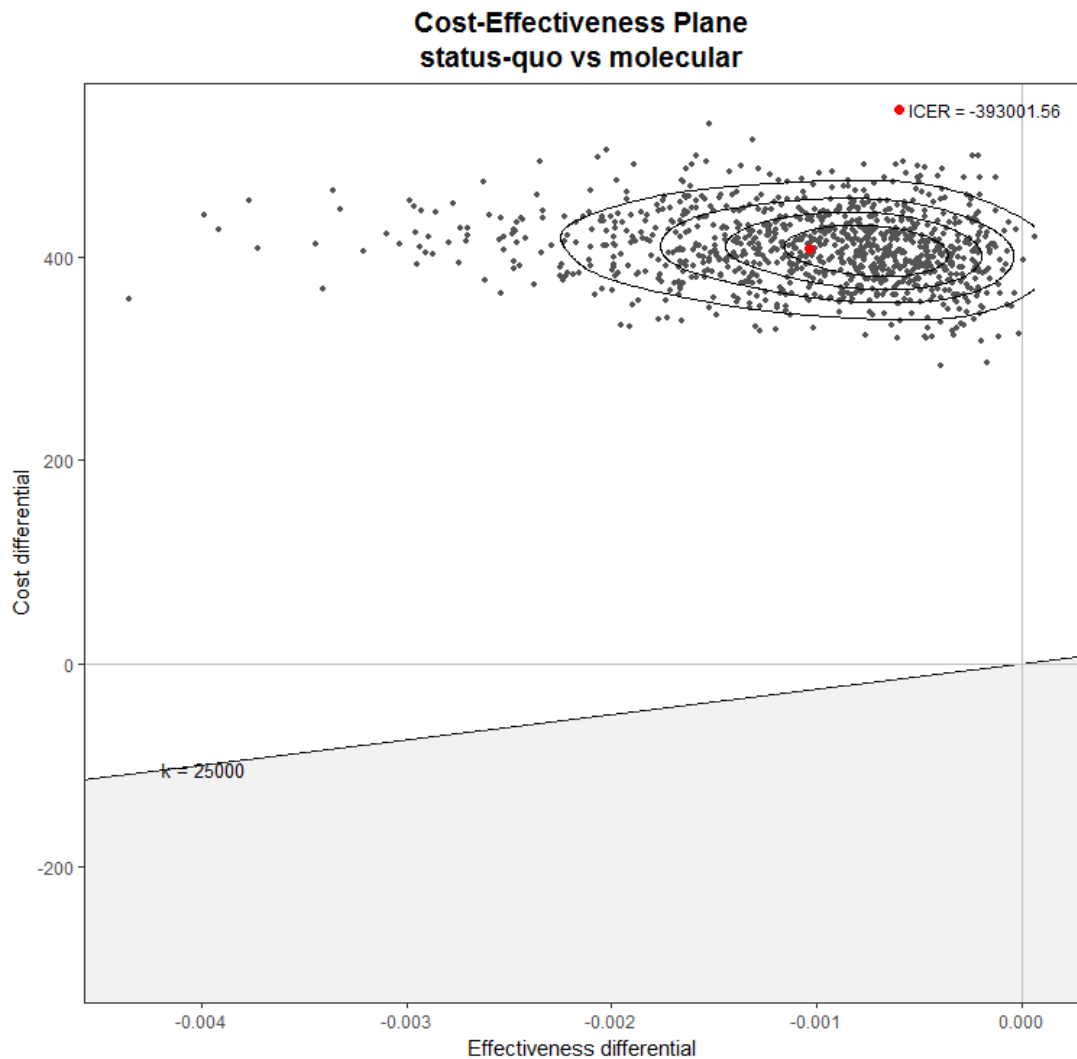


FIGURE 15. Cost-effectiveness planes for status-quo against molecular rule-out test scenario, with contour lines.

TABLE 22. : Cost-effectiveness statistics for molecular rule-out scenarios

Scenario	Incremental QALY gain	Incremental cost incurred (£)	Probability cost-effective	
			WTP = £20000/QALY	WTP = £30000/QALY
Total	-0.0010	408	0	0

Discussion

This health economic analysis has demonstrated that the use of those tests which performed best in classifying all TB cases in our VANTDET sub-studies (i.e. the Sweeney 4 gene transcriptomic signature, the %HLA-DR cellular immune signature, and the CXCL9 qRT-PCR test for ruling out active TB) cannot be considered cost-effective given the assumption of the modelling. This is due in part to the significant cost incurred by these tests, which in their current proposed format, would be very costly. Furthermore, the sensitivity and specificity provided by these tests were sub-optimal, meaning that they did not improve effectiveness when introduced into the current diagnostic pathway. For all tests, extra costs were incurred, and there was a minor reduction in QALYs, likely due to the increase delay in diagnosis to the introduction of new tests into the standard pathway. This was the case even when tests were considered as either rule-in or rule out tests.

It should be noted that, since all of the technologies and signatures are being investigated in research laboratory settings, and have not begun to be commercialised in any way, the cost estimates had a high degree of uncertainty, which impacted on the cost-effectiveness analysis. The effectiveness for the %HLA-DR cellular immune assay, for which we propose using in the context of IGRA+ TB suspects only, the cost-effectiveness plane demonstrated reasonable effectiveness, but a high cost differential. If the test accuracy of this signature could improve slightly, alongside a significant reduction in cost, it might be that this approach could be a cost-effective for diagnosing TB in the IGRA positive population. It is feasible that the cost of some of these technologies will become more economical in the coming years. Furthermore, it is likely that the costs of a commercially made assay using simpler but comparable technologies (i.e. using a simple flow cytometry method vs complex multi-colour analysis) would be lower.

This health economic analysis has some limitations. Given the lack of accurate data on novel test costs, we have relied on estimates based on current laboratory research costs and timings, which may not reflect the likely costs of these tests after they have been developed commercially. As such, there is a high level of uncertainty in our analyses. The uncertainty about the test sensitivities and specificities is represented as marginal distributions when there is a correlation between the two values that may be represented in a joint distribution if the appropriate information is available. Further, we highlight that one of the main assumptions

of the analysis is that we consider a relatively short timeframe up to TB diagnosis and thus cannot extend the conclusions beyond this.

Recommendations

In these analyses, we have considered the introduction of new tests into the pathway of a large number of TB suspects, in complex populations with a diverse mix of TB and OD patients with wide ranging symptoms, disease severities and clinical suspicion of TB. It is likely that a blood-based test for a specific subset of TB, which could be used to assist in diagnosis against a specific subset of OD patients, would be a more cost-effective approach for these blood-based novel tests. In addition, such a test may have improved test performance when used in combination with other tests. We have not yet evaluated such scenarios in this study, but it is a direction for future research on novel tests for diagnosing TB.

For rule-out tests, a test with a combination of high sensitivity and specificity, low cost and rapid speed of result, is required to be effective or cost-effective as a screening tool or as a test to be used at the beginning of a diagnostic pathways for a TB suspect¹⁵.

Future research on the cost-effectiveness of tests would benefit from looking at long term impacts beyond initial diagnosis, and ideally would consider the infectious nature of the condition.

Chapter 8. General Discussion

Principle findings

In this large study to validate new technologies for the diagnostic evaluation of active TB, we successfully evaluated 4 promising approaches for the diagnosis of active TB, all of which used blood samples: transcriptomics, proteomics, flow cytometry and two immunological rule-out tests. In each sub-study, we were able to evaluate multiple previously-identified signatures and assess their diagnostic performance.

For the transcriptomic signature sub study, we report that diagnostic performances ranging from 0.81-0.84 for all previously reported gene expression signatures, for the detection of active TB in the full cohort (using LDA). The performance for all TB and for culture confirmed TB was not found to be clinically useful. Although performance for either culture confirmed TB, or for PTB, was highest, the scenario in which a transcriptomic test is likely to be useful is where current tests fail. As such, despite a maximum AUC of 0.80 for identifying highly probable TB, the Sweeney 4-gene signature may be a useful clinical test for patients who test negative with microbiological tests. For culture negative TB, this signature could provide a high NPV (0.95), which may be clinically useful.

None of the previously identified proteomic signatures were able to provide high diagnostic accuracy for detecting active TB cases, with the maximum test accuracy being for the Levin 4 protein signature, detected by SLEDI TOF, which provided an AUC of 0.74 for culture confirmed TB using the DRS method. We were able to compare this signature across 2 proteomic technology platforms: SELDI TOF and LC-MS and showed relatively poor corroboration between the 2 platforms. We were able to assess the performance of 4 published signatures using LC-MS, none of which performed well for detecting TB (not culture confirmed or PTB groups).

Of the six previously identified cellular immune signatures for active TB, the %HLA-DR signature was found to perform the best, providing an AUC of 0.90 for all TB in the HIV-population. The %HLA-DR and %TEFF tests provide high PPV for all TB and for culture confirmed TB in IGRA+ TB suspects, and as such, may be clinically useful. Only the signatures which measured the phenotype (either memory/differentiation phenotype, or

activation status) provided discriminatory ability. Those signatures measuring function alone did not perform well.

We tested two novel approaches for the development of a rule-out test. We found that detection of CXCL9 by qRT-PCR of stimulated PBMC provided a sensitive assay which could detect a high proportion of TB cases. However, the assay has a high failure rate and was also hampered by poor specificity. The use of CXCL9 for discriminating between TB and OD/LTBI in immune assays should be considered.

None of the signatures validated were found to be costs effective when they evaluated for their use to rule in or rule out active TB, in relation to the current standard diagnostic pathway in the UK. The signature which performed best was the %HLA-DR flow cytometry signature for the ruling in of all TB in the IGRA+ population. However, this signature resulted in significant cost incurred and a slight reduction in QALY.

Strengths and limitations

In IDEA, we consecutively recruited a unique, large, multicenter, prospective cohort from, where individuals were suspected of having active TB as part of routine clinical practice. The cohorts we used in these validation studies (within the exception of the rule out test where an artificial cohort was selected) represent the full spectrum of clinical disease reflecting heterogeneity of all clinical phenotypes and other differential diagnoses. The study participants, with suspected TB symptoms but unknown diagnosis at point of enrolment, represent the uncertainty faced by clinicians. The high resource setting provides a high quality diagnostic facility and reliable and accurate diagnosis for culture unconfirmed, highly probable TB.

Ultimately, our study provided a real-life road test for the published signatures and evaluated their clinical utility, a crucial test before they enter next phases of biomarker development. We present the performance of the previously published signatures in our hands. The previously proposed thresholds for each signature were not reported here, since overall these thresholds provided sub-optimal performance in our cohort. In our experience with diagnostic tests where the outcome is a continuous, the threshold for optimal diagnostic performance does change between laboratories and cohorts, for a number of reasons, particularly when these tests are yet to be standardised and where the cohorts are so variable. We therefore took several approaches to define optimal thresholds and their associated performance

characteristics. Our findings suggest that the performance of all signatures and technologies in detecting TB in the HIV+ population was very poor, much worse than in the HIV- population. This is not in line with other studies, where performance of signatures in HIV+ has either been comparable to HIV- populations, or marginally worse^{19, 31, 52}. We were likely under powered to fully assess diagnostic performance in the HIV+ population, due to the small number of these patients in the parent IDEA cohort. As such, further validation is required in this population, in particular for the proteomic and cellular immune signature validations. Within the IDEA cohort, there is a small number of patients who are classified as ‘clinically indeterminate’ (Dosanjh 3) who were not included within this study, since assessing the performance of new tests within this population would give no known assessment of performance against gold standard/current practice. Nevertheless, they represent a sub-set of TB suspects who present in real life clinical practice. We did, however, included this population when all calculations of pre-test probability for TB subsets were made.

In addition to providing detailed statistics on the diagnostic accuracy of the previous published signatures, we herein present complex health economic analyses, where the most promising signatures were evaluated for cost effectiveness for either the rule-in or rule-out of active TB. However, we did not present health economic analysis for all sub-groups (i.e culture confirmed TB alone, or PTB alone), and it is likely that effectiveness in these scenarios will improve since the cohort of patients becomes smaller and the clinical question narrower.

We are lacking some analyses. In particular, we do not compare different signatures with each other using a formal statistical comparison. Furthermore, there is a need to do further sub-group analyses using these data. For example, to perform health economic analysis in particular clinical scenarios (i.e. within the smear negative groups), since a conclusion of this study is that these new tests are likely to be useful in specific types of TB-suspects.

Recommendations for clinical practice

Given the poor performance of each of the new technologies for detecting the full spectrum of active TB cases in this study, and their lack of cost effectiveness for this purpose, it is unlikely that the findings of this project will immediately have implications for changing current practice when it comes to the diagnostic pathways for TB suspects in the UK. None of the tests evaluated had sufficiently high diagnostic accuracy to provide a clinically useful

rule-in or rule-out function over and above the clinical utility of the currently available diagnostic tests. However, diagnostic assessment of certain key groups of patients could potentially be improved by the flow-cytometric and the transcriptomic tests. Smear microscopy results are available on the same day that the patient presents to secondary care and IGRA results are usually available by the next day. The role of new tests needs to be considered in the context of patients with suspected TB where both these test results are known. In patients with suspected TB who are IGRA-positive and smear-negative patients there is commonly a high index of suspicion for TB but a definitive diagnosis is not possible until culture results become available, typically after 2-3 weeks. A test that could rule-in or rule-out TB in these patients is thus an important unmet clinical need.

The %HLA-DR and %T_{EFF} flow-cytometry tests gave 94% PPV in the smear-negative IGRA-positive group, providing what is effectively a rule-in test and facilitating a decision to start treatment for presumptive TB. The test could thus form part of the diagnostic assessment pathway forming part of an algorithm alongside smear microscopy and IGRA. Interestingly, two transcriptomic signatures provided high NPVs of 89-90% for culture-positive TB in the smear-negative population (unstratified by IGRA result). These NPVs may thus provide clinicians with sufficient basis for withholding treatment (or at least delaying it until culture results are available) in smear-negative TB suspects who test negative with these signatures.

Recommendations for future research

Smear still remains a quick, cheap and specific method for quickly ruling in TB in cases with high bacterial load. Therefore a new test is not required for this sub-population. However, it remains the case that research into new approaches to diagnosing TB focus on the performance of the test in this population. The clinical need is in fact in the smear negative patients where there is currently not a quick test available for either rule in or rule out of active TB. This is particularly the case in low prevalence settings such as the UK, where a high proportion of those diagnosed with TB are paucibacillary and/or EPTB.

We have demonstrated that the transcriptomic test is able to provide a reasonably high NPV in those TB suspects who have a negative smear result. These tests are likely to be cheap and rapid if developed commercially and, if found to be reproducible using a rapid qRT-PCR approach, could potentially be developed into near-patient or point-of-care tests. For the reasons already discussed, we propose that future research into the use of these tests focuses on the smear-negative TB suspects, and includes those with clinically diagnosed (i.e culture-

negative) TB. The high accuracy for diagnosing TB using the Adekambi and Pollock flow cytometry assays reported in chapter 5 suggests a clinically useful role for these tests to help rule-in a diagnosis of TB in smear-negative IGRA-positive patients where this is currently a clear unmet clinical need. Accordingly, further development, standardisation and simplification of these assays should be prioritised in order to deliver an in vitro diagnostic (IVD) for routine clinical use. An automated flow cytometer set-up which allows for simple assay preparation and automated gating and analysis would help to make the tests suitable for routine diagnostic laboratories. It is unlikely that these tests would be applicable to low-income high burden settings, given their complexity and costs. We also propose that the tests are further validated within the specific sub-populations for which they would be useful, i.e smear negative TB suspects.

An overall conclusion from this study is that blood-based tests for TB can perform best when specific sub-types of TB are considered, and when specific subsets of OD can be excluded. As such, we recommend that future studies exploring these technologies focus on specific sub-types of TB. Most usefully, those groups where a diagnostic test is most needed, i.e EPTB, culture/smear negative TB.

It appears that while the ‘omic technologies are hampered by specific differential diagnoses (e.g Sarcoidosis), for the cellular immune signatures, where immune cells are stimulated with Mtb antigens before the test is performed, there is no sub-group within the OD population which hampered diagnostic performance. As such, we conclude that immune based tests are more useful when all TB is considered (with the exception, perhaps, of those infected with HIV).

Conclusion

In conclusion, in this report we presented the validation of new technologies for the diagnostic evaluation of TB, and have fully evaluated their use in a real life clinical setting. Overall, the diagnostic performance of all previously identified ‘signatures’ of TB was lower than previously reported. This likely reflects the nature of the cohort we used, which includes the harder to diagnose groups, such as culture unconfirmed TB, and EPTB, which were underrepresented in previous cohorts.

We have addressed all of our primary objectives. Some analyses are outstanding; importantly, we are yet to fully address our secondary objective of deriving novel signatures for active TB

1 using the data we generated in the sub-studies. This is an area of ongoing investigation by our
2 group and was beyond the scope of this report. This study provided a real-life road test for
3 these technologies and highlights the need for new diagnostic and signatures for TB to be
4 derived and evaluated in cohorts which included the full spectrum of TB cases, and where the
5 hard to diagnose sub-groups of TB are considered.

6

Acknowledgements

We would like to thank the following individuals for helpful in providing either technical assistance, analytical support and/or advice on the interpretation of data during the project:

Mr Elliot Harper, Dr Yanping Guo, Dr Luis Berrocal Almanza, Dr Ishita Marwah, Mrs Amarjit Badhan, Dr Paul Golby, Dr Hilary Whitworth, Dr Darren Chooneea, Dr Christopher Partlett, Dr Umar Niazi, Dr Lisa Haigh (Chemistry, Imperial College London), The IDEA team, Miss Gulfer Yackici, Miss Heather Lambie, Miss Helen Piotrowski, Mr Samuel Bremang.

We are grateful to the following centres for providing technical support in terms of conducting experiments:

Lille Transcriptomic Facility (Institute Pasteur Lille), Dundee Proteomics Facility, Proteomics Facility Imperial College London, St Mary's Flow Cytometry Facility (Imperial College London).

We thank Mr Dhansuklal Solanki, Mrs Claire Puddephatt, and Mrs Tingting Wu for administrative and financial management assistance.

We would also like to thank the patients recruited into the IDEA study, who contributed the samples which were used in this study.

Independent Scientific Advisory Board

Professor Sanjeev Krishna (Chair for ISAB meetings), Molecular Parasitology and Medicine, St George's University of London, London, UK

Dr Mahdad Noursadeghi (Clinical Senior Lecturer), Infection and Immunity, University College London, London, UK

Dr David Connell (Respiratory Registrar), The Royal Brompton Hospital, London, UK

Members of the Study Management Group

Professor Ajit Lalvani, Professor Onn Min Kon, Professor Peter Beverly, Dr Alice Halliday, Dr Long Hoang, Dr Pooja Jain, Dr Robert Parker, Mrs Mica Tolosa-Wright, Miss Tereza Masonou

Members of Wider Study Management Group

Professor Ajit Lalvani, Professor Onn Min Kon, Professor Peter Beverly, Dr Yemisi Takwoingi, Professor John Deeks, Dr Alice Halliday, Dr Long Hoang, Dr Robert Parker, Mica Tolosa-Wright, Miss Aime Boakye.

Contribution of authors

Dr Alice Halliday (Research Associate) led the cellular immune signature validation sub-study, oversaw the candidate rule out validation studies, managed the overall VANTDET project co-ordination, co-ordinated the write up of the report, and contributing to writing all sections.

Dr Pooja Jain (Bioinformatician) led the statistical analysis for the transcriptomics and proteomics sub-studies and conducted the main statistical analyses for these two chapters, and supported the writing of methods, results and discussion sections in these chapters.

Dr Long Hoang (Research Associate) led the validation of transcriptomic signature sub-studies, oversaw the validation of the molecular rule-out test, conducted statistical analysis and supporting the write up of the transcriptomic chapter.

Dr Robert Parker (Research Associate) led the validation of proteomic signatures sub-study, and led the writing of this chapter of the report.

Mrs Mica Tolosa-Wright (Research Technician), with the support of Alice Halliday and Long Hoang, led the validation of the molecular rule-out test sub-study, and led the writing of this chapter (6).

Miss Tereza Masonou (MSc student and Research Technician) supported the validation of cellular immune signature experimental work, analysis and interpretation of the data. She also supported the report write up by writing the methodology section for Chapter (5) and production of results tables.

1 Dr Nathan Green (Health Economics Statistician) led the health economic analysis for the
2 project, and led the write up of the

3 Miss Aime Boakye (IDEA Administrator) managed and coordinated the clinical database for
4 the IDEA patients, and coordinated the cohort selection for the separate sub-studies within
5 VANTDET. She ensured that the research staff remained blind to clinical data throughout the
6 study period, until analysis phase. She also supported and led many of the public engagement
7 and involvement activities.

8 Dr Shea Hamilton (Research Associate) performed the experimental work for the validation
9 of the 4 protein signature by SELD-TOF technology

10 Mr Vinay Mandagere designed and conducted the Quantiferon supernatant project in the
11 candidate rule out test chapter.

12 Dr Anastasia Fries designed and conducted the Quantiferon supernatant project in the
13 candidate rule out test chapter.

14 Dr Lachlan Coin analysed the SELDI-TOF signature validation data.

15 Dr Yemisi Takwoingi provided statistical analysis oversight, support, and advice on study
16 design throughout the project.

17 Dr Peter White was a co-investigator on the project, contributed to the study design and led
18 the statistical analysis of diagnostic pathways and the economic analysis.

19 Professor Michael Levin was a co-investigator on the project and oversaw the SELDI
20 proteomic signature validation

21 Professor John Deeks was a co-investigator on the project, contributed to the study design
22 and support of statistical analysis approaches

23 Professor Peter Beverley stood in for Professor Ajit Lalvani during his leave of absence (May
24 2015-May 2016), oversaw the project, provided clinical expertise, and contributed to the
25 study design, and data interpretation

26 Professor Onn Min Kon was a co-investigator on the project, and provided clinical expertise,
27 and contributed to the study design, data interpretations, execution and the writing of the
28 report.

1 Professor Ajit Lalvani (Professor and Consultant Physician Imperial College Healthcare) was
2 chief investigator, provided clinical expertise, and contributed to the study design, data
3 interpretations, execution and the writing of the report.

5 **Data Sharing Statement**

6 All data requests should be submitted to the corresponding author for consideration. Please
7 note exclusive use will be retained until the publication of major outputs. Access to
8 anonymised data may be granted following review.

9 **Funding**

10 This project was funded by the NIHR Efficacy and Mechanisms Evaluation (EME)
11 Programme and will be published in full. See the NIHR Journals Library EME project pages
12 for further information.

References

1. Organization WH. Global tuberculosis report 2017. 2017.
2. . *Tuberculosis in England 2017 report*. London: PHE publications; 2017.
3. Houben RM, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Med* 2016;**13**:e1002152. <https://doi.org/10.1371/journal.pmed.1002152>
4. Behr MA, Edelstein PH, Ramakrishnan L. Revisiting the timetable of tuberculosis. *BMJ* 2018;**362**:k2738. <https://doi.org/10.1136/bmj.k2738>
5. Uplekar M, Weil D, Lonnroth K, Jaramillo E, Lienhardt C, Dias HM, *et al*. WHO's new End TB Strategy. *The Lancet* 2015;**385**:1799-801. [https://doi.org/https://doi.org/10.1016/S0140-6736\(15\)60570-0](https://doi.org/10.1016/S0140-6736(15)60570-0)
6. Department of Health. Stopping Tuberculosis in England. An Action Plan from the Chief Medical Officer. 2004.
7. (UK) NifHaCE. *Tuberculosis: NICE guideline*; 2016.
8. Dorman SE, Schumacher SG, Alland D, Nabeta P, Armstrong DT, King B, *et al*. Xpert MTB/RIF Ultra for detection of Mycobacterium tuberculosis and rifampicin resistance: a prospective multicentre diagnostic accuracy study. *Lancet Infect Dis* 2018;**18**:76-84. [https://doi.org/10.1016/S1473-3099\(17\)30691-6](https://doi.org/10.1016/S1473-3099(17)30691-6)
9. NICE. *Clinical diagnosis and management of tuberculosis, and measures for its prevention and control*. National Institute of Clinical Excellence; 2011. URL: <http://www.nice.org.uk/guidance/index.jsp?action=byID&o=13422> (accessed).
10. Dosanjh DP, Hinks TS, Innes JA, Deeks JJ, Pasvol G, Hackforth S, *et al*. Improved diagnostic evaluation of suspected tuberculosis. *Ann Intern Med* 2008;**148**:325-36.
11. Yemisi Takwoingi HW, Melanie Rees-Roberts, Amarjit Badhan, Christopher Partlett, Nathan Green, Aime Boakye, Heather Lambie, Luigi Marongiu, Mark Jit, Peter White, Jonathan Deeks, Onn Min Kon, Ajit Lalvani. Interferon-gamma release assays for Diagnostic Evaluation of Active tuberculosis (IDEA): test accuracy study and economic evaluation. *Health Technology Assessment* 2017;**Uncorrected Proof**. <https://doi.org/10.3310/hta21XXX>
12. Thillai M, Pollock K, Pareek M, Lalvani A. Interferon-gamma release assays for tuberculosis: current and future applications. *Expert Rev Respir Med* 2014;**8**:67-78. <https://doi.org/10.1586/17476348.2014.852471>
13. Hilary S Whitworth AB, Aime A Boakye, Yemisi Takwoingi, Melanie Rees-Roberts, Christopher Partlett, Heather Lambie, John Innes, Graham Cooke, Marc Lipman, Christopher Conlon, Derek Macallan, Felix Chua, Frank Post, Martin Wiselka, Gerrit Woltmann, Jonathan J Deeks, Onn Min Kon & Ajit Lalvani on behalf of the IGRAs for Diagnostic Evaluation of Active TB (IDEA) Study Group. An observational cohort study to evaluate the clinical utility of current and second-generation interferon-gamma release-assays in diagnostic evaluation of tuberculosis. In. *Lancet Infectious Diseases*; 2018.
14. Abubakar I, Drobniewski F, Southern J, Sitch AJ, Jackson C, Lipman M, *et al*. Prognostic value of interferon-gamma release assays and tuberculin skin test in predicting the development of active tuberculosis (UK PREDICT TB): a prospective cohort study. *Lancet Infect Dis* 2018;**18**:1077-87. [https://doi.org/10.1016/S1473-3099\(18\)30355-4](https://doi.org/10.1016/S1473-3099(18)30355-4)
15. WHO. *High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting* Geneva; 2014.
16. Pareek M, Bond M, Shorey J, Seneviratne S, Guy M, White P, *et al*. Community-based evaluation of immigrant tuberculosis screening using interferon γ release assays and tuberculin skin testing: observational study and economic analysis. *Thorax* 2013;**68**:230-9. <https://doi.org/10.1136/thoraxjnl-2011-201542>

17. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010;**466**:973-7. <https://doi.org/10.1038/nature09247>
18. De Groote MA, Sterling DG, Hraha T, Russell TM, Green LS, Wall K, *et al.* Discovery and Validation of a Six-Marker Serum Protein Signature for the Diagnosis of Active Pulmonary Tuberculosis. *J Clin Microbiol* 2017;**55**:3057-71. <https://doi.org/10.1128/JCM.00467-17>
19. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, *et al.* Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med* 2013;**10**:e1001538. <https://doi.org/10.1371/journal.pmed.1001538>
20. Achkar JM, Cortes L, Croteau P, Yanofsky C, Mentinova M, Rajotte I, *et al.* Host Protein Biomarkers Identify Active Tuberculosis in HIV Uninfected and Co-infected Individuals. *EBioMedicine* 2015;**2**:1160-8. <https://doi.org/10.1016/j.ebiom.2015.07.039>
21. Millington KA, Fortune SM, Low J, Garces A, Hingley-Wilson SM, Wickremasinghe M, *et al.* Rv3615c is a highly immunodominant RD1 (Region of Difference 1)-dependent secreted antigen specific for Mycobacterium tuberculosis infection.
22. Kasprovicz VO, Mitchell JE, Chetty S, Govender P, Huang KH, Fletcher HA, *et al.* A molecular assay for sensitive detection of pathogen-specific T-cells. *PLoS One* 2011;**6**:e20606. <https://doi.org/10.1371/journal.pone.0020606>
23. Harari A, Rozot V, Bellutti Enders F, Perreau M, Stalder JM, Nicod LP, *et al.* Dominant TNF-alpha+ Mycobacterium tuberculosis-specific CD4+ T cell responses discriminate between latent infection and active disease. *Nat Med* 2011;**17**:372-6. <https://doi.org/10.1038/nm.2299>
24. Pollock KM, Whitworth HS, Montamat-Sicotte DJ, Grass L, Cooke GS, Kapembwa MS, *et al.* T-cell immunophenotyping distinguishes active from latent tuberculosis. *J Infect Dis* 2013;**208**:952-68. <https://doi.org/10.1093/infdis/jit265>
25. Adekambi T, Ibegbu CC, Cagle S, Kalokhe AS, Wang YF, Hu Y, *et al.* Biomarkers on patient T cells diagnose active tuberculosis and monitor treatment response. *The Journal of Clinical Investigation* 2015;**125**:1827-38. <https://doi.org/10.1172/JCI77990>
26. Portevin D, Moukambi F, Clowes P, Bauer A, Chachage M, Ntinginya NE, *et al.* Assessment of the novel T-cell activation marker–tuberculosis assay for diagnosis of active tuberculosis in children: a prospective proof-of-concept study. *The Lancet Infectious Diseases*; **14**:931-8. [https://doi.org/10.1016/S1473-3099\(14\)70884-9](https://doi.org/10.1016/S1473-3099(14)70884-9)
27. Petruccioli E, Petrone L, Vanini V, Cuzzi G, Navarra A, Gualano G, *et al.* Assessment of CD27 expression as a tool for active and latent tuberculosis diagnosis. *Journal of Infection* 2015;**71**:526-33. <https://doi.org/http://dx.doi.org/10.1016/j.jinf.2015.07.009>
28. Anderson ST, Kaforou M, Brent AJ, Wright VJ, Banwell CM, Chagaluka G, *et al.* Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Engl J Med* 2014;**370**:1712-23. <https://doi.org/10.1056/NEJMoa1303657>
29. Maertzdorf J, McEwen G, Weiner J, 3rd, Tian S, Lader E, Schriek U, *et al.* Concise gene signature for point-of-care classification of tuberculosis. *EMBO Mol Med* 2016;**8**:86-95. <https://doi.org/10.15252/emmm.201505790>
30. Roe JK, Thomas N, Gil E, Best K, Tsaliki E, Morris-Jones S, *et al.* Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. *JCI Insight* 2016;**1**:e87238. <https://doi.org/10.1172/jci.insight.87238>
31. Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir Med* 2016;**4**:213-24. [https://doi.org/10.1016/S2213-2600\(16\)00048-5](https://doi.org/10.1016/S2213-2600(16)00048-5)
32. Hoang LT, Lynn DJ, Henn M, Birren BW, Lennon NJ, Le PT, *et al.* The early whole-blood transcriptional signature of dengue virus and features associated with progression to dengue shock syndrome in Vietnamese children and young adults. *J Virol* 2010;**84**:12982-94. <https://doi.org/JVI.01224-10> [pii]

- 10.1128/JVI.01224-10
33. Team RC. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* 2017.
34. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 2008;**24**:1547-8. <https://doi.org/10.1093/bioinformatics/btn224>
35. Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, Nielser HB, *et al.* A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol* 2002;**3**:research0048.
36. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;**3**:32-5.
37. Cuevas LE, Browning R, Bossuyt P, Casenghi M, Cotton MF, Cruz AT, *et al.* Evaluation of tuberculosis diagnostics in children: 2. Methodological issues for conducting and reporting research evaluations of tuberculosis diagnostics for intrathoracic tuberculosis in children. Consensus from an expert panel. *J Infect Dis* 2012;**205** Suppl 2:S209-15. <https://doi.org/10.1093/infdis/jir879>
38. Agranoff D, Fernandez-Reyes D, Papadopoulos MC, Rojas SA, Herbster M, Loosemore A, *et al.* Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. *Lancet* 2006;**368**:1012-21. [https://doi.org/10.1016/S0140-6736\(06\)69342-2](https://doi.org/10.1016/S0140-6736(06)69342-2)
39. Ratzinger F, Bruckschwaiger H, Wischenbart M, Parschalk B, Fernandez-Reyes D, Lagler H, *et al.* Rapid Diagnostic Algorithms as a Screening Tool for Tuberculosis: An Assessor Blinded Cross-Sectional Study. *PLOS ONE* 2012;**7**:e49658. <https://doi.org/10.1371/journal.pone.0049658>
40. Liu Q, Chen X, Hu C, Zhang R, Yue J, Wu G, *et al.* Serum protein profiling of smear-positive and smear-negative pulmonary tuberculosis using SELDI-TOF mass spectrometry. *Lung* 2010;**188**:15-23. <https://doi.org/10.1007/s00408-009-9199-6>
41. Zhang J, Wu X, Shi L, Liang Y, Xie Z, Yang Y, *et al.* Diagnostic serum proteomic analysis in patients with active tuberculosis. *Clin Chim Acta* 2012;**413**:883-7. <https://doi.org/10.1016/j.cca.2012.01.036>
42. Sandhu G, Battaglia F, Ely BK, Athanasakis D, Montoya R, Valencia T, *et al.* Discriminating active from latent tuberculosis in patients presenting to community clinics. *PLoS One* 2012;**7**:e38080. <https://doi.org/10.1371/journal.pone.0038080>
43. Liu J, Jiang T, Wei L, Yang X, Wang C, Zhang X, *et al.* The discovery and identification of a candidate proteomic biomarker of active tuberculosis. *BMC Infect Dis* 2013;**13**:506. <https://doi.org/10.1186/1471-2334-13-506>
44. Song SH, Han M, Choi YS, Dan KS, Yang MG, Song J, *et al.* Proteomic profiling of serum from patients with tuberculosis. *Ann Lab Med* 2014;**34**:345-53. <https://doi.org/10.3343/alm.2014.34.5.345>
45. Xu D, Li Y, Li X, Wei LL, Pan Z, Jiang TT, *et al.* Serum protein S100A9, SOD3, and MMP9 as new diagnostic biomarkers for pulmonary tuberculosis by iTRAQ-coupled two-dimensional LC-MS/MS. *Proteomics* 2015;**15**:58-67. <https://doi.org/10.1002/pmic.201400366>
46. Li C, He X, Li H, Zhou Y, Zang N, Hu S, *et al.* Discovery and verification of serum differential expression proteins for pulmonary tuberculosis. *Tuberculosis (Edinb)* 2015;**95**:547-54. <https://doi.org/10.1016/j.tube.2015.06.001>
47. Chegou NN, Sutherland JS, Malherbe S, Crampin AC, Corstjens PL, Geluk A, *et al.* Diagnostic performance of a seven-marker serum protein biosignature for the diagnosis of active TB disease in African primary healthcare clinic attendees with signs and symptoms suggestive of TB. *Thorax* 2016;**71**:785-94. <https://doi.org/10.1136/thoraxjnl-2015-207999>
48. Reuschl AK, Edwards MR, Parker R, Connell DW, Hoang L, Halliday A, *et al.* Innate activation of human primary epithelial cells broadens the host response to Mycobacterium tuberculosis in the airways. *PLoS Pathog* 2017;**13**:e1006577. <https://doi.org/10.1371/journal.ppat.1006577>
49. Herberg JA, Kaforou M, Wright VJ, *et al.* Diagnostic test accuracy of a 2-transcript host rna signature for discriminating bacterial vs viral infection in febrile children. *JAMA* 2016;**316**:835-45. <https://doi.org/10.1001/jama.2016.11236>

50. Rozot V, Patrizia A, Vigano S, Mazza-Stalder J, Idrizi E, Day CL, *et al.* Combined Use of Mycobacterium tuberculosis-Specific CD4 and CD8 T-Cell Responses Is a Powerful Diagnostic Tool of Active Tuberculosis. *Clin Infect Dis* 2014; 10.1093/cid/ciu795. <https://doi.org/10.1093/cid/ciu795>
51. Wilkinson KA, Oni T, Gideon HP, Goliath R, Wilkinson RJ, Riou C. Activation Profile of Mycobacterium tuberculosis–Specific CD4+ T Cells Reflects Disease Activity Irrespective of HIV Status. *American Journal of Respiratory and Critical Care Medicine* 2016;**193**:1307-10. <https://doi.org/10.1164/rccm.201601-0116LE>
52. Riou C, Berkowitz N, Goliath R, Burgers WA, Wilkinson RJ. Analysis of the Phenotype of Mycobacterium tuberculosis-Specific CD4+ T Cells to Discriminate Latent from Active Tuberculosis in HIV-Uninfected and HIV-Infected Individuals. *Front Immunol* 2017;**8**:968. <https://doi.org/10.3389/fimmu.2017.00968>
53. Sester U, Fousse M, Dirks J, Mack U, Prasse A, Singh M, *et al.* Whole-blood flow-cytometric analysis of antigen-specific CD4 T-cell cytokine profiles distinguishes active tuberculosis from non-active states. *PloS one* 2011;**6**:e17813. <https://doi.org/10.1371/journal.pone.0017813>
54. Day CL, Abrahams DA, Lerumo L, Janse van Rensburg E, Stone L, O’rie T, *et al.* Functional Capacity of Mycobacterium tuberculosis-Specific T Cell Responses in Humans Is Associated with Mycobacterial Load. *The Journal of Immunology* 2011;**187**:2222-32. <https://doi.org/10.4049/jimmunol.1101122>
55. Goletti D, Butera O, Vanini V, Lauria FN, Lange C, Franken KLMC, *et al.* Response to Rv2628 latency antigen associates with cured tuberculosis and remote infection. *European Respiratory Journal* 2010;**36**:135-42. <https://doi.org/10.1183/09031936.00140009>
56. Takasaki J, Manabe T, Morino E, Muto Y, Hashimoto M, Iikura M, *et al.* Sensitivity and specificity of QuantiFERON-TB Gold Plus compared with QuantiFERON-TB Gold In-Tube and T-SPOT.TB on active tuberculosis in Japan. *Journal of Infection and Chemotherapy* 2018;**24**:188-92. <https://doi.org/https://doi.org/10.1016/j.jiac.2017.10.009>
57. Lalvani A, Millington KA. T Cells and Tuberculosis: Beyond Interferon-gamma. *The Journal of infectious diseases* 2008;**197**:941-3. <https://doi.org/10.1086/529049>
58. Blauenfeldt T, Heyckendorf J, Graff Jensen S, Lange C, Drabe C, Hermansen TS, *et al.* Development of a One-Step Probe Based Molecular Assay for Rapid Immunodiagnosis of Infection with *M. tuberculosis* Using Dried Blood Spots. *PLoS ONE* 2014;**9**:e105628. <https://doi.org/10.1371/journal.pone.0105628>
59. Tebruegge M, Dutta B, Donath S, Ritz N, Forbes B, Camacho-Badilla K, *et al.* Mycobacteria-Specific Cytokine Responses Detect Tuberculosis Infection and Distinguish Latent from Active Tuberculosis. *Am J Respir Crit Care Med* 2015;**192**:485-99. <https://doi.org/10.1164/rccm.201501-0059OC>
60. Wang S, Diao N, Lu C, Wu J, Gao Y, Chen J, *et al.* Evaluation of the Diagnostic Potential of IP-10 and IL-2 as Biomarkers for the Diagnosis of Active and Latent Tuberculosis in a BCG-Vaccinated Population. *PLOS ONE* 2012;**7**:e51338. <https://doi.org/10.1371/journal.pone.0051338>
61. Mihret A, Bekele Y, Bobosha K, Kidd M, Aseffa A, Howe R, *et al.* Plasma cytokines and chemokines differentiate between active disease and non-active tuberculosis infection. *Journal of Infection* 2013;**66**:357-65. <https://doi.org/https://doi.org/10.1016/j.jinf.2012.11.005>
62. Roy S, Guler R, Parihar SP, Schmeier S, Kaczowski B, Nishimura H, *et al.* Batf2/Irf1 Induces Inflammatory Responses in Classically Activated Macrophages, Lipopolysaccharides, and Mycobacterial Infection. *The Journal of Immunology* 2015;**194**:6035-44. <https://doi.org/10.4049/jimmunol.1402521>
63. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, *et al.* Consolidated Health Economic Evaluation Reporting Standards (CHEERS) Statement. *Value in Health*; **16**:e1-e5. <https://doi.org/10.1016/j.jval.2013.02.010>
64. Peter White MJ. Appendix I: Imperial College – LTBI treatment report. 2015. URL: <https://www.nice.org.uk/guidance/NG33/documents/tuberculosis-update-appendix-i2> (accessed).

- 1 65. Michelle Sleed JB, Martin Knapp, Colette McAuley, Nyree McCurry. *Estimating the unit costs*
2 *for home-start support*; 2004.
- 3 66. NICE. *Cost Effectiveness Analysis of Interferon Gamma Release Assay (IGRA) Testing for*
4 *Latent Tuberculosis*; 2010.
- 5 67. Graves AJ, Padilla MG, Hokey DA. OMIP-022: Comprehensive assessment of antigen-specific
6 human T-cell functionality and memory. *Cytometry A* 2014;**85**:576-9.
7 <https://doi.org/10.1002/cyto.a.22478>
- 8 68. Monitor, NHS, England. National Tariff Payment System 2014/15. 2014.
- 9 69. Hughes R, Wonderling D, Li B, Higgins B. The cost effectiveness of Nucleic Acid Amplification
10 Techniques for the diagnosis of tuberculosis. *Respiratory Medicine*;**106**:300-7.
11 <https://doi.org/10.1016/j.rmed.2011.10.005>
- 12 70. Drobniewski F, Cooke M, Jordan J, Casali N, Mugwagwa T, Broda A, *et al*. Systematic review,
13 meta-analysis and economic modelling of molecular diagnostic tests for antibiotic resistance in
14 tuberculosis. *Health Technol Assess* 2015;**19**:1-188, vii-viii. <https://doi.org/10.3310/hta19340>

15

Appendix 1 Prevalence data from the IDEA study

TABLE 23. Calculated prevalence data for TB populations, using the IDEA study data. Note these estimates include the ‘clinically indeterminate’ patients within the parent populations.

Target population	n	Parent Population	n	Prevalence
All TB	363	Full cohort	845	43.0%
All TB (HIV-)	338	Full cohort (HIV-)	710	47.6%
All TB (HIV+)	32	Full cohort (HIV+)	201	15.9%
Culture confirmed TB	261	Full cohort	845	30.9%
Highly probable TB	102	Full cohort	845	12.1%
Culture confirmed TB (HIV-)	248	Full cohort (HIV-)	710	34.9%
Highly probable TB (HIV-)	90	Full cohort (HIV-)	710	12.7%
Highly probable TB	102	Culture negative TB suspects (full)	584	17.5%
Highly probable TB (HIV-)	90	Culture negative TB suspects (HIV-)	462	19.5%
Highly probable TB (HIV+)	14	Culture negative TB suspects (HIV+)	183	7.7%
All TB (T-SPOT+, full)	270	T-SPOT+ (full)	352	76.7%
Dosanjh 1 (T-SPOT+,)	201	T-SPOT+ (full)	352	57.1%
Dosanjh 2 (T-SPOT+-)	69	T-SPOT+ (full)	352	19.6%
All TB (T-SPOT+ and HIV+)	17	T-SPOT+ (HIV+)	39	43.6%

Appendix 2 Distribution of disease types in the transcriptomic validation cohort

TABLE 24. Distribution of severity and disease in the TB group

N (%)	Culture confirmed TB	Highly probable TB	Total
All TB	212 (70.4)	89 (29.6)	301 (100)
Culture negative	0 (0)	76 (25.2)	76 (25.2)
Smear positive TB	48 (22.6)	2 (2.2)	50 (16.6)
Smear negative TB	139 (65.6)	59 (66.3)	198 (65.8)
Smear not tested	25 (8.3)	28 (9.3)	53 (17.6)
PTB	85 (40.1)	16 (18.0)	101 (33.6)
Extra pulmonary TB	100 (47.2)	64 (71.9)	164 (54.5)
PTB + EPTB	27 (12.7)	9 (10.1)	36 (12.0)
<i>Site of infection*</i>			
Abdomen	6 (2.8)	2 (2.2)	8 (2.7)
Bones	5 (2.4)	1 (1.1)	6 (2)
Brains	1 (0.5)	4 (4.5)	5 (1.7)
Chest	1 (0.5)	1 (1.1)	2 (0.7)
Lungs	112 (52.8)	24 (27.0)	136 (45.2)
Lymph node	87 (41.0)	44 (49.4)	131 (43.5)
Miliary TB (disseminated)	11 (5.2)	0 (0)	11 (3.7)
Pericardium	3 (1.4)	2 (2.2)	5 (1.7)
Pleura	13 (6.1)	10 (11.2)	23 (7.6)
Spine	10 (4.7)	6 (6.7)	16 (5.3)
Other	12 (5.7)	10 (11.2)	22 (7.3)

1 TABLE 25. Distribution of alternative diagnoses in the OD group

Diagnosis	N (%)
Lower respiratory track infection	79 (24.2)
Pneumonia	66 (20.2)
Non Pneumonia LRTI	13 (4.0)
Sarcoidosis	34 (10.4)
Cancer	26 (8.0)
Self resolved illness	17 (5.2)
Chest Infection	11 (3.4)
Upper respiratory track infection	10 (3.1)
Other*	169 (51.7)

2

3

4

Appendix 3 Additional information on transcriptomic signatures used for validation studies.

TABLE 26. Summary of previously identified signatures, including cohort characteristics and measures of diagnostic performance.

Study, ref	Berry et al., Nature, 2010 ¹⁷		Kaforou et al, PLoS Medicine, 2013 ¹⁹		Anderson et al., NEJM, 2014 ²⁸	Maertzdorf et al., EMBO Mol Med, 2016 ²⁹	Roe et al., JCI Insight, 2016 ³⁰	Sweeney et al., Lancet Resp 2016 ³¹
Study design	Case-control (training/test and validation)		Case-control (training/test and validation)		Children with suspected TB (training/test and validation)	Case-control (training/test and qPCR validation)	Case-control	Meta-analysis
Statistical methods	LDA		Elastic Net		Elastic Net	Random Forest	SVML	DerSimonian-Laird, Fisher's sum of logs
Location	UK / South Africa		South Africa /Malawi		South Africa /Malawi /Kenya	India /The Gambia / Uganda	UK	Multiple/meta analysis
Case type training /test cohort (n)	Training PTB + (n=54)	Test PTB + (n=33)	Training PTB + (n=194)	Test	ATB +/- (n=42/ 27)	PTB + (n=113)	ATB + (n=46)	ATB +/- (n=296)
Non-TB group training /test cohort (n)	Latent TB (n=69)/ Healthy (n=24)	Other inflammatory diseases (n=172)	Latent TB (n=167)	Other disease (n=175)	Latent TB (n=57, 14) and other disease (n=175, 64)	Healthy controls (inc. latent) (n=76)	Other Fever (n=70)	Various cohorts inc OD and latent TB
No. genes	393	86	27	44	51	4	4	3
Validated AUC*	N/A	0.83	0.98	0.95	0.89	0.98	0.95	0.84
Sensitivity/ Specificity	94.1 /96.7	90	95 /90	93 /91	82.9 /82.8	88 /75	?	
Routine Clinical Practice	X	X	X	X	Yes	X	X	X
Immuno-comp included?	X	X	HIV+/-	HIV+/-	HIV+/-	X	X	HIV+/-
EPTB included?	X	X	?	X	Yes	X	Yes	Yes
Culture –ve inc.?	X	X	X	X	Yes	X	X	Some

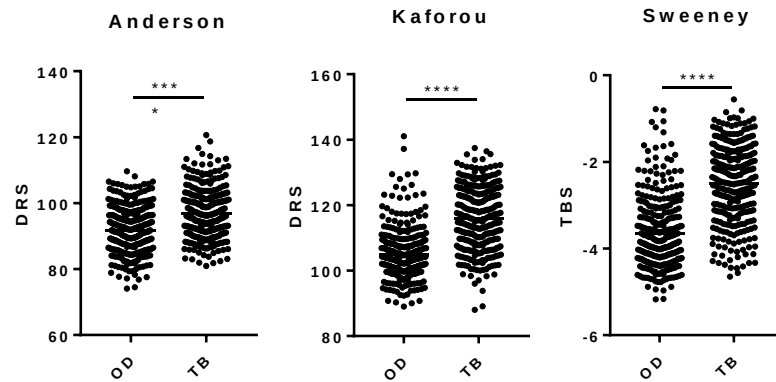
Abbreviations: LDA – Linear Discriminant Analysis; SVML – Support Vector Machine Learning; PTB - Pulmonary Tuberculosis; ATB – Active Tuberculosis; +/- = culture positive/negative; EPTB – Extrapulmonary Tuberculosis.

Appendix 4 Additional analyses for validation of transcriptomic ‘score’ signatures

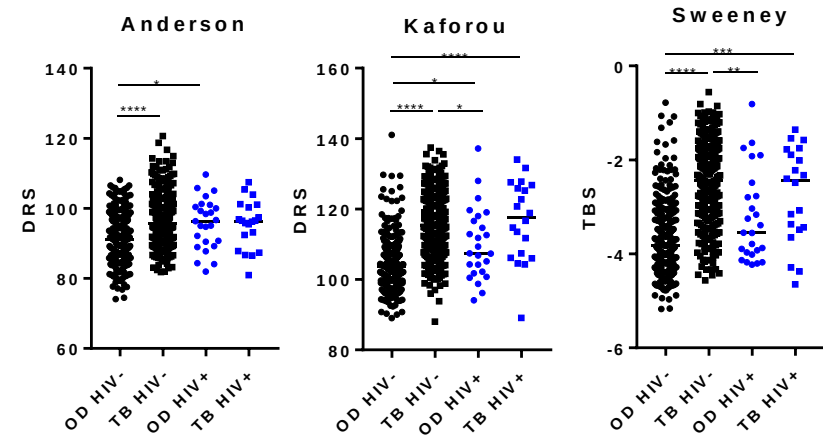
TABLE 27. Cross tabulation of previously published transcriptomic score signature results by the reference standard. The score signatures being assessed are: Anderson²⁸; Kaforou¹⁹; Sweeney³¹

	Scores	All TB (Dosanjh 1&2)	Culture-confirmed TB (Dosanjh 1)	Highly probable TB (Dosanjh 2)	OD (Dosanjh 4)
	n	301	212	89	327
Anderson	Median (range)	95.77 (80.95-120.6)	96.12 (80.95-120.6)	95.38 (81.83-118.7)	91.21 (74.06-109.7)
	Missing, n	0	0	0	0
Kaforou	Median (range)	115.5 (88.01-137.5)	117.5 (88.01-136.4)	111.7 (89.07-137.5)	103.8 (89.02-141.1)
	Missing, n	0	0	0	0
Sweeney	Median (range)	-2.431 (-4.648—0.556)	-2.181 (-4.412—0.556)	-2.813 (-4.648—1.139)	-3.817 (-5.175-0.780)
	Missing, n	0	0	0	0

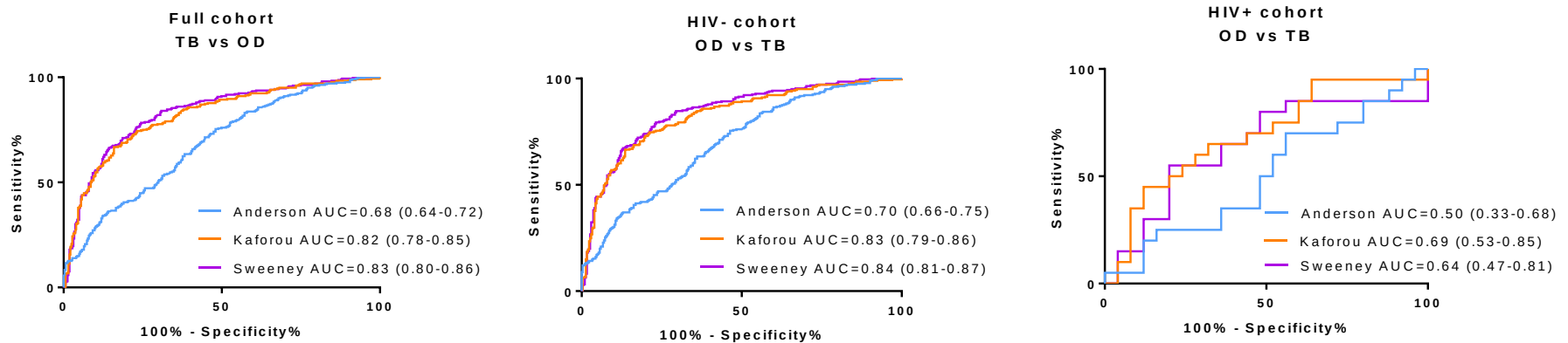
A



B



C



1

2

3

FIGURE 16. Performance of published ‘score’ based signatures in the full cohort, stratified by HIV infection status. A) Dot plot displaying the score values using each signature in the full cohort stratified by Other Disease (OD; Dosanjh 4) (n=327) group or the TB (Dosanjh 1 and 2) groups (n=301). B) Groups

1 were further stratified by HIV status. C) ROC curves displaying diagnostic performance of scores in either the full cohort, HIV- or HIV+ cohorts. The
2 score signatures being assessed are: Anderson²⁸; Kaforou¹⁹; Sweeney³¹ Statistical differences between score values between groups were determined using
3 either the Mann-Whitney (when only two groups are compared) or Kruskal-Wallis with Dunn's post hoc test (for multiple comparisons) tests for 3 or more
4 groups (* p<0.05; ** p<0.005; *** p<0.0005; p***** p<0.0001).

TABLE 28. Performance of score signatures based on Dosanjh category in the full cohort. The score signatures being assessed are: Anderson²⁸; Kaforou¹⁹; Sweeney³¹

Comparison	Signature	Patients (n)	Controls (n)	AUC
1 vs 4	Anderson	212	327	0.69 (0.65-0.74)
	Kaforou	212	327	0.85 (0.81-0.88)
	Sweeney	212	327	0.86 (0.83-0.89)
2 vs 4	Anderson	89	327	0.66 (0.60-0.72)
	Kaforou	89	327	0.74 (0.68-0.80)
	Sweeney	89	327	0.75 (0.70-0.81)
1/2 vs 4A	Anderson	301	46	0.71 (0.63-0.78)
	Kaforou	301	46	0.86 (0.82-0.91)
	Sweeney	301	46	0.88 (0.84-0.92)
1/2 vs 4B	Anderson	301	36	0.72 (0.63-0.81)
	Kaforou	301	36	0.85 (0.78-0.91)
	Sweeney	301	36	0.88 (0.82-0.93)
1/2 vs 4C	Anderson	301	175	0.65 (0.60-0.70)
	Kaforou	301	175	0.78 (0.74-0.83)
	Sweeney	301	175	0.80 (0.76-0.84)
1/2 vs 4D	Anderson	301	70	0.73 (0.66-0.79)
	Kaforou	301	70	0.85 (0.79-0.89)
	Sweeney	301	70	0.84 (0.79-0.89)

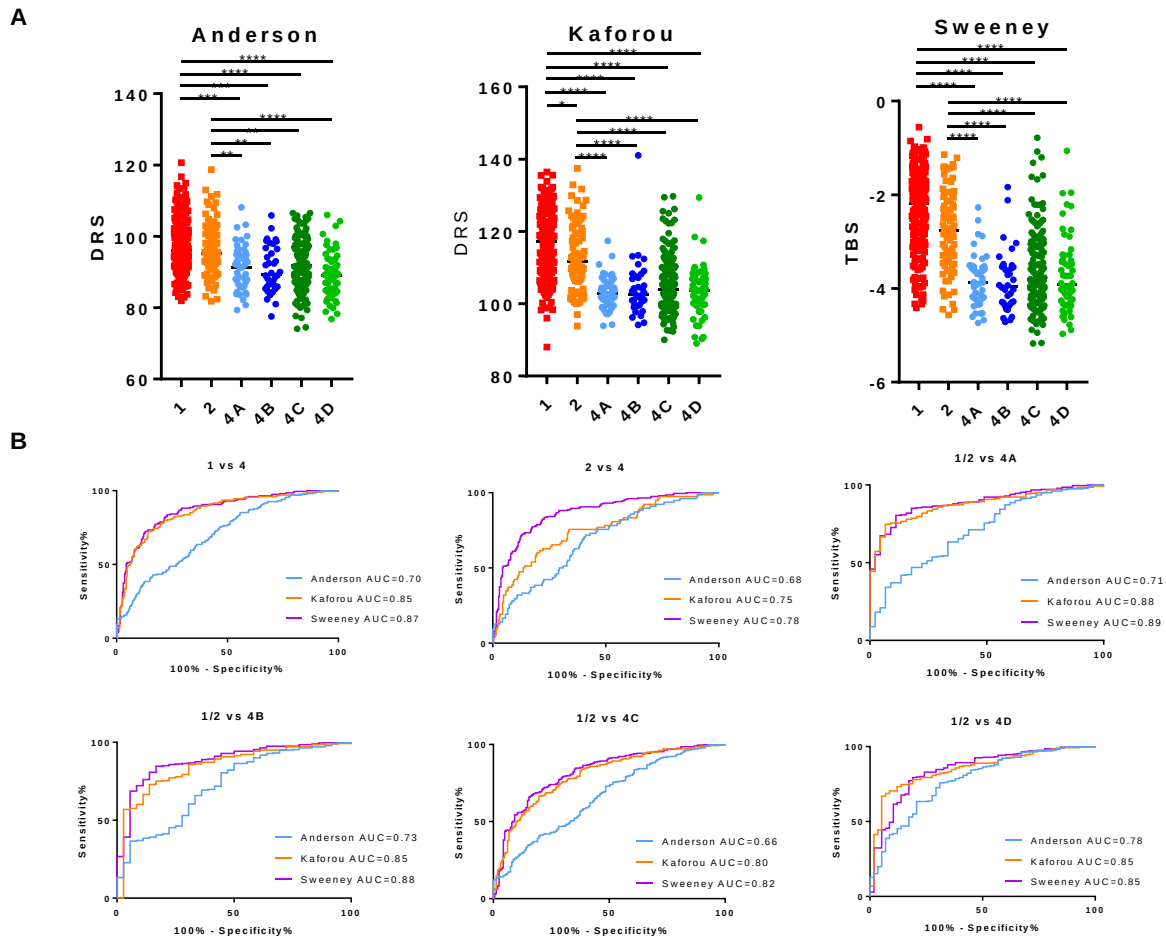


FIGURE 17. Performance of published SCORE signatures by Dosanjh category in the HIV-population. The score signatures being assessed are: Anderson²⁸; Kaforou¹⁹; Sweeney³¹. Diagnostic groups were compared using a Kruskal Wallis test with Dunn's post hoc test for multiple comparisons.

TABLE 29. Performance of score signatures based on Dosanjh category, HIV- cohort only/ The score signatures being assessed are: Anderson²⁸; Kaforou¹⁹; Sweeney³¹

Comparison	Signature	Patients (n)	Controls (n)	AUC
1 vs 4	Anderson	203	302	0.70 (0.66)
	Kaforou	203	275	0.85 (0.82)
	Sweeney	203	302	0.87 (0.83)
2 vs 4	Anderson	78	302	0.68 (0.62)
	Kaforou	78	275	0.75 (0.69)
	Sweeney	78	302	0.78 (0.72)
1/2 vs 4A	Anderson	281	45	0.71 (0.63)
	Kaforou	281	45	0.88 (0.83)

	Sweeney	281	45	0.89 (0.85)
1/2 vs 4B	Anderson	281	36	0.73 (0.64)
	Kaforou	281	36	0.85 (0.78)
	Sweeney	281	26	0.88 (0.83)
1/2 vs 4C	Anderson	281	163	0.66 (0.61)
	Kaforou	281	163	0.80 (0.75)
	Sweeney	281	163	0.82 (0.78)
1/2 vs 4D	Anderson	281	58	0.78 (0.71)
	Kaforou	281	58	0.85 (0.80)
	Sweeney	281	58	0.85 (0.79)

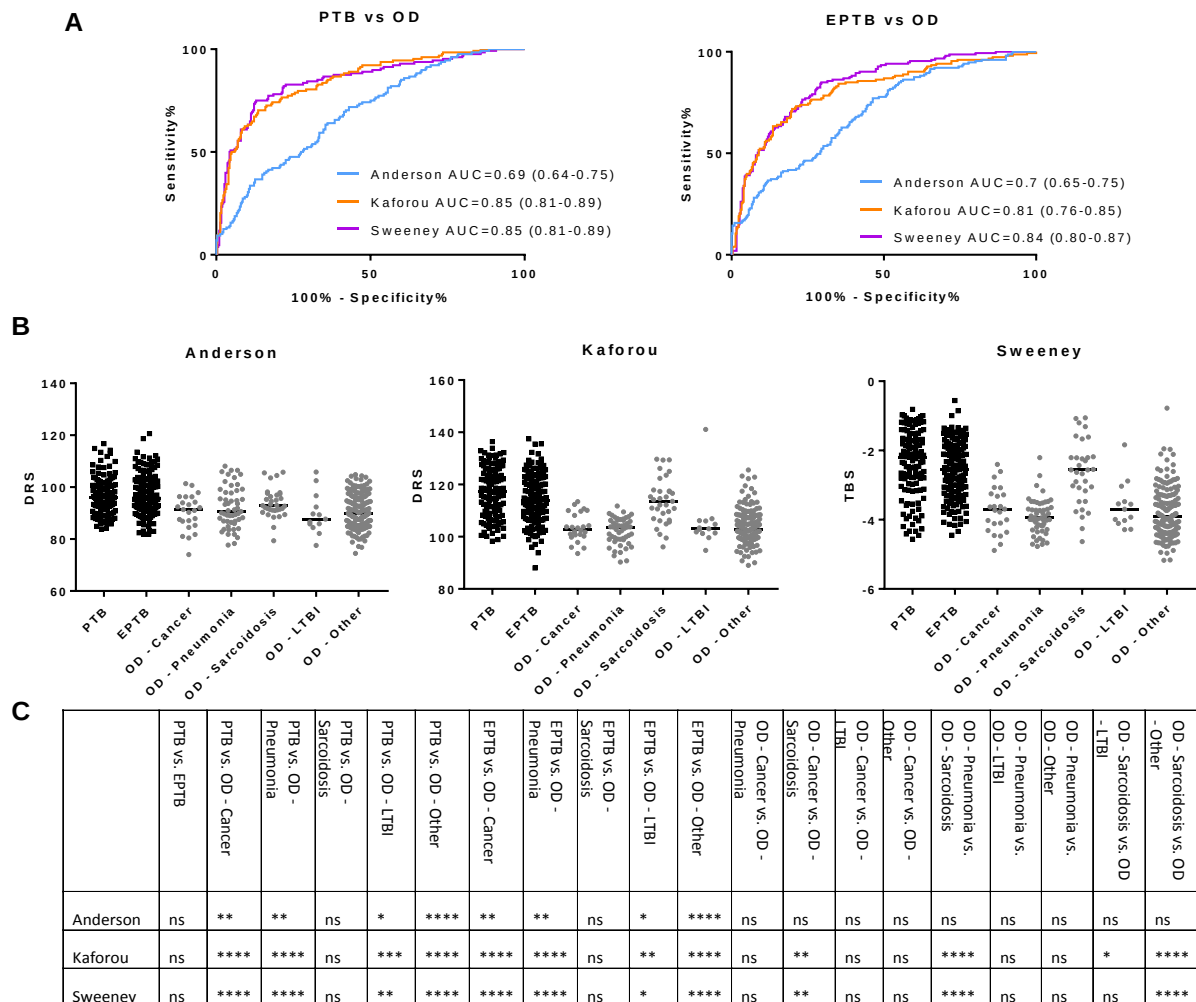


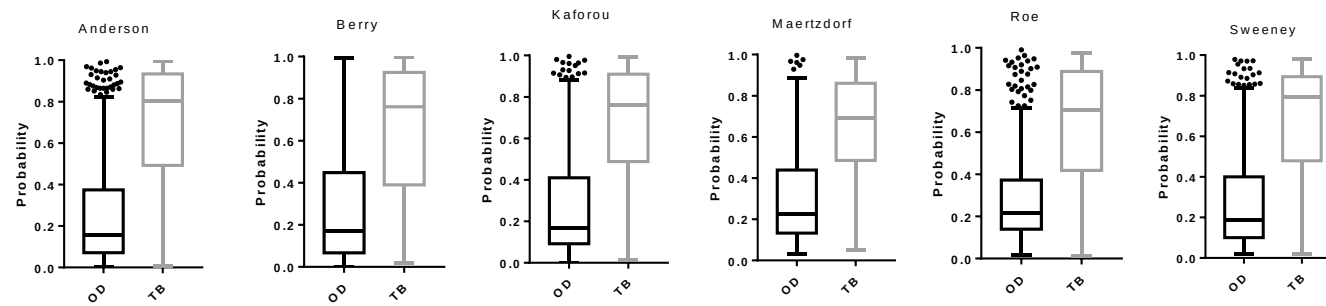
FIGURE 18. Performance of score signatures in PTB and EPTB, and in OD stratified by disease types. The diagnostic accuracy for the score signatures for detecting either PTB or EPTB was

1 performed for each signatures (A). The score values by disease sub-type within the TB and OD
2 groups is shown in the dot plots (B), scores were compared between groups using a Kruskal-
3 Wallis Test, and the results of the Dunn's multiple comparison post-hoc tests are shown in the
4 table (C). The score signatures being assessed are: Anderson²⁸; Kaforou¹⁹; Sweeney³¹. Statistical
5 differences between score values between groups were determined using the Kruskal-Wallis with
6 Dunn's post hoc test for multiple comparisons (* $p < 0.05$; ** $p < 0.005$; *** $p < 0.0005$; p****
7 $p < 0.0001$)

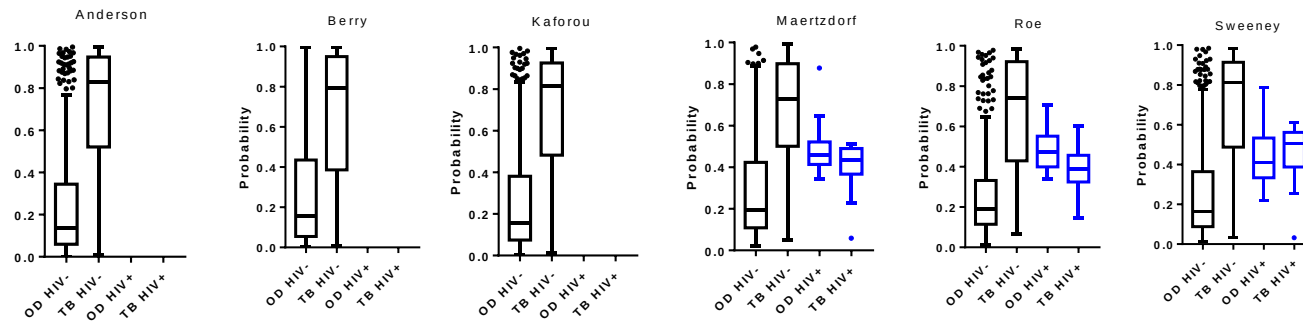
1 **Appendix 5 Additional analyses for validation of transcriptomic signatures using LDA**

2

A



B



C

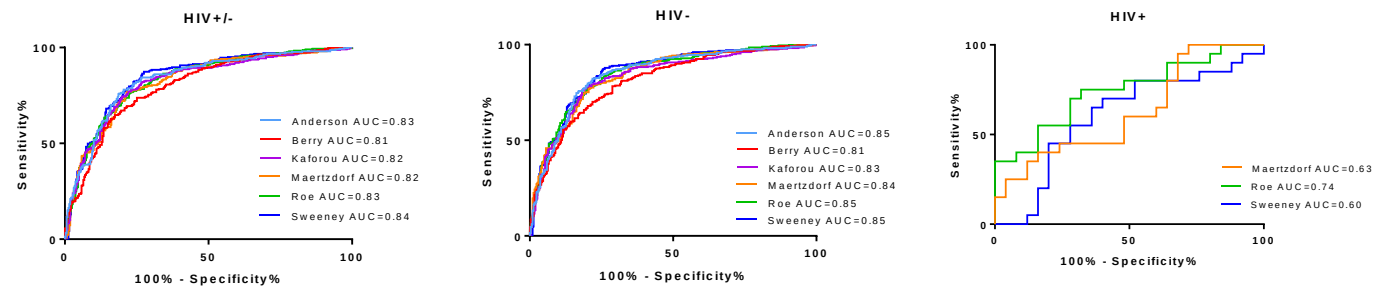


FIGURE 19. Validation of published transcriptomic signatures using LDA in full cohort (A), and stratified by HIV infection status (B). ROC curve analyses in the full cohort, HIV- and HIV+ populations are also shown (C). The signatures are referred to by their first author of the publication: Anderson²⁸; Berry¹⁷; Kaforou¹⁹; Maertsdorf²⁹; Roe³⁰; Sweeney³¹

TABLE 30. Performance of all previously published transcriptomic signatures in full cohort using LDA

LDA Signature, ref	HIV	TB (n)	OD (n)	AUC (95 CI)	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV
Anderson ²⁸	-	281	302	0.85 (0.82-0.88)	>0.09* >0.82 ^{\$} >0.40 ^Y	95.0 (91.8-97.3%) 52.0 (45.9-57.9%) 82.2 (77.2-86.5%)	39.1 (33.5-44.8%) 90.0 (86.1-93.2%) 78.2 (73.1-82.7%)	0.59 0.83 0.77	0.90 0.67 0.83
	+	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	-/+	301	327	0.83 (0.80-0.87)	>0.09* >0.82 ^{\$} >0.36 ^Y	95.0 (91.8-97.3%) 47.8 (42.1-53.7%) 84.4 (79.8-88.3%)	33.6 (28.5-39.0%) 90.2 (86.5-93.2%) 73.4 (68.3-78.1%)	0.52 0.79 0.71	0.90 0.70 0.86
Berry ¹⁷	-	281	302	0.81 (0.78-0.85)	>0.08* >0.81 ^{\$} >0.37 ^Y	95.0 (91.8-97.3%) 47.7(41.7-53.7%) 78.3 (73.0-83.0%)	35.4 (30.0-41.1%) 90.0 (86.1-93.2%) 71/2 (65.7-76.2%)	0.57 0.81 0.71	0.89 0.65 0.8
	+	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	-/+	301	327	0.80 (0.77-0.84)	>0.10* >0.83 ^{\$} >0.43 ^Y	95.0 (91.8-97.3%) 42.9 (37.2-48.6%) 73.8 (68.4-79.5%)	35.5 (30.3-40.9%) 90.2 (86.5-93.2%) 74.9 (69.9-79.5%)	0.52 0.77 0.69	0.90 0.68 0.79
Kaforou ¹⁹	-	281	302	0.83 (0.80-0.87)	>0.09* >0.80 ^{\$} >0.43 ^Y	95.0 (91.8-97.3%) 52.3 (46.3-58.3%) 78.3 (73.0-83.0%)	29.5 (24.4-35.0%) 90.0 (86.1-93.2%) 80.0 (64.8-84.2%)	0.55 0.80 0.83	0.87 0.71 0.68
	+	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	-/+	301	327	0.82 (0.79-0.86)	>0.11* >0.87 ^{\$} >0.38 ^Y	95.0 (91.8-97.3%) 48.8 (43.0-54.6%) 82.0 (77.3-86.2%)	30 (25.1-35.3%) 90.2 (86.5-93.2%) 73.0 (67.9-77.8%)	0.51 0.79 0.70	0.89 0.70 0.84
Maertsdorf ²⁹	-	281	302	0.84 (0.81-0.88)	>0.18* >0.73 ^{\$}	95.0 (91.8-97.3%) 49.8 (43.8-55.8%)	46.4 (40.6-52.2%) 90.0 (86.1-93.2%)	0.62 0.79	0.91 0.70

					>0.44 ^Y	79.0 (73.8-83.6%)	77.8 (72.7-82.3%)	0.82	0.66
	+	20	25	0.63 (0.46-0.9)	<0.50* <0.38 ^{\$} >0.52 ^Y	95.0 (75.1-99.9%) 25.0 (8.7-49.1%) 100 (83.2-100%)	32.0 (15.0-53.5%) 92.0 (74.0-99.0%) 28.0 (12.1-49.4%)	0.21 0.37 0.21	0.97 0.87 1.0
	-/+	301	327	0.82 (0.79-0.86)	>0.18* >0.72 ^{\$} >0.46 ^Y	95.0 (91.8-97.3%) 46.1 (40.8-52.3%) 77.1 (71.9-81.7%)	37.6 (32.3-43.1%) 90.2 (86.5-93.2%) 78.2 (73.4-82.3%)	0.53 0.78 0.73	0.91 0.69 0.82
Roe ³⁰	-	281	302	0.85 (0.82-0.88)	>0.16* >0.68 ^{\$} >0.31 ^Y	95.0 (91.8-97.3%) 55.5 (49.5-61.4%) 84.0 (79.2-88.10%)	40.4 (34.8-46.2%) 90.0 (86.1-93.2%) 73.5% (68.2-78.4%)	0.59 0.84 0.74	0.90 0.69 0.84

	-/+	301	327	0.83 (0.80-0.86)	>0.18* >0.70 ^{\$} >0.40 ^Y	95.0 (91.9-97.3%) 50.5 (44.7-56.3%) 76.7 (71.6-81.4%)	39.1 (33.8-44.7%) 90.2 (86.5-93.2%) 77.0 (72.1-81.5%)	0.54 0.80 0.72	0.91 0.71 0.81
Sweeney ³¹	-	302	281	0.85 (0.82-0.89)	>0.15* >0.78 ^{\$} >0.32 ^Y	95.0 (91.9-97.3%) 52.7 (46.7-58.6%) 87.9 (83.5-91.5%)	47.0 (41.3-52.8%) 90.0 (86.1-93.2%) 73.8 (68.5-78.7%)	0.62 0.83 0.75	0.91 0.68 0.87
	+	20	25	0.60 (0.43-0.78)	>0.26* >0.71 ^{\$} >0.43 ^Y	95.0 (75.1-99.9%) 0.00 (0.0-16.8%) 70.0 (45.7-88.1%)	8.0 (0.99-26.0%) 92.0 (74.0-99.0%) 60.0 (38.7-78.9%)	0.16 0.00 0.25	0.89 0.83 0.92
	-/+	301	327	0.84 (0.81-0.87)	>0.15* >0.78 ^{\$} >0.33 ^Y	95.0 (91.9-97.2%) 50.8 (40.0-56.6%) 87.4 (83.1-90.9%)	43.1 (37.7-48.7%) 90.2 (86.5-93.2%) 72.5 (67.3-77.3%)	0.56 0.80 0.71	0.92 0.71 0.89
	+	20	25	0.74 (0.60-0.89)	<0.57* <0.40 ^{\$} >0.45 ^Y	95.0 (75.1-99.9%) 40.0 (19.1-64.0%) 75.0 (50.9-91.3%)	20.0 (6.8-40.7%) 92.0 (74.0-99.0%) 68.0 (46.5-85.1%)	0.18 0.49 0.31	0.95 0.89 0.94

1

2

TABLE 31. Performance of all transcriptomic signatures in detecting either culture-confirmed TB (Dosanjh 1) or culture-unconfirmed TB (Dosanjh 2) using LDA according to Dosanjh criteria. The area under the curve (AUC) and 95% confidence intervals (95% CI) are reported.

Comparison	Signature	Patients (n)	Controls (n)	AUC (95% CI)
Dosanjh 1 vs 4	Anderson ²⁸	203	302	0.86 (0.83-0.90)
	Berry ¹⁷	203	302	0.84 (0.81-0.88)
	Kaforou ¹⁹	203	302	0.86 (0.82-0.89)
	Maertsdorf ²⁹	203	302	0.86 (0.83-0.89)
	Roe ³⁰	203	302	0.87 (0.84-0.90)
	Sweeney ³¹	203	302	0.87 (0.84-0.91)
Dosanjh 2 vs 4	Anderson ²⁸	78	302	0.78 (0.73-0.84)
	Berry ¹⁷	78	302	0.68 (0.61-0.75)
	Kaforou ¹⁹	78	302	0.73 (0.66-0.80)
	Maertsdorf ²⁹	78	302	0.79 (0.74-0.85)
	Roe ³⁰	78	302	0.79 (0.74-0.85)
	Sweeney ³¹	78	302	0.80 (0.74-0.85)

TABLE 32. Diagnostic performance of previously published signatures for detection of PTB or EPTB, by LDA

Comparison	Signature	Patients (n)	Controls (n)	AUC
PTB vs OD	Anderson ²⁸	128	302	0.84 (0.80-0.89)
	Berry ¹⁷	128	302	0.83 (0.79-0.87)
	Kaforou ¹⁹	128	302	0.84 (0.80-0.89)
	Maertsdorf ²⁹	128	302	0.85 (0.81-0.89)
	Roe ³⁰	128	302	0.86 (0.82-0.90)
	Sweeney ³¹	128	302	0.86 (0.82-0.90)
EPTB vs OD	Anderson ²⁸	153	302	0.84 (0.80-0.87)
	Berry ¹⁷	153	302	0.78 (0.73-0.82)
	Kaforou ¹⁹	153	302	0.82 (0.78-0.86)
	Maertsdorf ²⁹	153	302	0.83 (0.79-0.87)
	Roe ³⁰	153	302	0.84 (0.80-0.87)
	Sweeney ³¹	153	302	0.85 (0.81-0.89)

Appendix 6 Supporting data for proteomic signature validation

TABLE 33. Details of proteins within the SELDI-TOF Signature discovered by Hamilton and Levin et al.

Surface	MW (Da)	Protein ID
Q10 pH 9.5 low	4357 and 4468	? Patent pending
CM10 pH 6.0 low	8613	? Patent pending
IMAC	5099/5092	? Patent pending
CM10 pH 6.0	13,762 and 13,801	? Patent pending

Analysis for bridging study

TABLE 34. Correlation matrix showing r and p-value (Pearson's) for SELDI-TOF normalised intensity and LC-MS normalised ratio for protein 1-3 in the Levin Signature

	P1 (LC-MS)	P2 (LC-MS)	P3 (LC-MS)
P1 (SELDI-TOF)	0.17, 0.027	-0.215, 0.00060	-0.14, 0.12
P2 (SELDI-TOF)	0.15, 0.085	-0.15, 0.16	0.02, 0.68
P3 (SELDI-TOF)	0.12, 0.30	-0.13, 0.83	0.07, 0.17

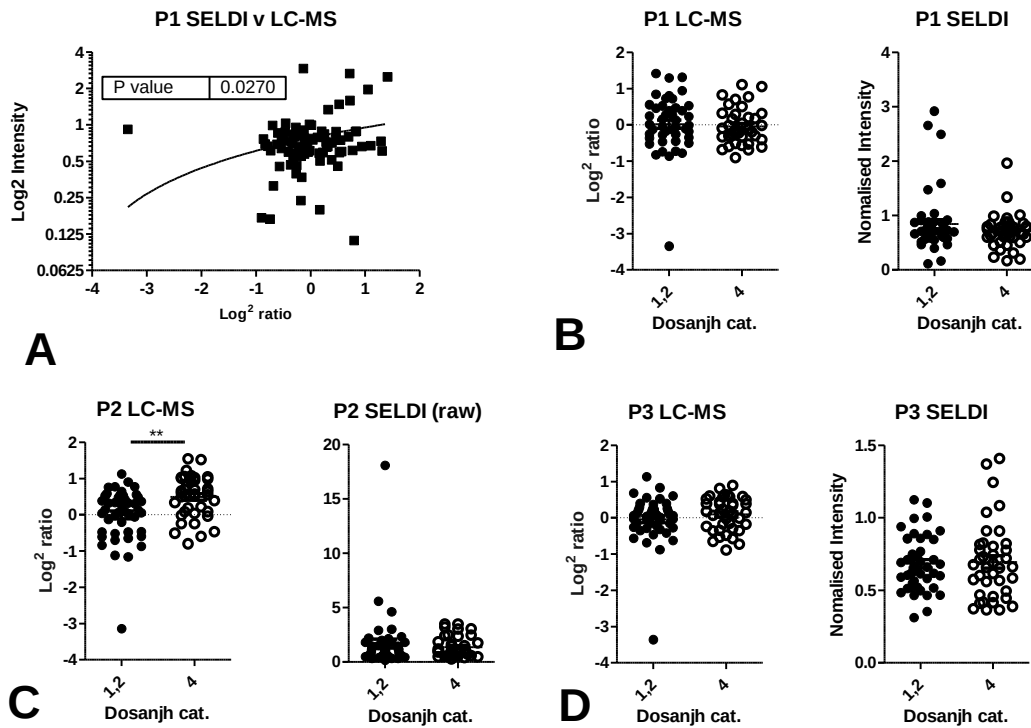


FIGURE 20. Comparison of three proteins from Levin signature with LC-MS data. A. Correlation of P1 SELDI-TOF intensity for P1 with LC-MS ratio. C-D. Dot plots for Protein 1-3 in LC-MS and SELDI-TOF data; abundance was compared between groups using a Mann-Whitney U test; ** $p < 0.005$.

1 Abundance of signature proteins

2 TABLE 35. Abundance of signature proteins in full cohort, and comparisons between key patient groups. (log2 ratio of total intensity, normalised to internal control). P
3 values showing significant differences between protein abundance were calculated using the ANOVA and Welch test and were adjusted for the effect of multiple
4 hypothesis testing using the FDR (<0.1) method. The signatures from which the proteins are derived are: Song⁴⁴; Xu⁴⁵; Levin (unpublished); Achkar²⁰.

	TB (Dosanjh 1 and 2) Vs OD (Dosanjh4)			Culture confirmed TB (Dosanjh1) Vs OD (Dosanjh 4)			Highly probable TB (Dosanjh2) Vs OD (Dosanjh 4)		
	ATB	OD	p.overall	ATB	OD	p.overall	ATB	OD	p.overall
Signature (Protein ID)	N=82	N=84		N=60	N=84		N=22	N=84	
Song. (P01009)	-0.10 [-0.98;4.63]	-0.27 [-2.08;4.36]	0.007	-0.07 [-0.98;4.63]	-0.27 [-2.08;4.36]	0.001	N=22	N=84	
Xu. (P14780_MMP9)	-0.19 [-1.29;4.10]	-0.20 [-2.97;4.30]	0.612	-0.16 [-1.29;4.10]	-0.20 [-2.97;4.30]	0.335	-0.24 [-0.58;0.20]	-0.27 [-2.08;4.36]	0.668
Xu. (P06702_S100A9)	-0.21 [-0.96;4.21]	-0.30 [-2.40;3.85]	0.036	-0.16 [-0.96;4.21]	-0.30 [-2.40;3.85]	0.008	-0.20 [-0.56;0.00]	-0.20 [-2.97;4.30]	0.503
Xu. (P08294_SOD)	-0.28 [-1.53;4.26]	-0.17 [-0.72;4.56]	<0.001	-0.28 [-1.53;4.26]	-0.17 [-0.72;4.56]	<0.001	-0.27 [-0.58;0.00]	-0.30 [-2.40;3.85]	0.858
Levin. (Lev 1)	-0.23 [-1.06;4.36]	-0.11 [-2.23;4.56]	<0.001	-0.23 [-1.06;4.36]	-0.11 [-2.23;4.56]	<0.001	-0.30 [-0.55;0.49]	-0.17 [-0.72;4.56]	0.019
Levin (Lev 2)	-0.20 [-0.94;4.42]	-0.19 [-2.21;4.44]	0.233	-0.20 [-0.94;4.42]	-0.19 [-2.21;4.44]	0.795	-0.20 [-0.47;-0.05]	-0.11 [-2.23;4.56]	0.015
Levin. (Lev 3)	-0.21 [-0.97;4.51]	-0.23 [-1.89;4.48]	0.679	-0.17 [-0.97;4.51]	-0.23 [-1.89;4.48]	0.206	-0.23 [-0.44;-0.09]	-0.19 [-2.21;4.44]	0.018
Levin (Lev 4)	-0.21 [-1.30;4.22]	-0.21 [-2.80;4.63]	0.475	-0.20 [-1.30;4.22]	-0.21 [-2.80;4.63]	0.59	-0.26 [-0.46;-0.02]	-0.23 [-1.89;4.48]	0.152
Achkar (O00391)	-0.20 [-1.29;4.31]	-0.17 [-2.89;4.39]	0.086	-0.19 [-1.29;4.31]	-0.17 [-2.89;4.39]	0.12	-0.23 [-0.91;0.34]	-0.21 [-2.80;4.63]	0.493
Achkar (P02654)	-0.23 [-1.08;4.32]	-0.10 [-2.49;4.68]	<0.001	-0.25 [-1.08;4.32]	-0.10 [-2.49;4.68]	<0.001	-0.21 [-0.43;0.00]	-0.17 [-2.89;4.39]	0.252
Achkar. (P07359)	-0.19 [-1.22;4.34]	-0.18 [-2.48;4.47]	0.361	-0.19 [-1.22;4.34]	-0.18 [-2.48;4.47]	0.404	-0.19 [-0.44;0.01]	-0.10 [-2.49;4.68]	0.04
Achkar (P08571)	-0.19 [-1.31;4.49]	-0.19 [-2.73;4.47]	0.804	-0.19 [-1.31;4.49]	-0.19 [-2.73;4.47]	0.644	-0.18 [-0.45;-0.04]	-0.18 [-2.48;4.47]	0.549
Achkar (P12955)	-0.22 [-1.33;3.74]	-0.18 [-2.63;3.99]	<0.001	-0.23 [-1.33;3.74]	-0.18 [-2.63;3.99]	<0.001	-0.20 [-0.50;0.01]	-0.19 [-2.73;4.47]	0.773
Achkar (P14151)	-0.18 [-1.14;4.39]	-0.19 [-2.49;4.53]	0.608	-0.17 [-1.14;4.39]	-0.19 [-2.49;4.53]	0.364	-0.19 [-0.42;0.02]	-0.19 [-2.49;4.53]	0.717
Achkar (P22105)	-0.21 [-1.23;4.33]	-0.15 [-2.56;4.52]	0.023	-0.22 [-1.23;4.33]	-0.15 [-2.56;4.52]	0.009	-0.19 [-0.44;0.01]	-0.15 [-2.56;4.52]	0.694

Achkar (P49747)	-0.24 [-1.17;4.31]	-0.12 [-2.56;4.59]	<0.001	-0.24 [-1.17;4.31]	-0.13 [-2.56;4.59]	0.001	-0.25 [-0.45;0.03]	-0.13 [-2.56;4.59]	0.062
Achkar (P49908)	-0.23 [-1.17;4.40]	-0.14 [-2.54;4.52]	0.001	-0.23 [-1.17;4.40]	-0.15 [-2.54;4.52]	0.003	-0.22 [-0.45;-0.03]	-0.15 [-2.54;4.52]	0.046
Achkar (P51884)	-0.24 [-1.07;4.39]	-0.12 [-2.16;4.53]	<0.001	-0.26 [-1.07;4.39]	-0.12 [-2.16;4.53]	<0.001	-0.19 [-0.49;0.01]	-0.12 [-2.16;4.53]	0.084

1

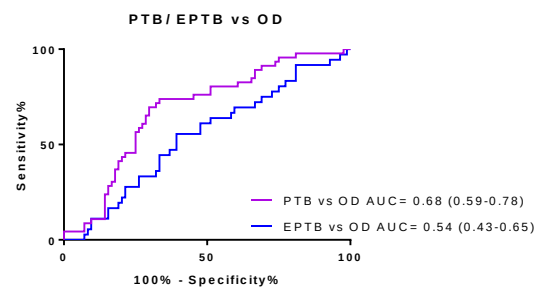
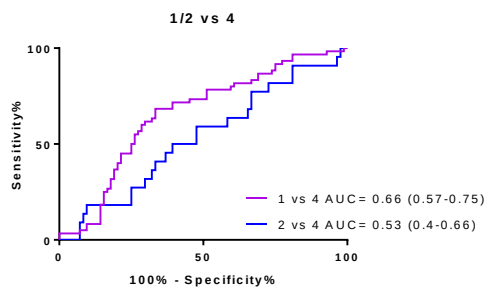
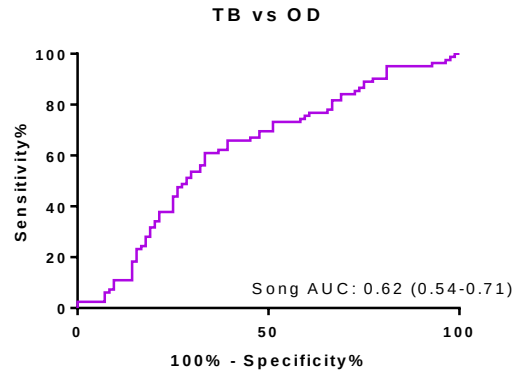
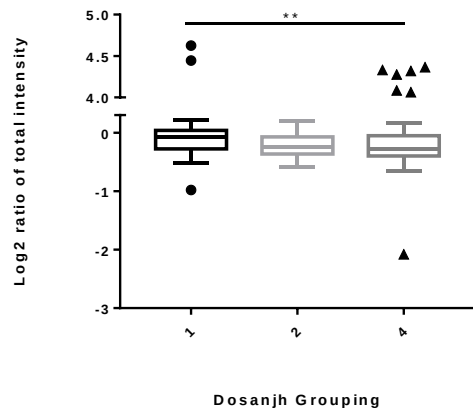


FIGURE 21. Abundance of SERPINA1 (alpha-1-antitrypsin or P01009). Medians were compared between groups using the Kruskal-Wallis test (** $p < 0.005$).

Missing and imputed data

Spreadsheets detailing the missing data for each signature are available on request.

TABLE 36. Performance of LC-MS technology in detecting proteins from key previously published signatures, including the proportion of missing data in the cohorts. The coverage of proteins using either the depleted dataset (n=168) or crude dataset (n=267) is presented.

Study	Published Signature	Depleted LC-MS MissingValues (%)	Crude_Dundee LC-MS MissingValues (%)
Song et al ⁴⁴	P01009	0	0
Xu et al ⁴⁵	P06702	21.43	61.65
Xu et al	P08294	0.6	Not Measured
Xu et al	P14780	10.71	Not Measured
Levin et al	Lev1	0	0
Levin et al	Lev2	0	0
Levin et al	Lev3	0	11.11
Levin et al	Lev4	0	65.59
Achkar et al ²⁰	O00391	0	Not Measured
Achkar et al	P02654	0	0
Achkar et al	P07359	0	Not Measured
Achkar et al	P08571	0	96.77
Achkar et al	P12955	21.43	Not Measured
Achkar et al	P14151	0	Not Measured
Achkar et al	P22105	0	Not Measured
Achkar et al	P49747	0	Not Measured
Achkar et al	P49908	0	Not Measured
Achkar et al	P51884	0	0

1 TABLE 37. Diagnostic performance of Song⁴⁴ signature in the validation of TB in the full LC-MS
2 validation cohort (n=168) and in key sub-groups of TB.

	TB vs OD	1 vs 4	2 vs 4	PTB v OD	EPTB vs OD
Patients (n)	82	60	22	46	36
Control (n)	84	84	84	84	84
AUC (95% CI)	0.62 (0.54-0.71)	0.66 (0.57-0.75)	0.53 (0.4-0.66)	0.68 (0.59-0.78)	0.54 (0.43-0.65)
Cut-off	-0.44 [*]	-0.42 [*]	-0.53 [*]	-0.4 [*]	-0.53 [*]
	-0.11 [§]	-0.11 [§]	-0.11 [§]	-0.11 [§]	-0.11 [§]
Sensitivity (95% CI)	95 (88%-99%) [*]	95 (86%-99%) [*]	95 (77%-100%) [*]	96 (85%-99%) [*]	97 (85%-100%) [*]
	11 (5.1%-20%) [§]	8.3 (2.8%-18%) [§]	18 (5.2%-40%) [§]	11 (3.6%-24%) [§]	11 (3.1%-26%) [§]
Specificity (95% CI)	19 (11%-29%) [*]	19 (11%-29%) [*]	3.6 (0.74%-10%) [*]	25 (16%-36%) [*]	3.6 (0.74%-10%) [*]
	90 (82%-96%) [§]	90 (82%-96%) [§]	90 (82%-96%) [§]	90 (82%-96%) [§]	90 (82%-96%) [§]

3
4
5
6

Appendix 7 Additional analyses for diagnostic evaluation of serum protein signatures for detection of active TB using LDA

TABLE 38. Diagnostic performance of serum proteomic signatures in full cohort n=166

Signature, ref	TB (n)	OD/ LTBI (n)	AUC (95 CI)	Cut-off	Sensitivity (95 CI)	Specificity (95 CI)
Xu ⁴⁵	82	84	0.66 (0.58-0.75)	>0.29* >0.65 ^{\$}	95.1 (88-98.7) 8.5 (3.5-16.8)	14.3 (7.6-23.6) 90.5 (82.1-95.8)
Levin (unpublished)	82	84	0.64 (0.55-0.72)	>0.34* >0.63 ^{\$}	95 (88-98.6) 25.6 (16.6-36.4)	14.1 (7.5-23.3) 90.5 (82.1-95.8)
Achkar ²⁰	82	84	0.68 (0.60-0.76)	>0.18* >0.77 ^{\$}	95.1 (88-98.7) 20.7 (12.6-31.1)	11.9 (5.9-20.8) 90.5 (82.1-95.8)

TABLE 39. Diagnostic performance of Xu⁴⁵, Levin (unpublished) and Achkar²⁰ serum proteomic signatures in key patient sub-groups in full cohort, using LDA. We also performed LDA for the Song signature using this approach, but it should be noted that since this protein only comprises of one protein, this statistical approach is not recommended.

Comparison	Signature	Patients (n)	Controls (n)	AUC
1 vs 4	Xu	60	84	0.70 (0.61-0.79)
	Levin	60	84	0.67 (0.58-0.76)
	Achkar	60	84	0.70 (0.62-0.79)
2 vs 4	Xu	22	84	0.59 (0.46-0.72)
	Levin	22	84	0.57 (0.45-0.69)
	Achkar	22	84	0.55 (0.41-0.69)
EPTB vs OD	Xu	36	84	0.61 (0.50-0.71)
	Levin	36	84	0.51 (0.40-0.62)
	Achkar	36	84	0.60 (0.49-0.70)
PTB vs OD	Xu	46	84	0.67 (0.57-0.78)
	Levin	46	84	0.69 (0.59-0.78)
	Achkar	46	84	0.70 (0.60-0.79)

Appendix 8 Validation of published proteomic signatures using LC-MS by LDA, with the removal of Sarcoidosis

TABLE 40. Performance of Xu⁴⁵, Levin (unpublished) and Achkar²⁰ signatures for detecting all TB, using LDA, with patients with Sarcoidosis removed.

Signature, ref	TB (n)	OD (n)	AUC (95 CI)	Cut-off	Sensitivity (95 CI)	Specificity (95 CI)
Xu	82	79	0.68 (0.60-0.76)	>0.30 >0.65	95.1 (88-98.7) 11 (5.1-19.8)	13.9 (7.16-23.6) 91.1 (82.6-96.4)
Levin	82	79	0.65 (0.56-0.73)	>0.35 >0.69	95.1 (88-98.7) 19.5 (11.6-29.7)	19 (11-29.4) 91.1 (82.6-96.4)
Achkar	82	79	0.69 (0.60-0.76)	>0.21 >0.76	95.1 (88-98.7) 24.4 (15.6-35.1)	13.9 (7.2-23.6) 91.1 (82.6-96.4)

TABLE 41. Performance of Xu⁴⁵, Levin (unpublished) and Achkar²⁰ signatures for detecting key subgroups of TB, using LDA, with Sarcoidosis removed.

Comparison	Signature	TB (n)	OD (n)	AUC
1 vs 4	Xu	60	79	0.72 (0.63-0.81)
	Levin	60	79	0.68 (0.59-0.77)
	Achkar	60	79	0.71 (0.63-0.80)
2 vs 4	Xu	22	79	0.60 (0.47-0.73)
	Levin	22	79	0.55 (0.43-0.67)
	Achkar	22	79	0.53 (0.39-0.67)
EPTB vs OD	Xu	36	79	0.63 (0.53-0.74)
	Levin	36	79	0.53 (0.42-0.64)
	Achkar	36	79	0.60 (0.49-0.70)
PTB vs OD	Xu	46	79	0.69 (0.59-0.79)
	Levin	46	79	0.70 (0.61-0.80)
	Achkar	46	79	0.70 (0.61-0.80)

Appendix 9 Additional information for flow cytometry

introduction and methods

TABLE 42. Summary of cellular immune signatures to be validated, with reference to the original publication and the study design and characteristics of the discovery cohorts. Abbreviations: C+/- = culture positive/negative; PTB = pulmonary TB; EPTB = extrapulmonary TB. Where two cohorts (discovery and validation) were described, the n and test accuracy was given for the validation cohort.

Publication	1 - Harari et al (2011) Nat Med ²³	2 - Pollock et al (2013) JID ²⁴	3 - Rozot et al (2014) CID ⁵⁰	4- Portevin et al (2014) Lancet ²⁶	5- Petruccioli et al (2015) ²⁷	6 - Adekambi et al (2016) JCI ²⁵
Signature	% of TNF α -only + cells	%CD45RA-CCR7-CD127- (T _{EFF})	SCORE (combination of %TNF α -only and presence of IFN γ response)	CD27 MFI Ratio	%CD45RA-CD27- or %CD45RA-CD27+	%HLA-DR+
Parent population	CD4+ Cytokine+ (IFN γ /IL-2/TNF α)	CD4+ TNF α -only	CD4+ Cytokine+ and CD8+	CD4+ IFN γ +	CD4+ IFN γ +	CD4+ IFN γ +
TB type (n) *	C+ PTB/EPTB (16)	C+ PTB (13)	C+/- PTB/EPTB (53)	Children C+/- PTB (6)	C+ PTB (13)	C+ PTB (26)
LTBI type (n) *	Healthy (77)	Healthy (HIV+/-) (21)	Healthy (141)	Healthy children (30)	Healthy (30)	Healthy (25)
HIV included?	No	Yes	No	No	No	No (in validation)
AUC *	0.99	0.99	0.89	0.93/0.88	0.85 and 0.91	1.0
Sensitivity *	100%	100%	81.1%	83.3%	69.2 / 76.9%	100 %
Specificity *	96%	92.9%	86.5%	96.8 / 86.7%	87 / 90%	96 %
Internal validation	Yes (AUC 0.82) and inc. C+/- and EPTB	No	No	Yes	No	Yes
External validation	Wilkinson et al 2016 (didn't work)	Halliday et al 2015 (AUC vs remote 0.99)	No	Petruccioli (AUC)	-	Wilkinson et al 2016 HIV+/- (AUC = 0.98) Riou et al 2017 HIV+/- (AUC 0.92/0.99)

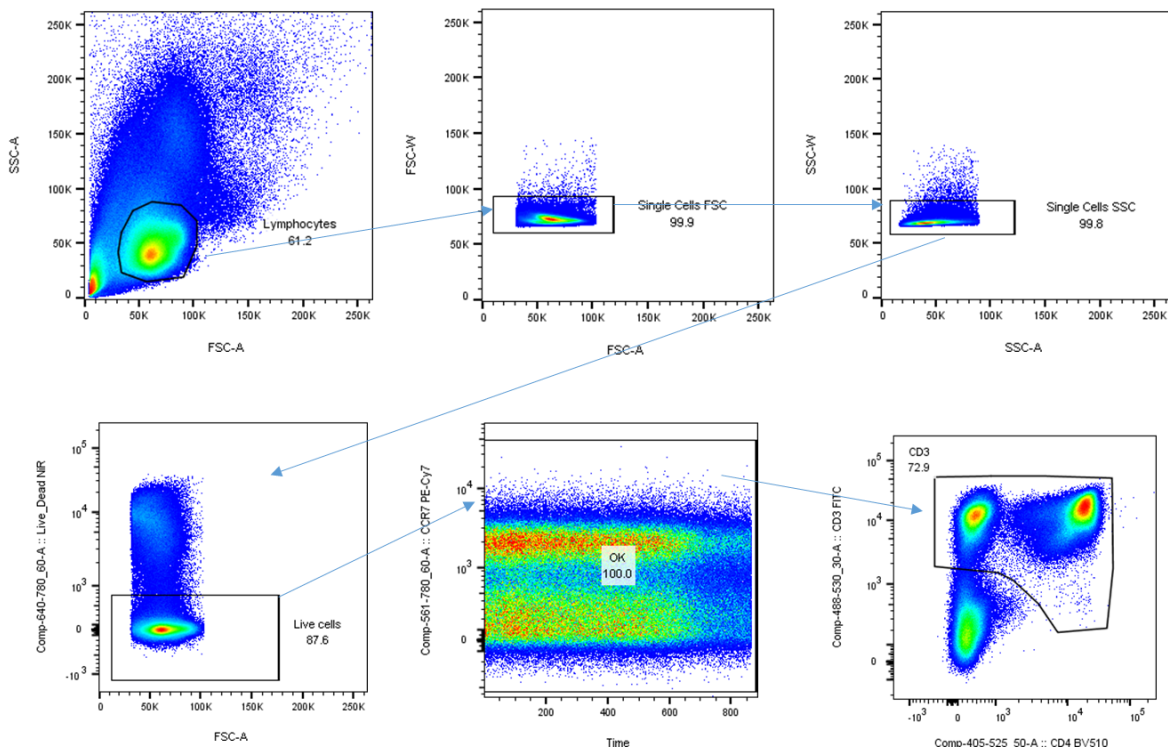
TABLE 43. Details of antibodies used and their final dilutions during staining.

Target	Surface/ IC	Species, isotype, clone	Fluorochrome	Bandpass Filter Fortessa A	Dilution	Company supplied
CD45RA	Surface	Mouse IgG ₁ , κ , HI30 RUO	BUV395	355-450/50	1 in 100	BD
CD3	Surface	Mouse IgG _{2a} , κ , HIT3a	FITC	488-530/30	1 in 30	Biolegend
CD4	Surface	Mouse IgG2b, κ ,OKT4	BV510	405-425/50	1 in 30	Biolegend
CD8	Surface	Mouse (BALB/c) IgG1, κ , SK1	BV605	405-560/520	1 in 30	Biolegend
CCR7	Surface	Rat IgG2a, κ , 3D12	PE-Cy7	561-780/60	1 in 15	BD
CD127	Surface	Mouse IgG ₁ , κ , HIL-7R-M21	PE	561-582/15	1 in 40	BD
CD27	Surface	Mouse IgG ₁ , κ , O323	BV785	405-655/8	1 in 40	Biolegend
HLA-DR	Surface	Mouse IgG2a, κ ,L243	PerCP-Cy5.5	488-710/50	1 in 40	Biolegend
IFN- γ	IC	Mouse IgG1, κ , 4S.B3	BV421	405-450/50	1 in 30	Biolegend
TNF- α	IC	Mouse IgG ₁ κ , MAb11	APC	640-670/14	1 in 50	Biolegend
IL-2	IC	Rat IgG2a, κ , MQ1-17H12	AF700	640-720/40	1 in 50	Biolegend

1

2

3



4

FIGURE 22. Initial basic gating strategies for the cellular immunological signatures as set on FlowJo version 10. Gates were set on the lymphocyte populations, live cells and on single cells and further gated on the CD3+ population.

The BV421 fluorochrome on the IFN- γ -specific antibody invariably showed a false positive signal in unstimulated samples. This increased background manifested as an aggregate population on the CD3+ population, an observation that has been noted for the BV421 fluorochrome and specific IFN γ antibodies elsewhere⁶⁷. The false positive could easily be differentiated from a true positive signal based on the fluorescence in other channels: the 'streak' was not positive for AF700 or APC, and showed an increasing fluorescence in the BV510 channel. The BV421 aggregate was therefore removed from each file before the final cytokine gates were applied, by selecting the 'streak' of artefact staining on the IFN- γ + TNF- α + plot within the CD3+ population.

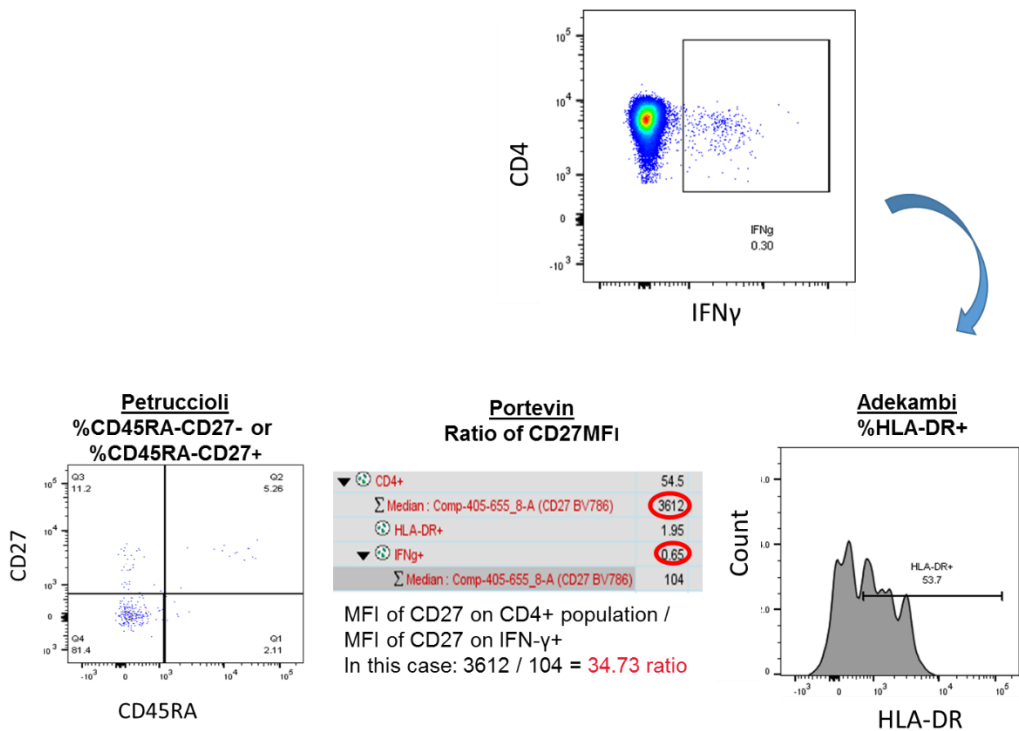


FIGURE 23. Gating strategies for the validating signatures exploring IFN γ + cytokine responses within the CD4+ population. Petruccioli²⁷ et al measures memory cells lost in differentiation, Portevin²⁶ et al incorporates an MFI ratio for measuring differentiation, and Adekambi²⁵ et al paper measures the %HLA-DR activation marker.

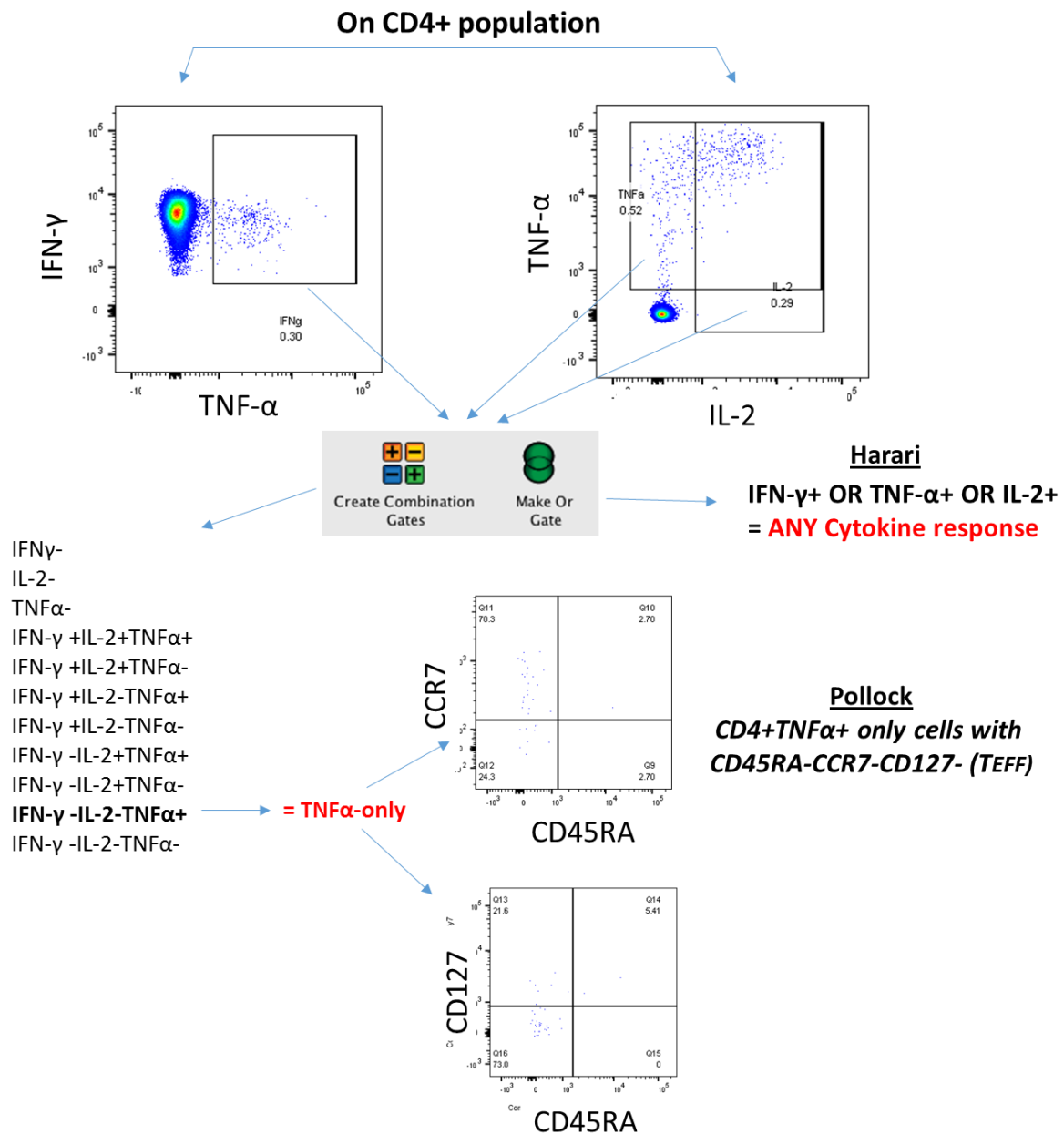


FIGURE 24. Gating strategy for validating Pollock²⁴ et al and Harari²³ et al signatures. Pollock et al investigates the effector phenotype within the CD4+ TNFα+only population (IFN-γ- IL-2- TNFα+) and Harari et al investigates %TNF-α+ within any cytokine producers in the CD4+ population.

Appendix 10 Additional cohort characteristics of the validation of cellular immune signatures sub-study

TABLE 44. Distribution of severity and disease in TB group of cellular immune signature validation cohort. For site of infection, some individuals will have TB in more than one of these sites, therefore the numbers do not equal the total amount in parent columns. ^{\$}In this table, those diagnosed with TB pulmonary and extrapulmonary involvement were classified as 'PTB + EPTB'; however, in the main analysis, the PTB + EPTB existed as a subset of the overall PTB group.

Type of TB		Culture confirmed TB (Dosanjh 1)	Highly Probably TB (Dosanjh 2)	All TB (Total)
All TB		60	32	92
Culture :	Positive	60 (100%)	0 (0%)	60 (65.2%)
	Negative	0 (0%)	27 (84.4%)	27 (29.4%)
	Not tested/failed	0 (0%)	5 (15.6%)	5 (5.4%)
Smear:	Positive	18 (30%)	2 (6.3%)	20 (21.8%)
	Negative	34 (56.7%)	25 (78.1%)	59 (64.1%)
	Not tested/failed	8 (13.3%)	5 (15.6%)	13 (14.1%)
PTB		30 (50%)	4 (12.5%)	34 (37%)
EPTB		19 (31.7%)	27 (84.4%)	46 (50%)
PTB + EPTB ^{\$}		11 (18.3%)	1 (3.1%)	12 (13%)
Site of infection*	Abdomen	1 (1.7%)	1 (3.1%)	2 (2.2%)
	Brains	1 (1.7%)	2 (6.2%)	3 (3.3%)
	Chest	1 (1.7%)	1 (3.1%)	2 (2.2%)
	Lungs	41 (68.3%)	5 (15.6%)	46 (50%)
	Lymph node	22 (36.7%)	16 (50%)	38 (41.3%)
	Miliary TB	3 (5%)	0 (0%)	3 (3.3%)
	Pericardium	2 (3.3%)	1 (3.1%)	3 (3.3%)
	Pleura	3 (5%)	2 (6.2%)	5 (5.4%)
	Spine	2 (3.3%)	2 (6.2%)	4 (4.3%)
	Other (skin, eye, liver, bowel, CNS, intestine, neck, peritoneum)	3 (5%)	4 (12.5%)	7 (7.6%)

TABLE 45. Distribution of alternative diagnoses in OD group (n=92) in the validation cohort for cellular immune signatures.

Diagnosis	N (%)
Cancer	8 (8.7%)
Pneumonia	11 (11.9%)
Sarcoidosis (including Sarcoidosis and other)	3 (3.3%)
LRTI/Chest Infection	9 (9.8%)
URTI	4 (4.3%)
Self-resolving illness	9 (9.8%)
Other	48 (52.2%)

*n=2 patients were administered with LTBI treatment

Appendix 11 Viability, cell recovery and cytokine responses to antigens

Recovery and viability

Vials of PBMCs from baseline blood samples of TB suspects were collected into the IDEA biobank between 2011-2013. The vials used in this sub-study were kept in cryogenic conditions for a median of 60 months (range 38-70 months) before being defrosted and used in experiments. After defrosting of cells and washing, cells were counted and compared to the total number of stored cells when cryopreserved. The mean recovery of live cells from cryopreserved PBMC vials for the full cohort was 48.1% (range 0-116.7%, median 48.4%) with a mean viability of 81% (range 0-100%, median 85%). The recovery rate from the HIV+ participants was lower than the HIV- (median of 48.96 in HIV- vs 33.33 in HIV+, $p = 0.01$), when compared using a Mann-Whitney U test. No differences were observed between TB and OD groups (whether or not HIV status was used to further stratify these groups or not).

Responses to antigen stimulations

Positive control antigen

PMA/ionomycin was used as a positive control stimulation, to ensure that the assay was successfully able to detect a definite response to stimulation, in the form of cytokine production. All of the samples stimulated with PMA/ionomycin produce the three cytokines to a level that was at least twice above background, displaying wide ranges of responses from less than 1% to >90% of cells producing cytokines ().

TABLE 46. Responses to PMA/ionomycin in full cohort. * 3 individuals (all HIV+) were not stimulated with PMA, as a means of reserving the few remaining cells for the unstimulated and PPD conditions.

	All TB (90)	Dosanjh 1 (58)	Dosanjh 2 (32)	OD/LTBI Dosanjh 4 (91)
Stimulated	90	58	32	88
Cytokine responses no. with response over background (%), range				
CD4+ IFN γ response	90 (100%) 0.28-46.9	58 (100%) 0.28-46.9%	32 (100%) 1.89-46.2%	88 (100%) 3.7-53.4%
CD4+ IL-2 response	90 (100%) 0.26-55	58 (100%) 0.26-47.4%	32 (100%) 8.23-55%	88 (100%) 20.9-63.5%
CD4+ TNF α response	90 (100%) 0.46-81.4	58 (100%) 0.46-81.4%	32 (100%) 16.5-75.9%	88 (100%)
CD4+ TNF α -only response	90 (100%) 0.12-30.5	58 (100%) 0.12-30.5%	32 (100%) 5.79-29.6%	87 (99%)
CD4+ Any cytokine response	90 (100%) 0.46-84.6%	58 (100%) 0.46-84.6%	32 (100%) 20.3-79.3%	87 (99%)

CD8+ IFN γ response	90 (100%) 1.82-89.8%	58 (100%) 1.82-89.8%	32 (100%) 11.2-79.2%	88 (100%)
CD8+ IL-2 response	90 (100%) 0.68-54.2%	58 (100%) 0.68-54.2	32 (100%) 0.94-50.5	88 (100%)
CD8+ TNF α + response	90 (100%) 5.99-95.3	58 (100%) 5.99-95.3	32 (100%) 16.4-91.9	88 (100%)

It was noted that the PMA/ionomycin condition reduced the viability of the lymphocyte population compared to the other conditions. Furthermore, this stimulation resulted in a massive reduction in the surface expression of CD4, such that the proportion of CD4+ cells as measured by flow cytometry staining was dramatically reduced; see *FIGURE 25*.

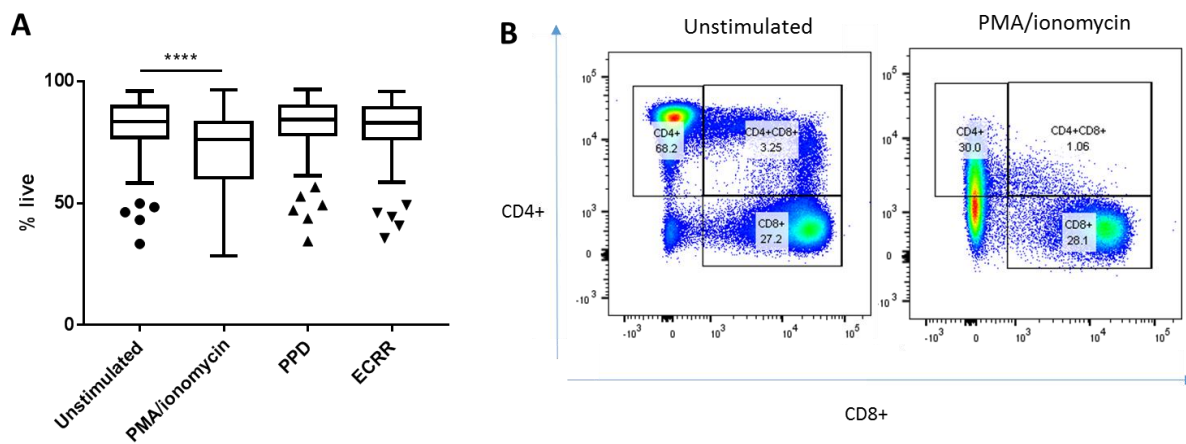


FIGURE 25. The effect of PMA/ionomycin on viability of lymphocytes (A) and CD4+ surface expression (B).

Purified peptide derivative (PPD)

All of the 181 samples which were plated were stimulated with PPD; the cytokine responses to this antigen are summarised in *TABLE 47*.

TABLE 47. Responses to PPD, stratified by Dosanjh criteria.

	All TB (90)	Dosanjh 1 (58)	Dosanjh 2 (32)	OD/LTBI Dosanjh 4 (91)
Stimulated	90	58	32	91
CD4+ IFN γ response	81 (90%)	51 (88%)	30 (94%)	77 (85) 0.005-8.3
CD4+ IL-2 response	78 (87%)	50	28	75
CD4+ TNF α response	84	54	30	81
CD4+ TNF α -only response	85	54	31	79
CD4+ Any cytokine response	82	52	30	78

CD8+ IFN γ response	33	21 (66%)	12	42
CD8+ IL-2 response	36	23	13	40
CD8+ TNF α + response	43	28	15	41

The effect of HIV infection status of response to PPD

TABLE 48. Responses to PPD, stratified by patient group and HIV infection status.

	TB HIV- (78)	TB HIV+ (12)	OD/LTBI HIV- (78)	OD/LTBI (13)
CD4+ IFN γ response	71	10	70	7
CD4+ TNF α -only response	74	11	72	7
CD4+ Any cytokine response	72	10	75	5

Mtb antigens

TABLE 49. Details of samples with positive CD4+ and CD8+ T cell cytokine responses to Mtb antigens.

One hundred and seven samples (59% of those assayed) were stimulated with the pooled RD-1-related Mtb antigens .

	All TB (90)	Dosanjh 1 (58)	Dosanjh 2 (32)	OD/LTBI Dosanjh 4 (91)
Stimulated	56	37	19	51
CD4+ IFN γ response	54	36	18	45
CD4+ IL-2 response	53	35	18	36
CD4+ TNF α response	55	36	19	38
CD4+ TNF α -only response	53	36	17	35
CD4+ Any cytokine response	53	35	18	39
CD8+ IFN γ response	30	21	9	22
CD8+ IL-2 response	31	23	8	24
CD8+ TNF α + response	30	19	11	17

TABLE 50. Responses to Mtb antigens, stratified by patient group and HIV infection status.

	TB HIV- (78)	TB HIV+ (12)	OD/LTBI HIV- (78)	OD/LTBI (13)
stimulated				
CD4+ IFN γ response	71	10	70	7
CD4+ TNF α -only response	74	11	72	7
CD4+ Any cytokine response	72	10	75	5

Appendix 12 Additional analyses for the validation of cellular immune signatures for the diagnostic evaluation of active TB

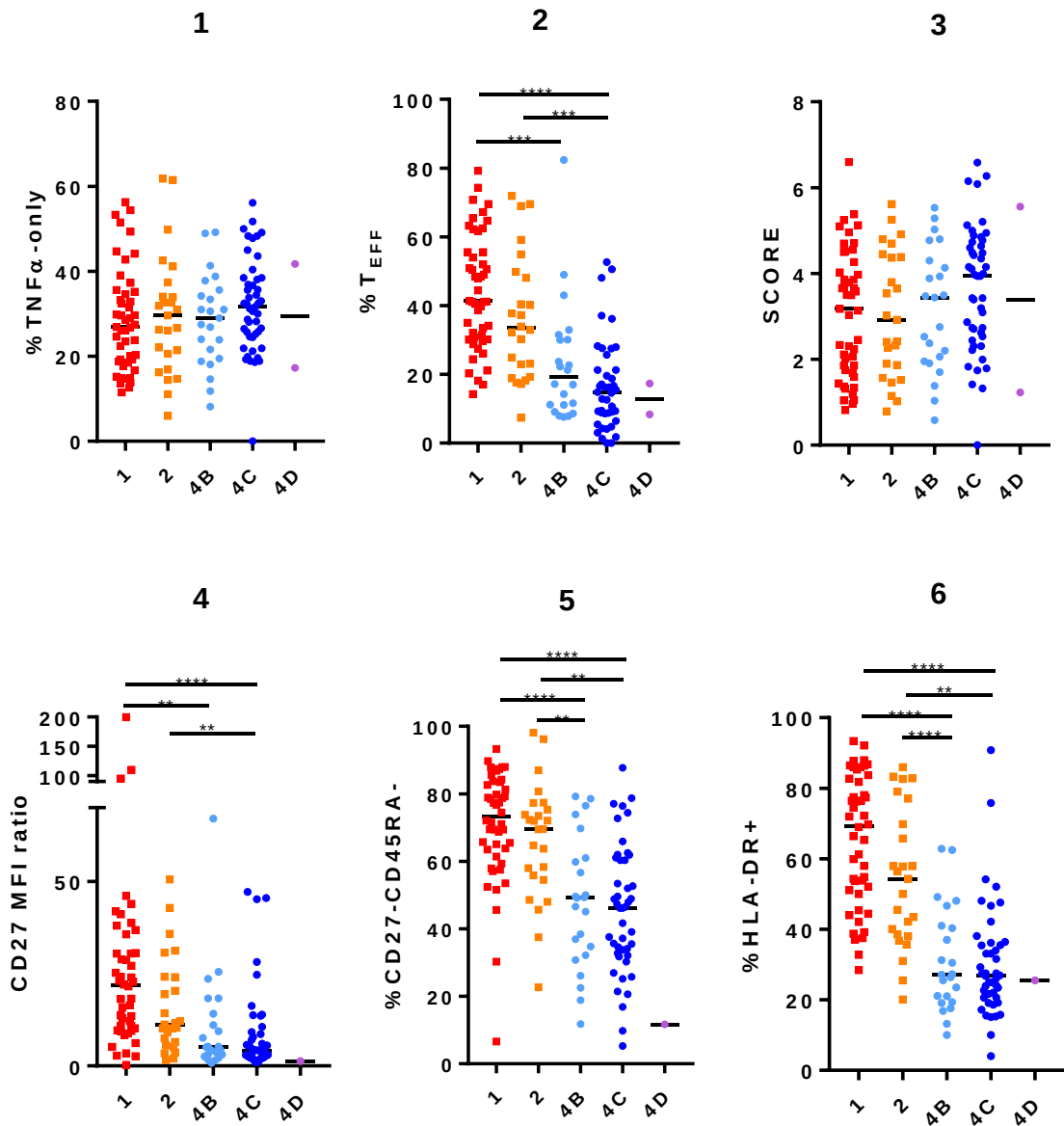


FIGURE 26. Performance of published cellular immune signatures according to Dosanjh classification (HIV- patients only). The signatures are as follows: 1) Harari²³ (%TNF α -only); 2) Pollock²⁴ (%T_{EFF}); 3) Rozot⁵⁰ (SCORE); 4) Portevin²⁶ (CD27 MFI Ratio); 5) Petruccioli²⁷ (%CD27-CD45RA-; 6) Adekambi²⁵ (%HLA-DR). Statistical differences between groups were determined using the Kruskal-Wallis with Dunn's post hoc test for multiple comparisons (* p<0.05; ** p<0.005; *** p<0.0005; p**** p<0.0001).

1 TABLE 51. Performance of cellular immune signatures for the detection of TB vs OD when Dosanjh 4 is further stratified by risk of latent TB. The signatures (Sig) are
2 numbered as follows: 1) Harari²³ (%TNF α -only); 2) Pollock²⁴ (%T_{EFF}); 3) Rozot⁵⁰ (SCORE); 4) Portevin²⁶ (CD27 MFI Ratio); 5) Petruccioli²⁷ (%CD27-CD45RA-; 6)
3 Adekambi²⁵ (%HLA-DR).. The number of samples per group, Area under the curve (AUC), cut-offs and sensitivity and specificity values are reported.

Comparison (Dosanjh)	Sig	TB (n)	OD/LTBI (n)	AUC (95% CI)	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV
1/2 vs 4B	1	72	23	0.51 (0.38-0.64)	<28.8* <14.5 ^{\$}	52.8 (40.7-64.7) 09.7 (04.0-19.0)	52.2 (30.6-73.2) 91.3 (72.0-98.9)	0.50 0.50	0.55 0.53
	2	72	22	0.80 (0.69-0.92)	>30.1* >43.7 ^{\$}	73.6 (61.9-83.3) 41.7 (30.2-53.9)	72.7 (49.8-89.3) 90.9 (70.8-98.9)	0.71 0.81	0.75 0.63
	3	72	23	0.52 (0.39-0.66)	<3.3* <1.4 ^{\$}	52.8 (40.7-64.7) 15.3 (07.9-25.7)	52.2 (30.6-73.2) 91.3 (72.0-98.9)	0.50 0.61	0.55 0.54
	4	71	23	0.75 (0.63-0.87)	>10.2* >23.8 ^{\$}	69.0 (56.9-79.5) 38.0 (26.8-50.3)	69.6 (47.1-86.8) 91.3 (72.0-98.9)	0.67 0.80	0.71 0.62
	5	71	23	0.80 (0.69-0.90)	>60.4* >76.6 ^{\$}	73.2 (61.4-83.1) 39.4 (28.0-51.8)	73.9 (51.6-89.8) 91.3 (72.0-98.9)	0.72 0.80	0.75 0.62
	6	71	23	0.88 (0.81-0.96)	>42.2* >49.6 ^{\$}	78.9 (67.6-87.7) 70.4 (58.4-80.7)	78.3 (56.3-92.6) 91.3 (72.0-98.9)	0.77 0.88	0.80 0.77
1/2 vs 4C	1	72	47	0.59 (0.49-0.70)	<29.7* <19.1 ^{\$}	56.9 (44.7-68.6) 25.0 (15.5-36.6)	57.5 (42.2-71.7) 91.5 (79.6-97.6)	0.55 0.73	0.60 0.57
	2	72	42	0.88 (0.82-0.95)	>25.9* >36.7 ^{\$}	79.2 (68.0-87.8) 56.9 (44.7-68.6)	78.6 (63.2-89.7) 90.5 (77.4-97.3)	0.77 0.84	0.81 0.70
	3	72	47	0.62 (0.51-0.72)	<3.5* <1.8 ^{\$}	52.8 (40.7-64.7) 25.0 (15.6-36.6)	53.2 (38.1-67.9) 91.5 (79.6-97.6)	0.51 0.73	0.55 0.57
	4	71	45	0.79 (0.70-0.88)	>9.2* >25.0 ^{\$}	77.5 (66.0-86.5) 32.4 (21.8-44.6)	77.8 (62.9-88.8) 91.1 (78.8-97.5)	0.76 0.77	0.79 0.60
	5	71	45	0.83 (0.75-0.90)	>59.8* >74.5 ^{\$}	73.2 (61.4-83.1) 42.3 (30.6-54.6)	71.1 (55.7-83.6) 91.1 (78.8-97.5)	0.70 0.81	0.75 0.63
	6	71	45	0.90 (0.84-0.96)	>39.7 >49.1 ^{\$}	81.7 (70.7-89.9) 70.4 (58.4-80.7)	82.2 (68.0-92.0) 91.1 (78.8-97.5)	0.81 0.88	0.83 0.77

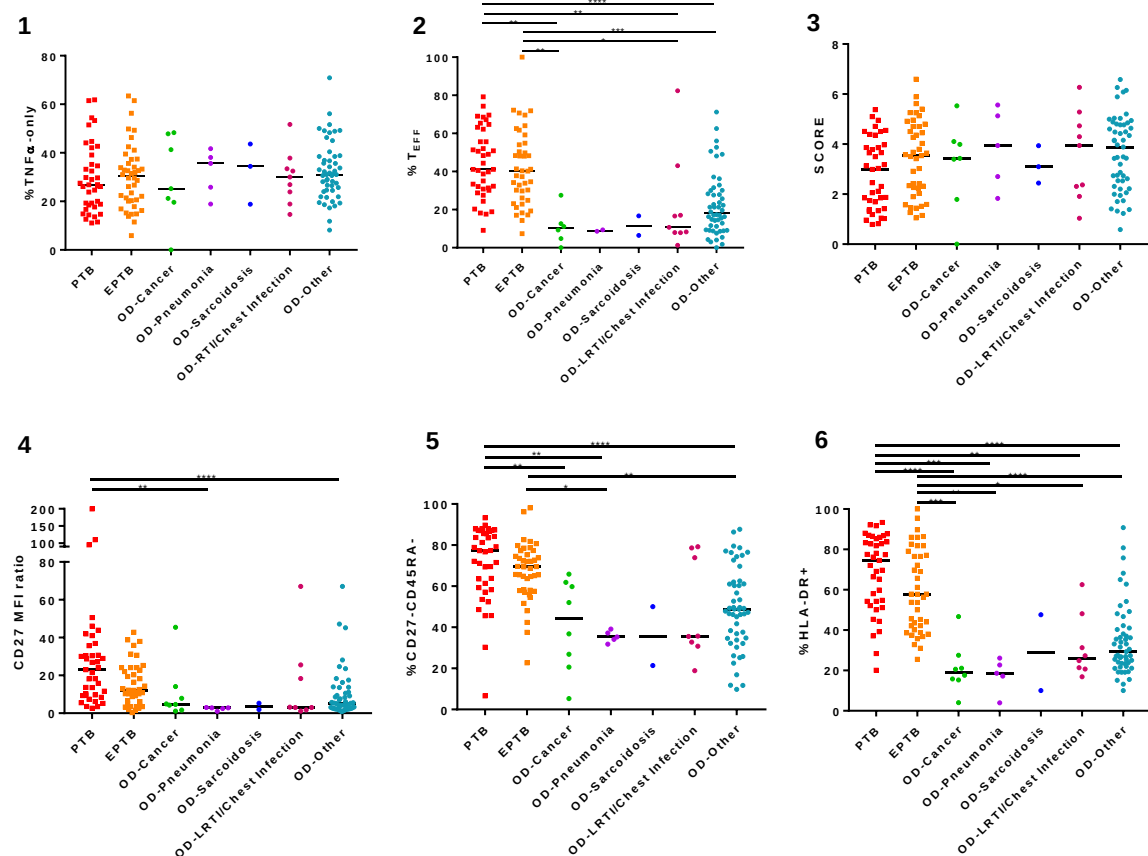


FIGURE 27. Performance of cellular immune signatures in different TB and OD/LTBI disease types.

The signatures are as follows: 1) Harari²³ (%TNF α -only); 2) Pollock²⁴ (%T_{EFF}); 3) Rozot⁵⁰ (SCORE); 4) Portevin²⁶ (CD27 MFI Ratio); 5) Petruccioli²⁷ (%CD27-CD45RA-; 6) Adekambi²⁵ (%HLA-DR). Statistical differences between groups were determined using the Kruskal-Wallis with Dunn's post hoc test for multiple comparisons (* p<0.05; ** p<0.005; *** p<0.0005; p**** p<0.0001)

TABLE 52. Performance of cellular immune signatures for diagnosing either PTB or EPTB. The signatures (Sig) are as follows: 1) Harari²³ (%TNF α -only); 2) Pollock²⁴ (%T_{EFF}); 3) Rozot⁵⁰ (SCORE); 4) Portevin²⁶ (CD27 MFI Ratio); 5) Petruccioli²⁷ (%CD27-CD45RA-; 6) Adekambi²⁵ (%HLA-DR). The number of samples per group, Area under the curve (AUC), cut-offs and sensitivity and specificity values are reported.

Comparison	Sig	TB (n)	OD/LTBI (n)	AUC (95% CI)	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)
PTB vs OD/LTBI	1	39	72	0.57 (0.45-0.69)	<28.5* <18.7 ^s	56.4 (39.6-72.2) 25.6 (13.0-90.3)	56.9 (44.7-68.6) 90.3 (81.0-96.0)
	2	39	66	0.88 (0.81-0.94)	>28.55* >37.9 ^s	82.05 (66.5-92.5) 61.5 (44.6-76.6)	81.8 (70.4-90.2) 90.9 (81.3-96.6)
	3	39	72	0.63 (0.52-0.74)	<3.2* <1.7 ^s	59.0 (42.1-74.4) 28.2 (15.0-44.9)	58.3 (46.1-69.9) 90.3 (91-96)
	4	39	69	0.83 (0.75-0.90)	>60.7* >25.1 ^s	74.4 (57.9-87) 46.2 (30.1-62.8)	75.4 (63.5-85.0) 91.3 (82.0-96.7)

	5	39	69	0.83 (0.75-0.92)	>60.7* >76.5 ^s	74.4 (57.9-87.0) 53.9 (37.2-69.9)	75.4 (63.5-85.0) 91.3 (82.0-96.7)
	6	39	69	0.93 (0.87-0.98)	>47.2* >49.6 ^s	84.6 (69.5-94.1) 84.6 (69.5-94.1)	85.5 (75.0-92.8) 91.3 (82.0-96.7)
EPTB vs OD/LTBI	1	25	72	0.53 (0.42-0.64)	<30.8* <18.7 ^s	50.0 (34.2-65.8) 21.4 (10.3-36.8)	50.0 (78.0-62.0) 90.3 (81.0-96.0)
	2	24	66	0.84 (0.77-0.92)	>25.9* >37.2 ^s	75.6 (59.7-87.6) 56.1 (39.8-71.5)	75.8 (63.6-85.5) 90.9 (81.3-96.6)
	3	25	72	0.51 (0.40-0.62)	<03.5* <01.6 ^s	47.6 (32.0-63.6) 19.1 (08.6-34.1)	48.6 (36.7-60.7) 90.3 (81.0-96.0)
	4	25	69	0.71 (0.61-0.81)	>08.2* >25.0 ^s	70.7 (54.5-83.9) 17.1 (07.2-32.1)	71.0 (58.9-81.3) 91.3 (82.0-96.7)
	5	25	69	0.80 (0.72-0.89)	>59.1* >76.5 ^s	70.7 (54.5-83.9) 24.4 (12.4-40.3)	71.0 (58.9-81.3) 91.3 (82.0-96.7)
	6	25	69	0.88 (0.82-0.94)	>39.5* >49.6 ^s	78.1 (62.4-89.4) 61.0 (44.5-75.8)	78.3 (66.7-87.3) 91.3 (82.0-96.7)

1

2

Appendix 13 Inclusion criteria of cellular immune signatures from previous publications

The publications proposing the cellular immune signatures each used different inclusion criteria and/or definitions of ‘responders’ in their studies (TABLE 53). To ensure a fair comparison of all signatures within this study, we applied universal inclusion criteria for the main analysis. However, we have also validated each signature according to the inclusion criteria reported in each of the reported publications which have shown promising diagnostic performance.

TABLE 53. Inclusion criteria for ‘responders’ and diagnostic performance analysis from previous publications. The signatures are as follows: 1) Harari²³ (%TNF α -only); 2) Pollock²⁴ (%T_{EFF}); 3) Rozot⁵⁰ (SCORE); 4) Portevin²⁶ (CD27 MFI Ratio); 5) Petruccioli²⁷ (%CD27-CD45RA-; 6) Adekambi²⁵ (%HLA-DR).

Parent population	Signature	Study	Inclusion criteria			
			Viability	Freq. of parent (%)	Above background	No. of cells in parent
CD4+ IFN γ /IL-2/TNF α +	% TNF α -only	Harari (2011)	-	>0.05	?	-
CD4+ TNF α -only	%CD45RA-CCR7-CD127-(%T _{EFF})	Pollock (2013)	-	>0.001	> x2	-
CD4+ and CD4+ IFN γ +	CD27 MFI Ratio	Portevin (2014)	-	>0.05		≥ 5
CD4+ IFN γ +	% CD27+/- CD45RA-	Petruccioli (2015)	-	>0.003	Ag-US >0.003	≥ 30
CD4+ IFN γ +	%HLA-DR+	Adekambi (2016)	75-95%	≥ 0.05	?	-
All	All phenotype signatures	VANTDET	Any	Any	> x2	≥ 10

Appendix 14 Performance of signatures after Mtb RD-1-related antigen stimulation

Performance of cellular immune signatures using Mtb-specific RD1-related antigens

Here we present the performance of the signatures when 4 Mtb RD-1-related antigens were used to stimulate the PBMCs, which was done in a large subset of the overall cohort, where PBMC numbers permitted. The 4 antigens used were ESAT-6, CFP-10, Rv3615c, Rv3879c.

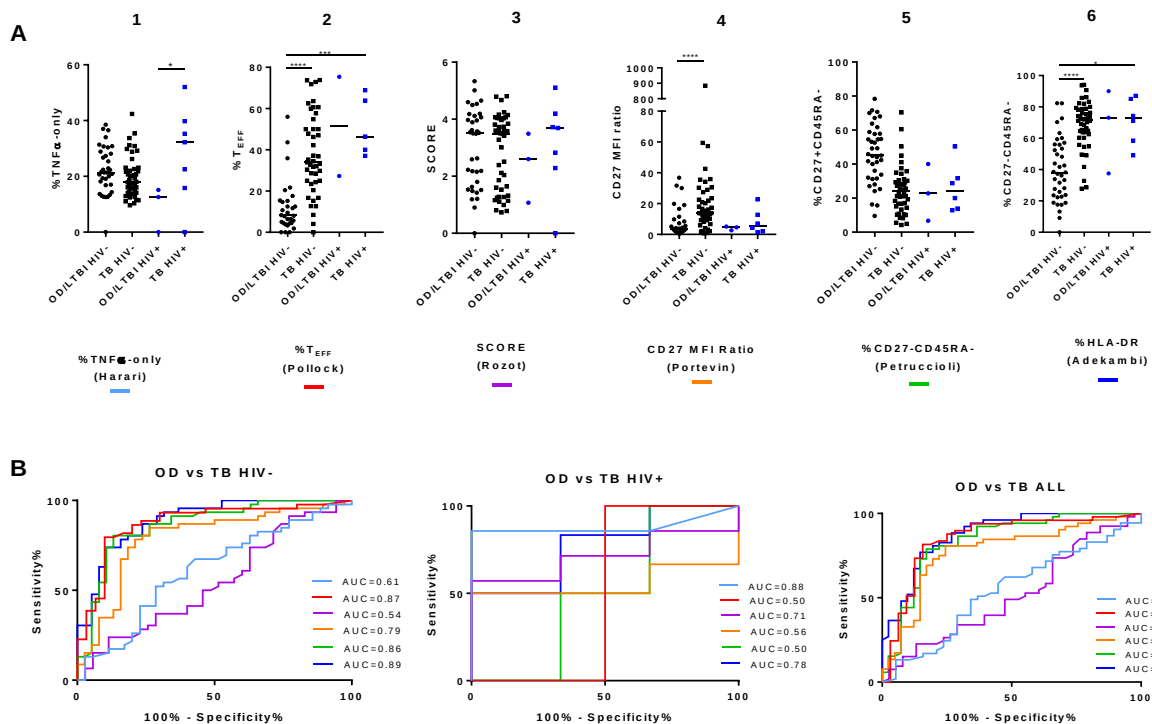


FIGURE 28. Performance of published signatures after stimulation with Mtb antigens. The signatures are as follows: 1) Harari²³ (%TNF α -only); 2) Pollock²⁴ (%T_{EFF}); 3) Rozot⁵⁰ (SCORE); 4) Portevin²⁶ (CD27 MFI Ratio); 5) Petruccioli²⁷ (%CD27-CD45RA-); 6) Adekambi²⁵ (%HLA-DR).

Overall, the performance of the signatures using the Mtb RD-1-related antigen stimulus was comparable to using PPD, and as with PPD, the %HLA-DR signature performed the best. The number of individuals within the HIV+ cohort was small, and thus little conclusion can be made about the performance of the signatures in this important sub-population of TB suspects.

Correlation between Mtb RD-1-related and PPD antigen responses

In order to check whether the PPD and Mtb RD1-related stimuli elicited a comparable response within the same individuals, we did a direct comparison within the individuals where both stimuli were used FIGURE 29.

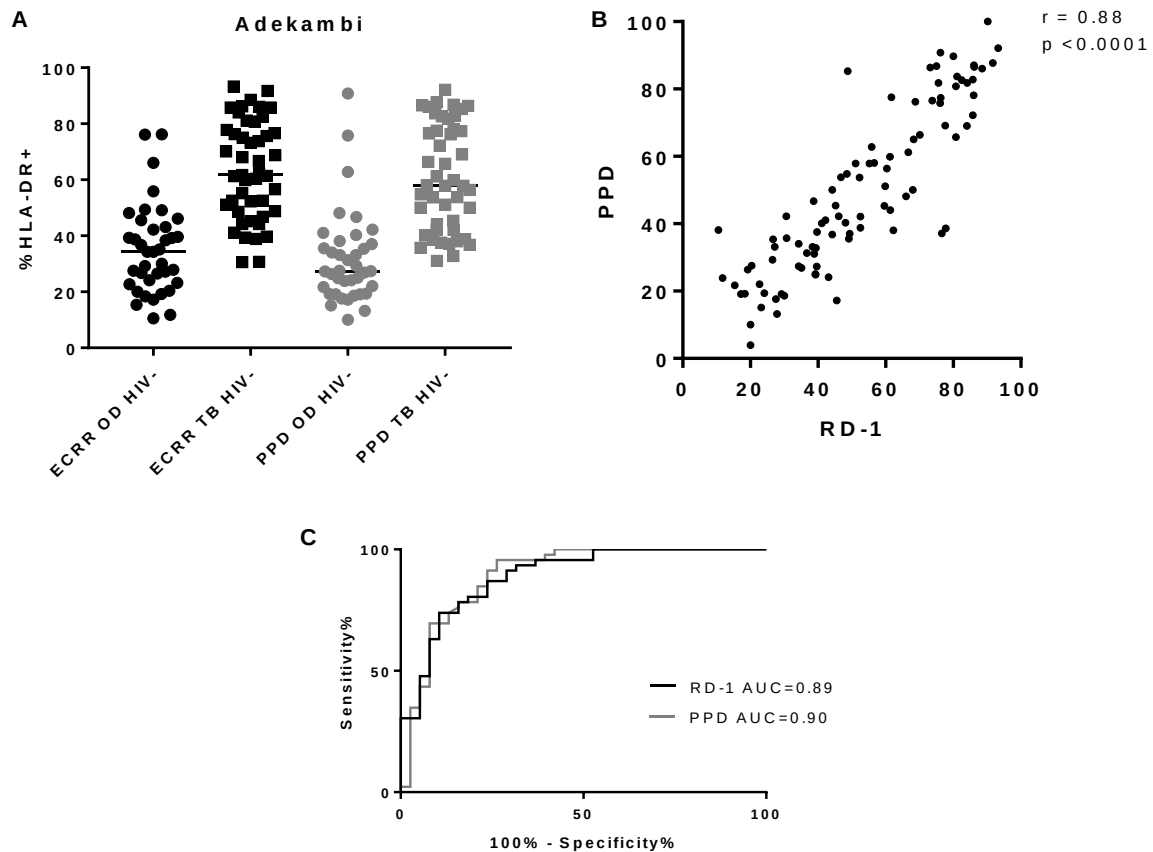


FIGURE 29. Correlation of %HLA-DR signature using PPD and RD-1 (ECRR) antigens. Agreement between %HLA-DR signature using the two conditions was shown, as demonstrated by similar group distributions, averages and variability (A), high correlation (statistics for Pearson correlation shown) (B), and comparable diagnostic accuracy for detection active TB (C).

Performance of signatures after ECRR stimulation

1

2 TABLE 54. Performance of published cellular immune signatures in the diagnosis of all active TB, stratified by HIV infection status.

Signature, ref	HIV	TB (n)	OD/ LTBI (n)	AUC (95 CI)	Cut-off	Sensitivity (95 CI)	Specificity (95 CI)	PPV	NPV
%TNF α -only ²³	-	46	35	0.61 (0.49-0.74)	<20.3* <12.7 ^{\$}	60.9 (45.4-75.9) 13.0 (4.9-26.3)	60.0 (42.1-76.2) 94.3 (80.8-99.3)	0.83 0.88	0.32 0.25
	+	7	3	0.88 (0.65-1.11)	>13.9 >15.5 ^{\$}	85.7 (42.1-99.6) 85.7 (42.1-99.6)	66.7 (09.4-99.2) 100 (29.2-100)	0.67 1.0	0.86 0.90
	-/+	53	38	0.54 (0.42-0.66)	<20.3* <12.5 ^{\$} <12.7 ^{\$}	56.6 (42.3-70.2) 13.2 (05.5-25.3) 13.2 (05.5-25.3)	55.3 (38.3-71.4) 94.7 (82.3-99.4) 89.5 (75.2-97.1)	0.49 0.89 0.81	0.63 0.25 0.24
	-	44	30	0.87 (0.79-0.96)	>17.2* >22.6 ^{\$}	81.8 (67.3-91.8) 80.0 (64.7-90.2)	80.0 (61.4-92.3) 90.0 (73.5-97.9)	0.93 0.96	0.57 0.57
%TEFF ²⁴ (note HIV+ the same whether * or \$, not a typo)	+	5	2	0.5 (-0.19-1.19)	>43.2 >55.2* >55.2 ^{\$}	60.0 (14.6-94.7) 40.0 (05.3-85.3) 40.0 (05.3-85.3)	50.0 (01.3-98.7) 50.0 (01.3-98.7) 50.0 (01.3-98.7)	0.48 0.38 0.38	0.62 0.52 0.52
	-/+	49	32	0.85 (0.76-0.95)	>21.1* >36.4 ^{\$}	81.6 (68.0-91.2) 51.0 (36.3-65.6)	81.3 (63.6-92.8) 90.6 (75.0-98.0)	0.93 0.95	0.57 0.36
	-	46	35	0.54 (0.42-0.67)	<3.5* <1.2 ^{\$}	50.0 (34.9-65.1) 15.2 (06.3-28.9)	51.4 (34.0-68.6) 91.4 (76.9-98.2)	0.77 0.85	0.24 0.25
SCORE ⁵⁰	+	7	3	0.71 (0.39-1.0)	>2.7* >3.6 ^{\$}	71.4 (29.0-96.3) 57.1 (18.4-90.1)	66.7 (09.4-99.2) 100 (29.2-100)	0.62 1.0	0.75 0.75
	-/+	53	38	0.51 (0.39-0.64)	<03.5* <1.18 ^{\$}	49.1 (35.1-63.2) 15.1 (6.7-27.6)	47.4 (31.0-64.2) 92.1 (78.6-98.3)	0.75 0.86	0.22 0.25
	-	46	38	0.79 (0.69-0.89)	>08.8* >20.1 ^{\$}	78.3 (63.6-89.1) 34.8 (21.4-50.3)	76.0 (62.7-90.5) 92.1 (78.6-98.3)	0.92 0.96	0.52 0.30
CD27 MFI Ratio ²⁶	+	6	3	0.56 (0.17-0.94)	>4.8* >6.1 ^{\$}	50.0 (11.8-88.2) 50.0 (11.8-88.2)	66.7 (9.4-99.2) 100 (29.2-100)	0.54 1.0	0.63 0.72
	-/+	52	41	0.78 (0.68-0.88)	>07.1* >19.5	75.0 (61.1-86.0) 32.7 (20.3-47.1)	75.6 (60.0-87.6) 90.2 (76.9-97.3)	0.91 0.92	0.80 0.29
	-	46	38	0.86 (0.78-0.95)	>58.3* >70.7 ^{\$}	80.4 (66.1-90.6) 54.4 (39.0-69.1)	81.6 (65.7-92.3) 92.1 (78.6-98.3)	0.93 0.96	0.56 0.38
% CD27-CD45RA- ²⁷	+	6	3	0.5 (0.01-0.99)	>73.7* >73.7 ^{\$}	50.0 (11.8-88.2) 50.0 (11.8-88.2)	66.7 (9.4-99.2) 66.7 (9.4-99.2)	0.54 0.54	0.63 0.63
	-/+	52	41	0.84 (0.75-0.93)	>59.0*	78.9 (65.3-88.9)	78.1 (62.4-89.4)	0.92	0.53

					>72.90	44.2 (30.5-58.7)	90.2 (76.9-97.3)	0.94	0.33
%HLA-DR ²⁵	-	46	38	0.89 (0.82-0.96)	>46.5	80.4 (66.1-90.6)	81.6 (65.7-92.3)	0.93	0.56
					>56.3	63.0 (47.6-76.8)	92.1 (78.6-98.3)	0.96	0.43
	+	6	3	0.78 (0.45-1.1)	>78.8*	66.7 (22.3-95.7)	66.7 (09.4-99.2)	0.61	0.72
					>82.5	50.0 (11.8-88.2)	100 (29.2 -100)	1.0	0.72
	-/+	52	41	0.88 (0.81-0.95)	>48.3	80.8 (67.5-90.4)	80.5 (65.1-91.2)	0.93	0.56
					>66.4	51.9 (37.6-66)	90.2 (76.9-97.3)	0.95	0.36

1

2

3

Appendix 15 Additional cohort characteristics for the molecular rule-out test assessment

TABLE 55. Distribution of disease types and severity within the TB patients in the GO/NO GO cohort for the validation of the molecular rule out tests.

	Active TB		Total N=46
	T-Spot + (n=24)	T-Spot – (n=22)	
Dosanjh			
Culture Confirmed (1)	21	16	37
Highly Probable TB (2)	4	6	10
Smear positive, n (%)	3	5	8
Smear not tested,	4	3	7
Pulmonary TB	5	12	17
Extrapulmonary TB	17	8	25
Pulmonary and extrapulmonary TB	2	2	4
Site of Infection			
Lung only	5	12	17
Lymph node only	14	5	1
Lung and Lymph node	1	1	2
Lung and other	1	1	2
Lymph node and other	1	1	2
Pleura	1	1	2

TABLE 56. Distribution of other diseases in the OD patients within the TB patients in the GO/NO GO cohort for the validation of the molecular rule out tests.

	Other Disease		Total n=48
	T-Spot + (n=22)	T-Spot – (n=26)	
Cancer	1	3	4
Sarcoidosis	0	2	2
Chest Infection	1	2	3
Pneumonia	6	6	12
Upper Respiratory Tract Infection	1	1	2
Lower Respiratory Tract Infection	0	1	1
LTBI	1		2
Other	12	11	23

Appendix 16 Additional data for the validation of candidate rule out tests

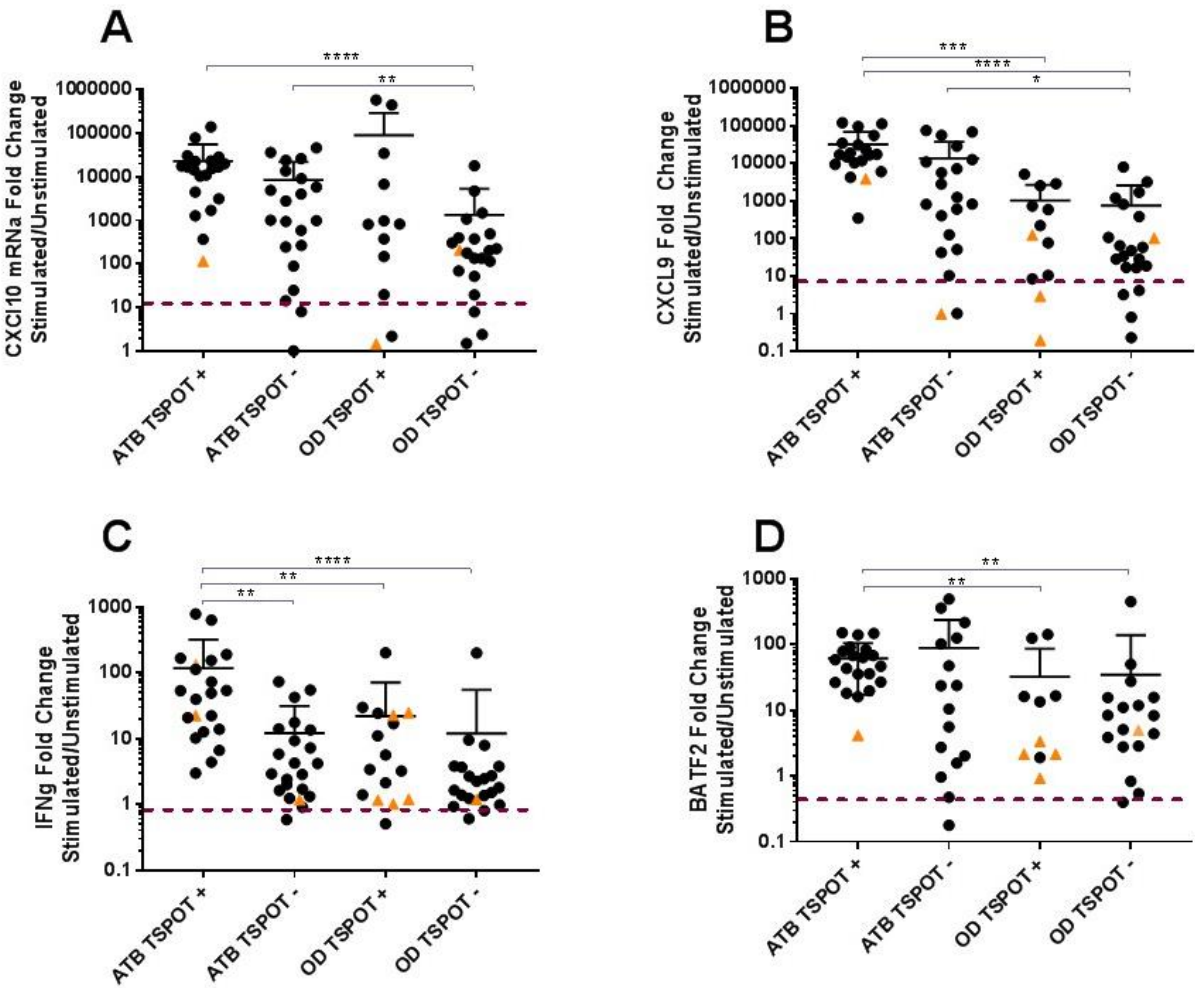


FIGURE 30. Relative abundance of CXCL10, CXCL9, IFN γ and BATF2 in key patient groups within the GO, NO/GO cohort for assessment of the candidate molecular rule-out test. The cut-offs shown are the level of expression required for the test to provide >95% sensitivity for all active TB cases. HIV+ individuals are shown in orange triangles. The gene expression cut-off values for achieving 95% sensitivity for all TB cases are displayed. Statistical differences between groups were determined using the Kruskal-Wallis with Dunn's post hoc test for multiple comparisons (* $p < 0.05$; ** $p < 0.005$; *** $p < 0.0005$; p**** $p < 0.0001$)

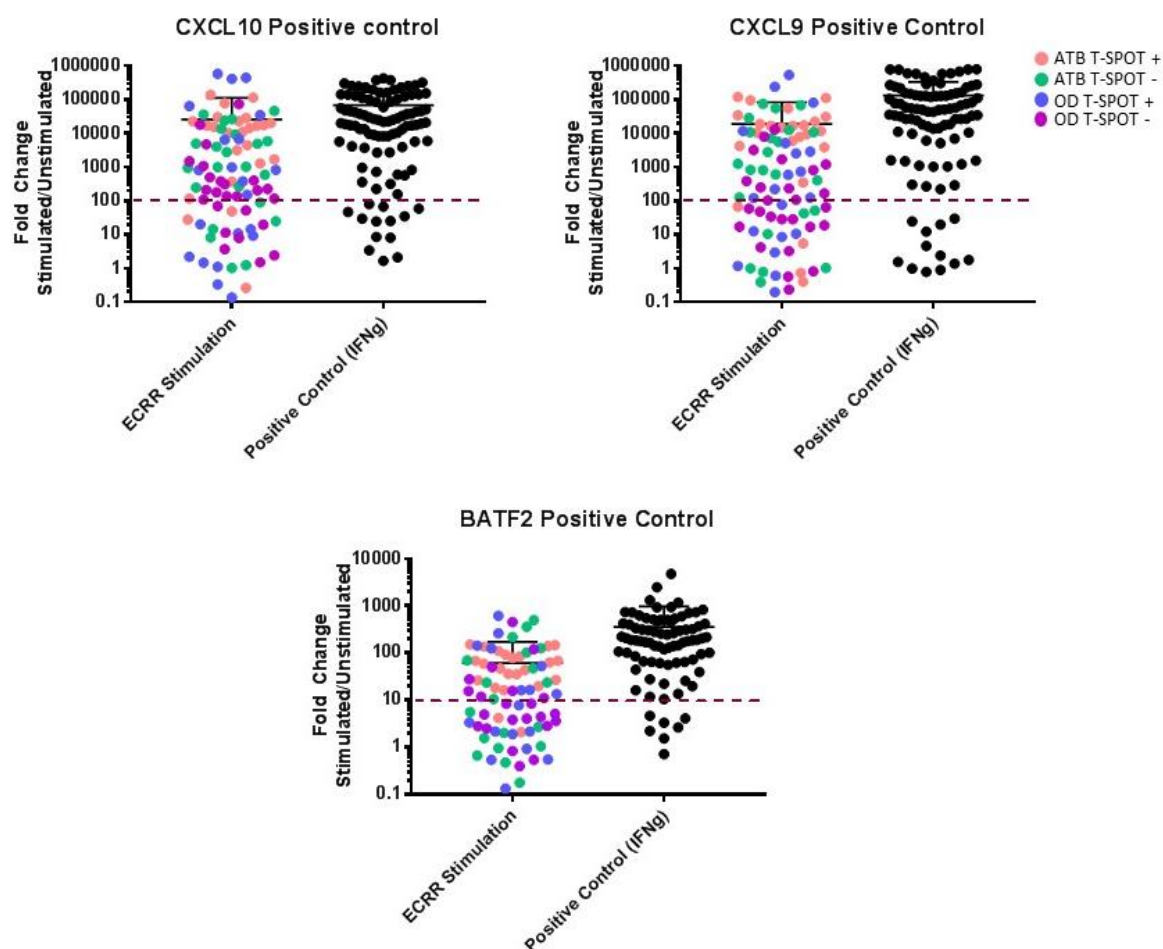


FIGURE 31. Selection of cut-off for target genes in response to IFN γ positive control condition.

TABLE 57. Measurement of IFN γ transcripts in stimulated PBMC samples using qRT-PCR.

	Active TB		Other Disease	
	T-SPOT +	T-SPOT -	T-SPOT +	T-SPOT-
n	24	22	22	26
Total for analysis	21	20	16	21
QC failed total, n :	3	2	6	5
<i>HPRT1</i> , n	3	2	6	5
<i>Indeterminate</i> , n	0	0	0	0
ECRR/ US fold change	122.09(0.03-787.82)	12.99 (0.90-70.02)	21.73 (0.51-207.58)	12.01 (0.61-200)
IFN- γ / US	21, 72.82 (0.06-344.81)	21, 49.76 (0.60-425.01)	18, 26.79 (0.72-179.44)	21, 25.52(0.40-449.46)

TABLE 58. Diagnostic accuracy data for candidate molecular rule out test: all TB vs all other disease. AUC, sensitivity and specificity values are shown, with 95% CI, and corresponding cut-offs.

	ATB TSPOT + (n)	ATB TSPOT - (n)	OTHER DISEASE TSPOT + (n)	OTHER DISEASE TSPOT- (n)	AUC (95% CI)	CUT OFF	SENSITIVITY (95% CI)	SPECIFICITY (95% CI)
T-SPOT	24	22	22	26	0.53(0.41-0.65)		52 (37% to 67%)	54 (39% to 69%)
CXCL10	19	19	12	21	0.72 (0.62-0.86)	11	95 (82% to 99%)	15 (5.1% to 32%)
CXCL9	18	18	12	21	0.84 (0.74-0.94)	1	97 (85% to 99%)	9.1 (1.9% to 24%)
IFNg	21	20	16	21	0.70(0.59-0.82)	0.85	95 (83% to 99%)	8.1 (1.7% to 22%)
BATF2	19	15	10	18	0.72 (0.59-0.8)	0.44	97 (85% to 100%)	3.6 (0.09% to 18%)

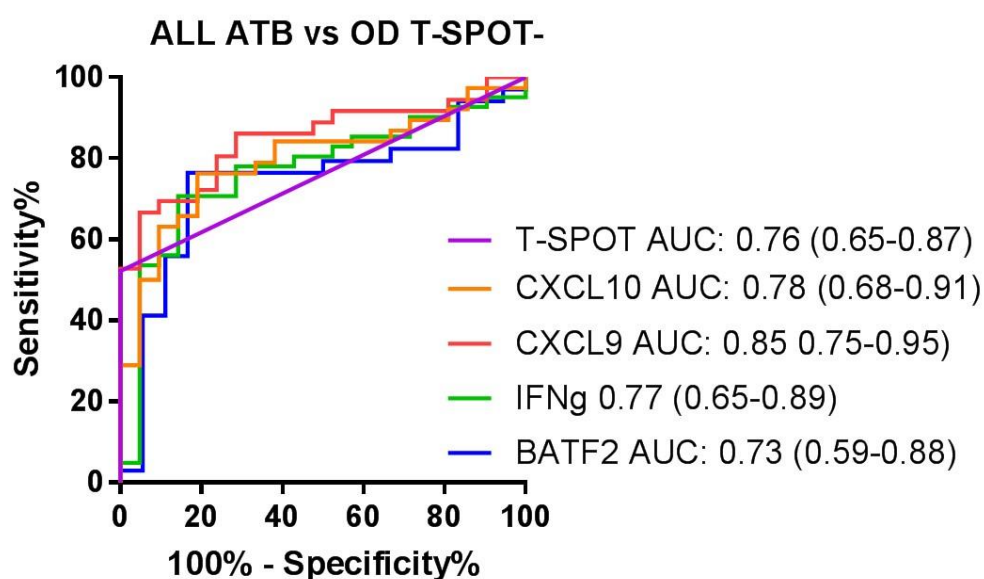


FIGURE 32. Diagnostic performance of CXCL10, CXCL9, IFNg and BATF2 Mtb-specific gene expression in all TB vs TSPOT- OD.

TABLE 59. Diagnostic accuracy data for candidate molecular rule out test: all TB vs OTHER DISEASE T-SPOT-. AUC, sensitivity and specificity values are shown, with 95% CI, and corresponding cut-offs.

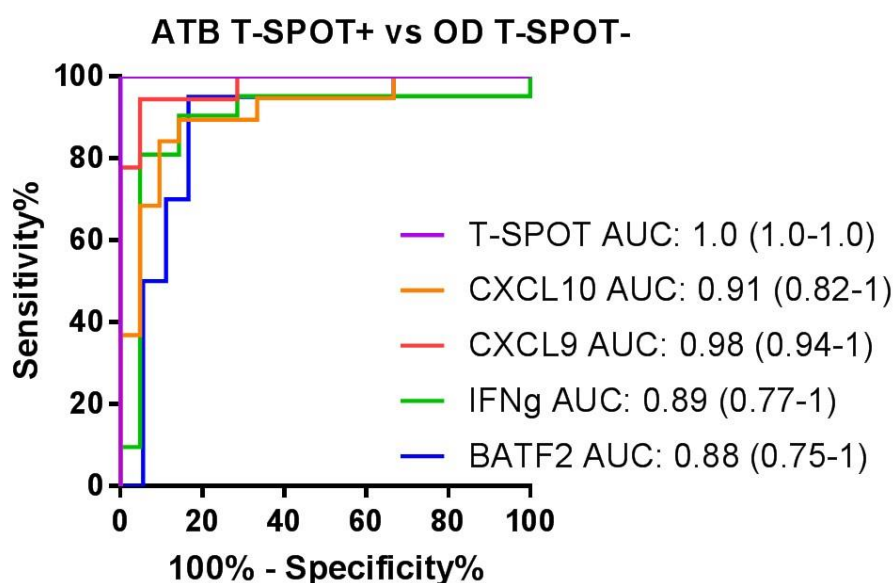
	ATB TSPOT +	ATB TSPOT	OTHER DISEASE	AUC (95% CI)	CUT OFF	SENSITIVITY (95% CI)	SPECIFICITY (95% CI)
--	----------------	--------------	------------------	-----------------	------------	-------------------------	-------------------------

	(n)	– (n)	TSPOT- (n)				
T-SPOT	24	22	26	0.76 (0.65-0.87)		52 (37% - 67%)	100 (87% - 100%)
CXCL10	19	19	21	0.78 (0.65-0.91)	11	95 (82% to 99%)	14 (3% to 36%)
CXCL9	18	18	21	0.85 (0.75-0.95)	1	97 (85% to 100%)	9.5 (1.2% to 30%)
IFN γ	21	20	21	0.77 (0.65-0.89)	0.85	95 (85% to 99%)	9.5 (1.2% to 30%)
BATF2	19	14	18	0.74 (0.59-0.88)	0.44	97 (85% to 100%)	5.6 (0.14% to 27%)

1

2 TABLE 60. Diagnostic accuracy data for candidate molecular rule out test: TB TSPOT+ vs OTHER
3 DISEASE TSPOT-. AUC, sensitivity and specificity values are shown, with 95% CI, and corresponding
4 cut-offs.

	ATB TSPOT+ (n)	OTHER DISEASE T-SPOT- (n)	AUC (95% CI)	CUT OFF	SENSITIVITY (95% CI)	SPECIFICITY (95% CI)
T-SPOT	24	26	1.0 (1.0-1.0)		100(86.28% to 100%)	100 (86.77% to 100%)
CXCL10	19	21	0.91 (0.81- 1.0)	340	95% (86.8% to 100%)	67 (43% to 85%)
CXCL9	18	21	0.98 (0.94- 1.0)	229	100 (81% to 100%)	71 (48% to 100%)
IFN-γ	21	21	0.89 (0.77- 1.0)	2.9	95 (76% to 100%)	71 (48% to 89%)
BATF2	19	18	0.88 (0.75- 1.0)	16	95 (75% to 100%)	83 (59% to 96%)



5

FIGURE 33. Diagnostic performance of CXCL10, CXCL9, IFN γ and BATF2 Mtb-specific gene expression in TSPOT+ TB vs TSPOT- OD.

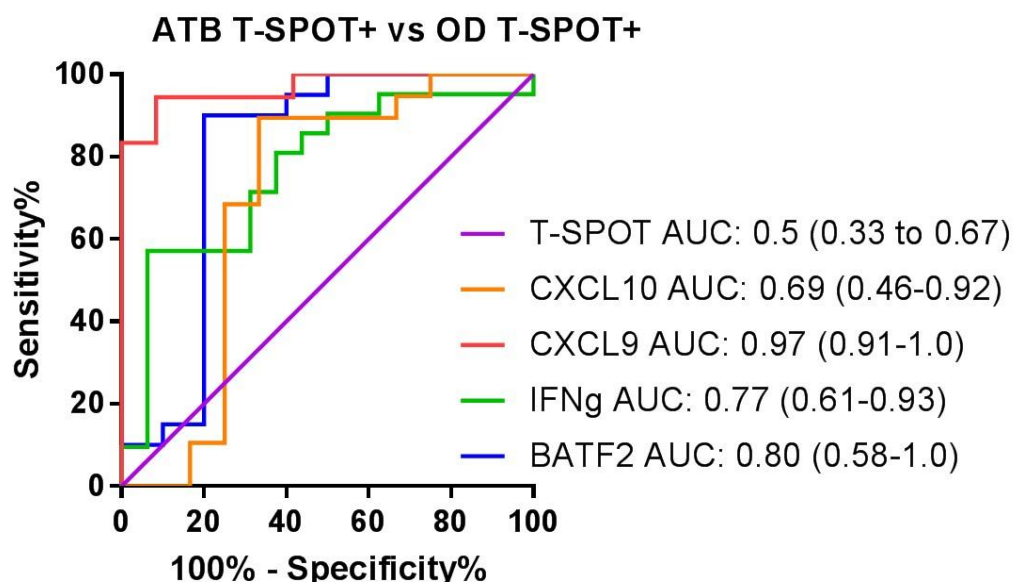


FIGURE 34. The performance of qRT-PCR detection of CXCL10, CXCL9, IFN γ , BATF2 and T-SPOT for discriminating between T-SPOT+ TB cases (n=19-24) and T-SPOT + OD/LTBI cases (n=10-22).

TABLE 61. Diagnostic accuracy data for candidate molecular rule out test: ATB T-SPOT + vs OD T-SPOT +. AUC, sensitivity and specificity values are shown, with 95% CI, and corresponding cut-offs.

	ATB TSPOT+ (n)	OTHER DISEASE T-SPOT+ (n)	AUC (95% CI)	CUT OFF	SENSITIVITY (95% CI)	SPECIFICITY (95% CI)
T-SPOT	24	22	0.50 (0.35-0.67)		0 (86.28% to 100%)	100 (85% to 100%)
CXCL10	19	12	0.69 (0.46-0.92)	260	95% (74% to 100%)	33 (9.9% to 65%)
CXCL9	18	12	0.97 (0.91-1.0)	286	100 (81% to 100%)	71 (48% to 89%)
IFN-γ	21	16	0.77 (0.61-0.93)	2.9	95 (76% to 100%)	71 (48% 89%)
BATF2	19	10	0.8 (0.58-1.0)	16	95 (75% to 100%)	60 (59% to 96%)

1 TABLE 62. Performance of IFN γ and CXCL10 analytes from Quantiferon supernatants detected by MSD
2 for the detection of all TB in the GO/NO GO cohort.

	ATB QFT + (n)	ATB QFT – (n)	OTHER DISEASE QFT + (n)	OTHER DISEASE QFT - (n)	AUC (95% CI)	CUT OFF	SENSITIVITY (95% CI)	SPECIFICITY (95% CI)
QFT	20	21	5	23	0.7 (0.5- 0.8)		49 (33%-65%)	63 (63%- 94%)
CXCL10	20	21	5	23	0.66 (0.53- 0.79)	-19	95 (83%-99%)	3.6 (0.09%- 18%)
IFN γ	19	21	5	22	0.69 (0.57- 0.82)	-18	95 (83%-99%)	3.7 (0.09%- 19%)

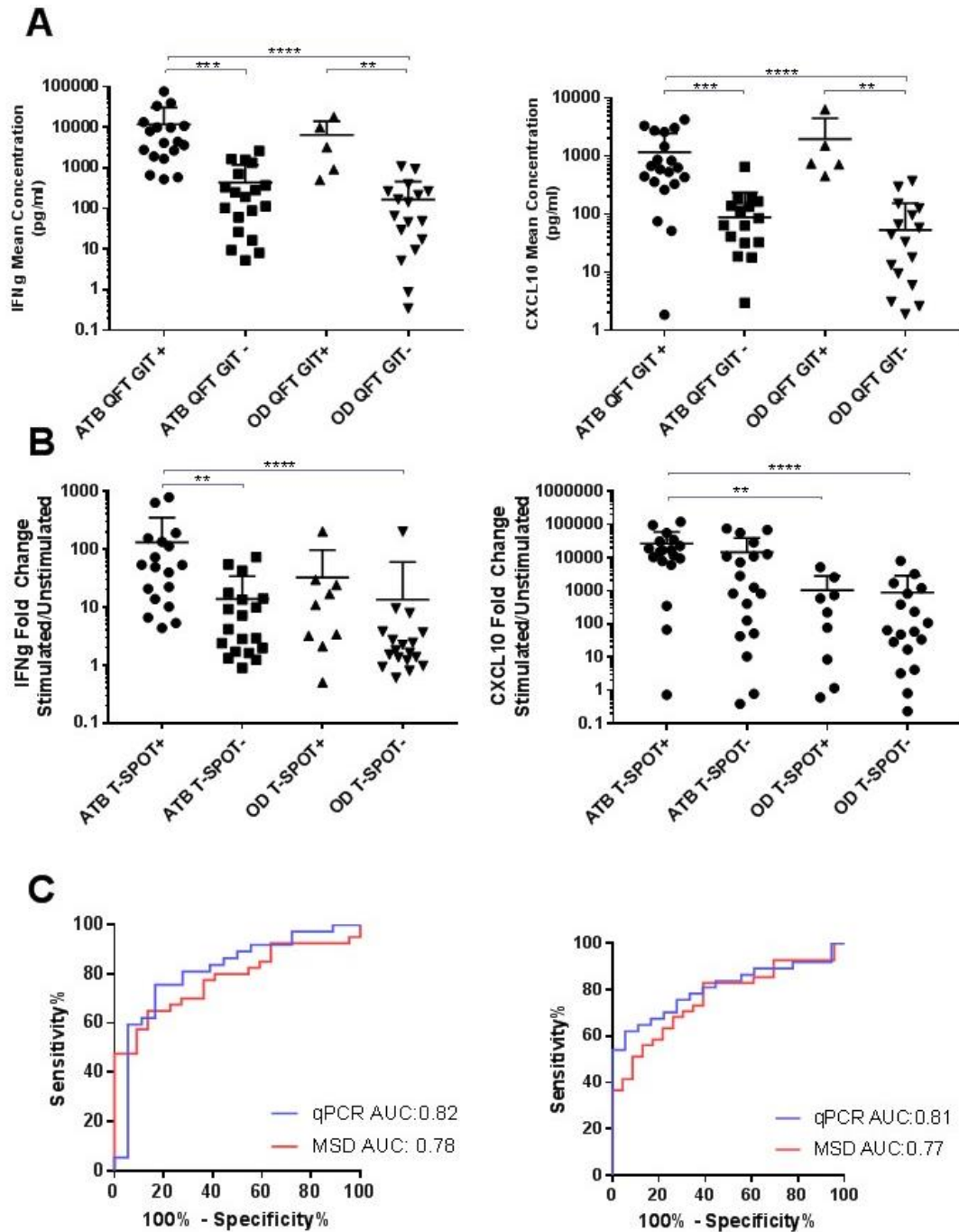


FIGURE 35. Comparison of detection of IFN γ or CXCL10 in Mtb-antigen stimulated samples, using either MSD from QFT supernatants (A) or qRT-PCR from stimulated PBMCs (B) and comparing between patient groups in the overlapping cohort on n=75 where both tests were performed. Protein levels and fold-changes in gene expression were compared between groups using the Kruskal-Wallis with Dunn's post hoc test for multiple comparisons (* p<0.05; ** p<0.005; *** p<0.0005; p**** p<0.0001) The diagnostic accuracy of using either IFN γ or CXCL10 using either platform, for detection all TB cases, was assessed by ROC curve analysis (C).

Appendix 17 Additional tables for health economic analysis – methodology

TABLE 63. Health impact of adverse event due to TB treatment. All values from ⁶⁴

Adverse event	Symbol	Cost of treatment (£)	QALY loss
Hepatotoxicity	q_h	Gamma(6.679, 87.889)	Gamma(65.753, 7×10^{-5})
Nausea and vomiting	q_n	63	Gamma(109.67, 3×10^{-6})

TABLE 64. Time to obtain a result and cost of new tests. Estimate staff costs using available knowledge/prediction of equivalent staff time, from VANTDET study team. *assuming £24.20 per hour staff cost for Research Associate or equivalent and Unif[6,8] hours staff time

Test	Unit cost (£)	Total staff cost (£)*	Time for a result (days)
Transcriptomic PCR test	16.5	Unif[145.2,193.6]	Unif[1, 1.5]
Cellular immune flow cytometry	150	Unif[145.2,193.6]	Unif[2, 7]
Molecular rule out RT-PCR test	Unif[27,35]	Unif[145.2,193.6]	Unif[2,5]

TABLE 65. Health care professional consultation visit monetary costs incurred. SE, standard error

Consultation type	Cost (£)	Distribution	Sources
First visit: respiratory medicine, multi-professional	241	Gamma(53.3, 4.52)	National tariff 2014–15, Annex 5A ⁶⁸ ⁶⁷
	167 (SE 33)		Hughes et al (2012) ⁶⁹
Follow up visit: respiratory medicine, multi-professional	143	Gamma(18.78, 7.62)	National tariff 2014–15, Annex 5A ⁶⁸
	167 (SE 33)		Hughes et al (2012) ⁶⁹

1 TABLE 66. Test and sampling procedure costs for cost-effectiveness calculations. SE, standard error

Test	Unit cost (£)	Distribution	Sources
Culture	22.29 (SE 2.23)	Gamma(100,0.22)	Drobniewski et al (2015) ⁷⁰
Sputum smear	7	Gamma(106,0.07)	NICE guidance NG33 (2016) ⁷
microscopy	1.56 (SE 0.68)		Hughes et al (2012) ⁶⁹
Tuberculin skin	17.48		Sutcliffe (2016)
test	16 [8, 32]	Uniform(8,36)	NICE guidance CG33 (2010) ⁶⁶
T-SPOT.TB	59.57		Sutcliffe (2016)
	55 [45, 99]	Uniform(50,106)	NICE guidance CG33 (2010) ⁷
QFT-IT	58 [29, 87]	Uniform(29,87)	Pareek et al ¹⁶
Chest X-ray	35		NICE guidance NG33 (2016) ⁷
	28 [19, 34]	Uniform(23,43)	NICE guidance CG33 (2010) ⁶⁶
Bronchoalveolar	23.24	NA	St Mary's R & D office
lavage	[11.62, 46.48]	Uniform(11.62, 46.48)	Proportions from NICE clinical guidelines (2006), Pareek et al ¹⁶ , NICE guidance CG33 (2010) ⁹
EBUS	2634	NA	St Mary's R & D office
Bronchoscopy	612	NA	St Mary's R & D office
procedure	[306, 1224]	Uniform(306,1224)	Proportions from NICE clinical guidelines (2006) (12), Pareek et al ¹⁶ , NICE guidance CG33 (2010) ⁹
Histology from	25	NA	St Mary's R & D office
biopsy	[12.5, 50]	Uniform(12.5,50)	Proportions from NICE clinical guidelines (2006), Pareek et al ¹⁶ , NICE guidance CG33 (2010) ⁹
Needle aspirate	90.21	NA	St Mary's R & D office
	[45.1, 180.42]	Uniform(45.1,180.42)	Proportions from NICE clinical guidelines (2006) (12), Pareek et al ¹⁶ , NICE guidance CG33 (2010) ⁹
Polymerase chain	202.45	NA	St Mary's R & D office
reaction	[101.2, 404.9]	Uniform(101.2,404.9)	Proportions from NICE clinical guidelines (2006) (12), Pareek et al ¹⁶ , NICE guidance CG33 (2010) ⁹

2

3

4

5

TABLE 67. Active TB treatment costs for the first 60 days. Final cost is the sum of all listed 60-day total drug costs.

Drug	Dosage (mg/day) ^a	Dosage by patient weight (mg/kg/day) ^a	Batch cost ^b (£)	Quantity per batch (capsule or tablet) ^b	Dosage per capsule or tablet ^b	60-day total cost (£)
Rifampicin	600	–	48	100	300	57.60
Isoniazid	300	–	19.24	56	50	123.69
Pyrazinamide	2000	–	38.34	30	500	306.72
Ethambutol	–	15	42.74	56	400	116.74

^aJoint Formulary Committee, 2016. 5.1.9 Antituberculosis drugs. In: JOINT FORMULARY COMMITTEE. *British National Formulary*. 69. London: BMJ Group and Pharmaceutical Press. ^bNICE guidance NG33 (2016)⁷

Distributional formulation of individual level/sample uncertainty

The number of individuals who enter the diagnostic pathway was defined as n . Of these, we defined n_+ as ATB cases and the remainder as non-ATB. We defined the probability of a given patient being an ATB case as n_+/n (and so the probability of non-ATB is $1 - (n_+/n)$). This can therefore be considered as a draw from a Bernoulli distribution and so for the total sample the Binomial distribution gives the number of ATB and non-ATB cases in a sample population, denoted with subscript TB and $\sim$ TB respectively, as

$$X_{TB} \sim \text{Bin}(n_+/n, n)$$

$$X_{\sim TB} = n - X_{TB}$$

that is,

$$p(X_{TB} | n_+/n) = \binom{n}{X_{TB}} (n_+/n)^{X_{TB}} (1 - n_+/n)^{(n - X_{TB})}.$$

By the same principle, patients are then randomly split between those that who test positive and negative.

$$X_+^{TB} \sim \text{Bin}(p_+, X_{TB})$$

$$X_-^{TB} = X_{TB} - X_+^{TB}$$

$$X_+^{\sim TB} \sim \text{Bin}(1 - p_+, X_{\sim TB})$$

$$X_-^{\sim TB} = X_{\sim TB} - X_+^{\sim TB}$$

This process results in a final random subdivision of the sample population into one of the four end states.

Use as rule-out test

Each of the terminal nodes (outcomes) of the decision tree has an associated cost and health utility (measured in QALYs). For the patient cohort these were defined as

$$C_{novel} = C_{ruleout} + \left(\frac{n_+}{n}\right)\bar{c}_+ + \left(\frac{n_-}{n}\right)(1 - p_-)\bar{c}_- \\ E_{novel} = qT_{ruleout} + \left(\frac{n_+}{n}\right)q(\bar{t}_+ + (1 - p_+)T_{followup}) + \left(\frac{n_-}{n}\right)q(1 - p_-)\bar{t}_-$$

Use as rule-in test

Cost-effectiveness outcomes

For the patient cohort the cost and health detriment were defined as

$$C_{novel} = C_{rulein} + \left(\frac{n_+}{n}\right)(1 - p_+)\bar{c}_+ + \left(\frac{n_-}{n}\right)((1 - p_-)C_{Tx} + p_-\bar{c}_-) \\ E_{novel} = qT_{rulein} + \left(\frac{n_+}{n}\right)(1 - p_+)q\bar{t}_- + \left(\frac{n_-}{n}\right)(p_-q\bar{t}_- + (1 - p_-)(q \\ + q_{Tx})T_{followup})$$

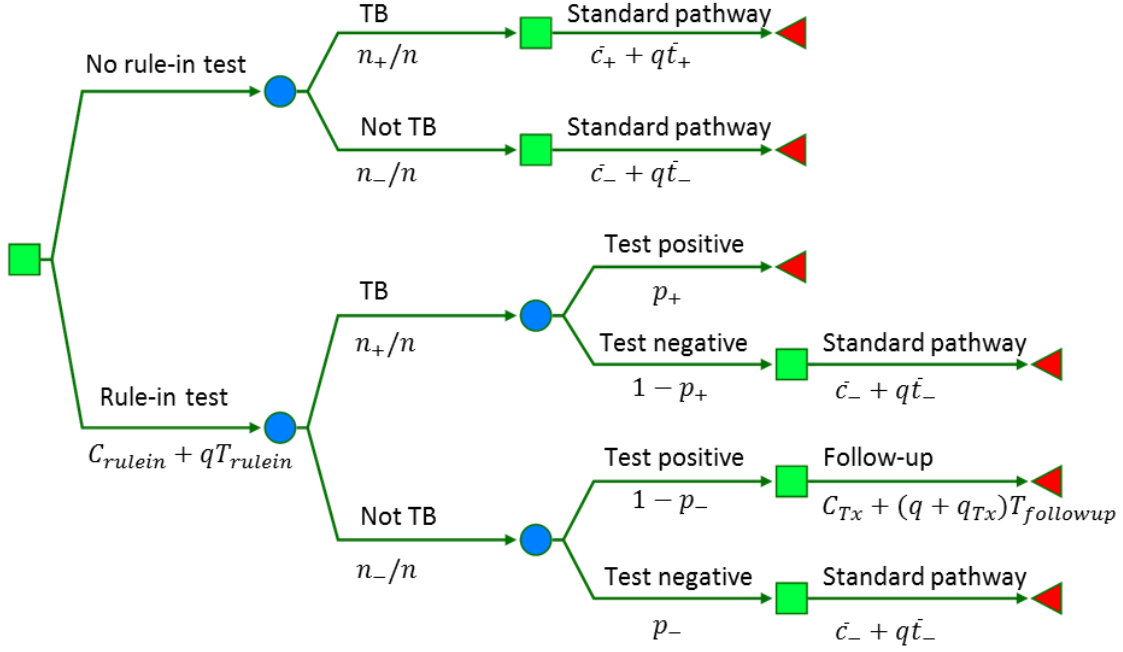


FIGURE 36. Decision tree comparing current practice (“No rule-in test”) with a diagnostic pathway incorporating an initial rule-in test (“Rule-in test”). The “Standard pathway” branch represents the range of variation observed in the patient cohort. Probabilities are shown below branches following a circular chance node and costs are below branches following a square decision node.

Use of dual testing as rule-in test

Cost-effectiveness outcomes

For the patient cohort the cost and health detriment were defined as

$$\begin{aligned}
 C_{novel} = & C_{rulein} + \left(\frac{n_+}{n}\right)(c_{IGRA} + \bar{c}_+)(p_+^{IGRA}(1 - p_+) + (1 - p_+^{IGRA})) \\
 & + \left(\frac{n_-}{n}\right)((1 - p_-^{IGRA})(1 - p_-)(c_{IGRA} + C_{Tx}) + (c_{IGRA} + \bar{c}_-)((1 - p_-^{IGRA})p_- \\
 & + p_-^{IGRA})) \\
 E_{novel} = & qT_{rulein} + \left(\frac{n_+}{n}\right)q\bar{t}_-(p_+^{IGRA}(1 - p_+) + (1 - p_+^{IGRA})) + \\
 & \left(\frac{n_-}{n}\right)(q\bar{t}_-((1 - p_-^{IGRA})p_- + p_-^{IGRA}) + (1 - p_-^{IGRA})(1 - p_-)(q + q_{Tx})T_{followup})
 \end{aligned}$$

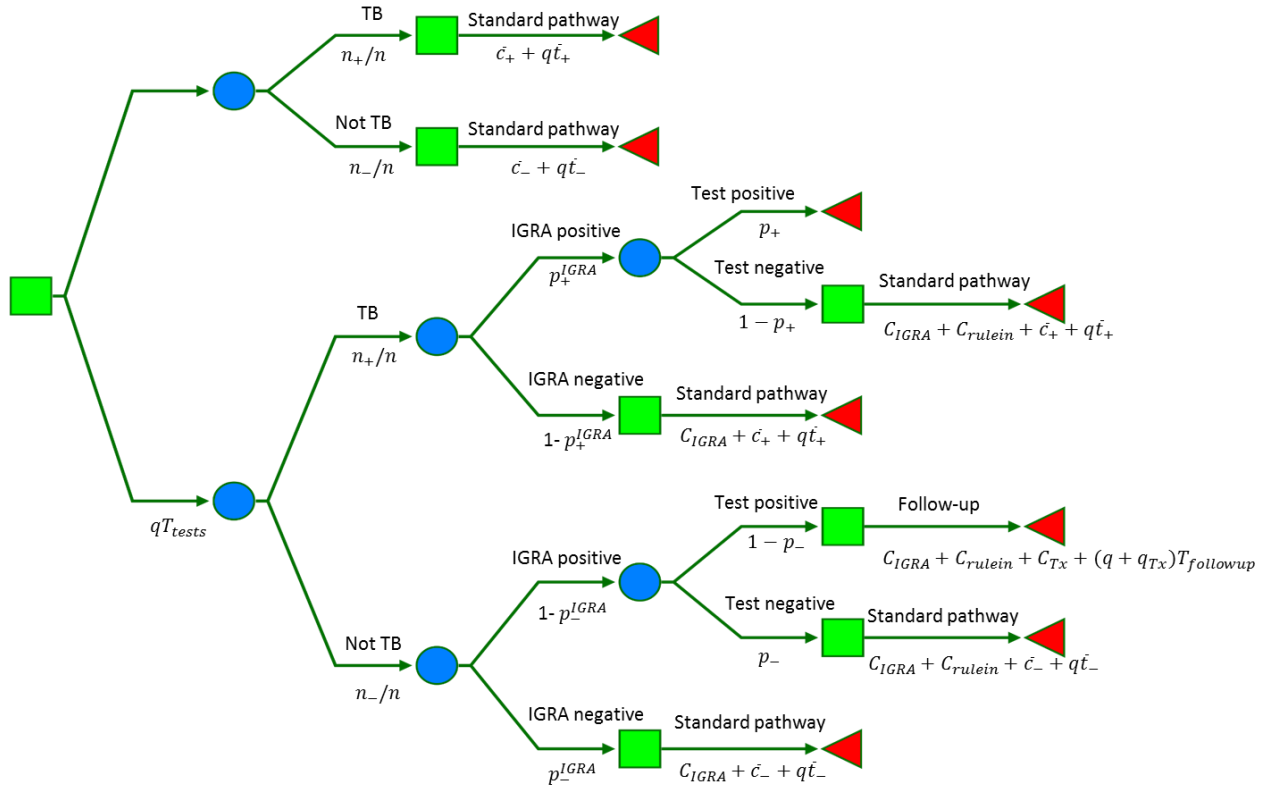


FIGURE 37. Decision tree comparing current practice (“No rule-in test”) with a diagnostic pathway incorporating an initial dual rule-in test (“Dual rule-in test”). The “Standard pathway” branch represents the range of variation observed in the patient cohort. Probabilities are shown below branches following a circular chance node and costs are below branches following a square decision node.