Aus dem Forschungszentrum Borstel
Leibniz-Lungenzentrum
Forschergruppe Klinische Infektiologie
Lehrstuhl für Respiratory Medicine & International Health
Universität zu Lübeck
Prof. Dr. med. Dr. h.c. C. Lange

TB22:

A transcriptomic 22-gene model to individualize treatment durations for patients with tuberculosis

Inauguraldissertation
zur Erlangung der Doktorwürde
der Universität zu Lübeck
– aus der Sektion Medizin –
vorgelegt von

Maja Reimann
aus
Rostock
Matrikelnummer: 72160

1. Berichterstatter/Berichterstatterin: Prof. Dr. med. Dr. h.c. Christoph Lange

2. Berichterstatter/Berichterstatterin: Prof. Dr. med. Jan Rupp

Tag der mündlichen Prüfung: 17.08.2022

Zum Druck genehmigt. Lübeck, den 17.08.2022

Promotionskommission der Sektion Medizin

Table of content

Abbreviation index

AIC	Akaike information criteria
CC	Culture conversion
DS-TB	Drug-susceptible tuberculosis
EOT	End of therapy
GIC	German identification cohort
GLM	Generalized linear model
GVC	German validation cohort
HIV	Human immunodeficiency virus
LASSO	Least absolute shrinkage and selection operator
IGRA	Interferon-γ release assay
MDR	Multidrug-resistant
MNI	Marius-Nasta-Institute
M. africanum	*Mycobacterium africanum*
M. bovis	*Mycobacterium bovis*
M. tuberculosis	*Mycobacterium tuberculosis*
PCA	Principal component analysis
PCR	Polymerase chain reaction
PPV	Positive predictive value
RMSE	Root mean square error
RVC	Romanian validation cohort
SGVC	Second German validation cohort
TB	Tuberculosis
TB22	22-gene transcriptomic signature model
TCC	Time to culture conversion
tmod	Transcriptional modules
TOS	Therapy outcome score
TPP	Target product profile
TPS	Therapy progression score
TTP+	Time to culture positivity
TNF	Tumour necrosis factor
WHO	World Health Organization
XDR	Extensively drug-resistant tuberculosis

Graphic Index

Tables

Supplement tables

Formulas

Zusammenfassung

Die Tuberkulose ist die häufigste zum Tod führende bakterielle Infektionskrankheit weltweit. Jährlich erkranken etwa 10 Millionen Menschen an einer Tuberkulose; 500.000 davon an einer multiresistenten Tuberkulose (MDR-TB) bei der die Bakterien gegenüber Rifampicin und Isoniazid resistent sind. Die MDR-TB ist mit einer langen Therapiezeit, hohen Therapiekosten und einer hohen Rate an unerwünschten Arzneimittelreaktionen verbunden. Durch eine standardisierte Therapiedauer erfahren viele Pateinten eine Überbehandlung. Es fehlt ein Biomarker der es erlaubt, die Dauer der Therapie zu individualisieren.

Ziel dieser Arbeit war die Entwicklung einer RNA-Signatur aus dem Blut der Patienten, die mit hoher Sicherheit eine Bakterien-freie Diagnose der Tuberkulose erlaubt und die sich als Biomarker zur Steuerung der Therapiedauer eignet. Hierfür wurden Tuberkulosepatienten in insgesamt sechs Kohorten prospektiv rekrutiert. Den Patienten wurde zu verschiedenen Zeitpunkten Blut für RNA Analysen abgenommen. In zwei Identifizierungskohorten wurde ein mehrstufiger Algorithmus zur Ermittlung des Therapieendes entwickelt, dieser wurde in vier Validierungskohorten überprüft. Ein weiterer Vergleich mit gesunden Probanden und Patienten mit anderen Erkrankungen diente der Überprüfung der diagnostischen Genauigkeit der Gensignatur.

In den Identifizierungskohorten wurde ein 22-Gen RNA Signatur Modell (TB22) entwickelt. Für die Unterscheidung zwischen kulturell bestätigter Tuberkulose und anderen Krankheiten ist die diagnostische Genauigkeit von TB22 mit den besten bislang publizierten RNA Signaturen vergleichbar (area under the curve [AUC] 0.89; CI: 0.82-0.96). In den Validierungskohorten bei Patienten mit sensibler Tuberkulose hat TB22 eine Genauigkeit für die Unterscheidung einer laufenden und erfolgreich beendeten Therapie mit einer AUC von 0.94; 95%-CI: 0.90 – 0.98). In vier Validierungskohorten mit MDR-TB lag die mittlere Therapiereduktion durch Anwendung von TB22 bei 218.0 Tagen, 34.2%, p<0.001, 211.0 Tagen, 32.9%, p<0.001, 161.0 Tagen, 23.4%, p=0.001 und 254 Tagen, 40.0%, p=0<0.001.

Im Vergleich mit 17 publizierten RNA Signaturen zeigte TB22 die größte Genauigkeit zur Unterscheidung zwischen laufender und erfolgreich beendeter Therapie (p-Wert: <0.001 – 0.070).

In dieser Arbeit wurde eine auf 22-Genen basierende RNA-Signatur (TB22) aus dem Blut von Tuberkulosepatienten identifiziert und validiert, die mit hoher diagnostischer Genauigkeit eine Bakterien-freie Diagnose der Tuberkulose erlaubt und die zum Therapiemonitoring der MDR-TB geeignet erscheint. Die klinische Anwendung von TB22 kann perspektivisch das Management der MDR-TB verändern.

Summary

Tuberculosis is the most common fatal bacterial infectious disease worldwide. Every year, about 10 million people contract tuberculosis; 500,000 of them have multidrug-resistant tuberculosis (MDR-TB) in which the bacteria are resistant to rifampicin and isoniazid. MDR-TB is associated with a long duration of therapy, high therapy costs, and a high rate of adverse drug reactions.

Due to a standardized duration of therapy, many patients experience overtreatment. There is a lack of a biomarker that allows individualizing the duration of therapy.

The aim of this work was the development of a RNA signature from the blood of patients, which allows with high certainty a bacteria-free diagnosis of tuberculosis and which is suitable as a biomarker to control the duration of therapy. For this purpose, tuberculosis patients were prospectively recruited in a total of six cohorts. Blood was drawn from the patients at different time points for RNA analyses. In two identification cohorts, a multistage algorithm was developed to determine the end of therapy, and this was tested in four validation cohorts. Further comparison with healthy subjects and patients with other diseases was used to verify the diagnostic accuracy of the gene signature.

A 22-gene RNA signature model (TB22) was developed in the identification cohorts. For discriminating between culturally confirmed tuberculosis and other diseases, the diagnostic accuracy of TB22 is comparable to the best RNA signatures published to date (area under the curve [AUC] 0.89; CI: 0.82-0.96). In validation cohorts in patients with drug-susceptible tuberculosis, TB22 has accuracy for discriminating ongoing and successfully completed therapy of AUC 0.94; 95%-CI: 0.90 – 0.98). In four validation cohorts with MDR-TB, the median treatment reduction with use of TB22 was 218.0 days, 34.2%, $p<0.001$; 211.0 days, 32.9%, $p<0.001$; 161.0 days, 23.4%, $p=0.001$; and 254 days, 40.0%, $p<0.001$.

Compared with 17 published RNA signatures, TB22 showed the greatest accuracy for discriminating between ongoing and successfully terminated therapy (p-value: <0.001 – 0.070)

In this work, a 22-gene based RNA signature (TB22) was identified and validated from the blood of tuberculosis patients, which allows bacteria-free diagnosis of tuberculosis with high diagnostic accuracy and appears suitable for therapy monitoring of MDR-TB. The clinical application of TB22 may perspectively change the management of MDR-TB.

1 Introduction

1.1 Background

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis* (*M. tuberculosis*) [1]. In the pre-COVID-19 era of the past decade, TB was the leading cause of mortality by a single infectious agent world-wide [2]. The World Health Organization (WHO) estimates that in the year 2019 approximately 10.0 million individuals developed TB [3, 4].

An estimated half million TB patients are infected with a multidrug-resistant (MDR) strain of *M. tuberculosis* [3, 4]. Multidrug resistance in TB is defined by a bacillary strain with resistance against isoniazid and rifampicin, the two most active anti-TB drugs [5].

The numbers of patients affected by MDR-TB have increased dramatically in the new century and patients are unequally distributed worldwide [6, 7]. The therapeutic outcome in the presence of MDR strains of *M. tuberculosis* highly depends on the extent of drug-resistance and the anti-TB medicines being available [8]. While overall 85 % of patients affected by TB achieve a favourable treatment outcome, only 57% of patients with MDR-TB can currently be treated successfully [2].

Patients affected by TB with a drug-susceptible strain of *M. tuberculosis* usually receive a standard therapy consisting of isoniazid, rifampicin, ethambutol and pyrazinamide for 2 months, followed by therapy with isoniazid and rifampicin for four-month. Results of a systematic review of individual patient data suggest that patients with MDR-TB have an optimal treatment outcome when they are treated with at least 4 active medicines over a minimum of 18 months [8], although results from recent studies show that high cure rates can be achieved with much shorter treatment durations when potent therapeutic regimens are administered [9-13]. Treatment of MDR-TB is not only longer than treatment for drug-susceptible TB, it is also much more expensive [14] and causes a high rate of adverse drug events [15].

The duration of anti-TB therapies could be substantially shortened in many patients with MDR-TB if a biomarker would be available that can safely indicate that treatment can be terminated without an increased risk of relapse. So far, microbiological parameters are not suitable as such a biomarker. The majority of patients with MDR-TB achieve *M. tuberculosis* culture conversion within 2 months of anti-TB drug therapy [16], but discontinuation of anti-TB treatment at this point causes a high rate of relapse [8].

Recently, there have been attempts to adapt the duration of therapy to the severity of the disease and the drug resistance profile of *M. tuberculosis* [16-18] but until now no biomarker has been identified and validated to guide physicians in the decision when to discontinue MDR-TB therapy [19]. Furthermore, a suitable therapy monitoring system has not been identified to

quickly obtain precise feedback regarding the progress of disease or the therapy response. This thesis addresses the development and validation of a biomarker for treatment monitoring and to individualize the duration of therapy for patients affected by MDR-TB.

1.2 The pathogen *Mycobacterium tuberculosis*

Hippocrates (~460 - 375 B.C.) used the term phthisis (consumption) for a disease associated with physical decay [20]. In 1689 Thomas Morton specified the term as a lung disease and characteristic lung lesions were named tubercle, from which Lukas Schönlein derived the term *tuberculosis* in 1832 [21].

In 1882, Robert Koch discovered that a micro-organism, *M. tuberculosis,* is the cause of TB [22]. Mycobacteria belong to the *Actinobacteria* [23]. *Mycobacterium* is the only class of the family *Mycobacteriaceae*. *M. tuberculosis*, *M. bovis*, *M. africanum, M. canetti, M. caprae, M. microti, M. mungi and M. pinnepedii* are representatives of this genus, which form the *M. tuberculosis*-complex due to their DNA homology of >95% [21, 24].They are non-spore producing rod bacteria [25]. Infections with *M. tuberculosis* strains are most common [26].

Following staining with carbol fuchsin the thick mycobacterial cell wall resists decolorizing by acid-alcohol and the mycobacteria appear as bright red rods under the light microscope.

The overall growth rate of *M. tuberculosis* is slow. Under optimal conditions, a division process takes 12-24 hours during active disease and no to little growth during the latent phase [27].

1.3 Epidemiology

It is estimated that ¼ of the world´s population is infected with *M. tuberculosis* [28], but the great majority of patients who is infected probably never develops active TB [29]. More than one quarter of all patients with TB world-wide live in India. More than one-third of affected patients live in India and China, the most populous countries in the world. More than two thirds of patients affected by TB live in 8 countries [6].

In Europe, the incidence of TB has been declining for decades. The WHO Office for Europe and the European Centre for Disease Control and Prevention (ECDC) report an incidence of 26/100,000 for the WHO Europe region in 2018 (compare: global incidence: 130/100,000) [6, 30]. However, the proportion of patients with MDR-TB is comparatively high here, and this is increasingly problematic, especially in the Eastern European countries [6, 30]. In Germany, since the introduction of mandatory electronic reporting in 2001, the number of TB cases also declined steadily, except for a slight rebound in 2015. In 2019, 4782 TB cases were reported, and 129 patients died [31, 32].

In some developed countries, there is a trend toward increased extra-pulmonary and disseminated TB cases [31]. In Germany, extra-pulmonary TB accounted for approximately 28% of all cases in 2019, with lymph node TB accounting for nearly half of those cases [33, 34]. The proportion of extra-pulmonary cases has also increased in Germany in recent years. In 2015, the proportion of extra-pulmonary cases was only 23% [35].

1.4 The disease tuberculosis

The disease TB is an airborne disease that is transmitted via aero droplets from persons with active TB. In case of infection via inhalation, the bacteria can still be killed by phagocytosis of alveolar macrophages or neutrophils [36]. If this does not succeed initially, they can survive in within dendritic cells and alveolar macrophages. In those who progress to active TB following primary infection, the onset of TB may occur between weeks and decades after primary infection. In immunosuppressed individuals, the risk of developing active TB is significantly increased and the onset period is shorter, than in immunocompetent individuals [37].

Latent infection with *M. tuberculosis* (LTBI) is asymptomatic [38]. LTBI is diagnosed by a positive reaction in the tuberculin skin test or interferon-γ release assays (IGRA) in the absence of active TB. Neither test is suitable for predicting the risk of active TB accurately [39]. When TB transitions to the active and contagious phase, symptoms appear usually insidiously. In children and immunosuppressed patients, however, the disease can also progress much more rapidly. The transition from latent to active TB resembles a continuum of bacterial and immunologic activity[40] . Therefore, in addition to latent and active TB, it is also appropriate to speak of incipient and subclinical TB [41]. While incipient TB is characterized by the fact that specific metabolic activity of the pathogen is measurable, in subclinical TB radiological or microbiological signs are already evident [41]. The symptoms of extra-pulmonary TB vary and depend largely on the organ system affected. If two or more organs are affected, it is called disseminated TB [31].

1.5 Clinical manifestations

Regarding the continuous spectrum, symptoms also appear gradually and with increasing intensity [42]. Symptoms for pulmonary TB include cough, night sweats, weight loss, and fatigue. Rarely haemoptysis occurs [31]. Pulmonary TB can be divided into parenchymal disease, endobronchial TB, and intrathoracic lymph node TB, which are similar overall but differ primarily in cough productivity [42].

Symptoms of extra-pulmonary TB are usually accompanied by similar nonspecific symptoms. Organ-specific symptoms can include painful lymph node swelling, joint pain or headache. Spinal

TB is additionally associated with spondylitis [43], and central nervous system involvement leads to TB meningitis and/or cerebral abscess formation [44]. In tuberculous pericarditis shortness of breath or cardiac arrhythmias may occur [45]. Urogenital TB is often asymptomatic besides occasional flank pain [42].

1.6 Tuberculosis diagnosis

The detection of *M. tuberculosis* is performed by laboratory diagnostics. For pulmonary affection, sputum is the most important sample type for the diagnosis. With Ziehl-Neelsen light microscopy, the specificity of TB is high, but the sensitivity shows a varying range of 20%-80% only [46]. Conventional fluorescence microscopy shows a higher sensitivity, but is less feasible in some settings due to higher costs and higher equipment requirements [46, 47]. Smear microscopy offers fast detection, but the sensitivity and specificity are poor. A polymerase chain reaction (PCR) test can also be used in the diagnosis. Here, the overall sensitivity of GeneXpert PCR test is 69.4% with a specificity of 98.8% in HIV-negative people [48].

Imaging techniques are often used to diagnose pulmonary TB. Typical changes are tree-in-bud signs and cavities [49]. If the findings are unclear, the x-ray can also be extended to include CT scans or MRI procedures in case of extra-pulmonary TB. Radiology provides important information, especially for diagnosis, since cultural results sometimes take longer to arrive. In addition, radiological findings give an indication of the severity of the disease with regard to the classification of pulmonary TB [50], further radiological changes may show progression to active TB. However, the same radiological changes are also an indication of the response to therapy, since the size and structure of the cavities would have to decrease in the course of therapy [51]. PET/CT scans can be used to determine the organ involvement for disseminated or extra-pulmonary TB [52].

Mycobacterial culture is the gold standard for the diagnosis of TB, but has a long turn-around time of up to 2 months. Culture-based test methods are not suitable for all patients because some cannot produce sputum or the sputum samples do not become positive [53].

Standard anti-TB drug-resistance testing is performed using culture-based susceptibility testing. However, genotypic methods allow for a rapid prediction of phenotypic drug susceptibility testing results[54]. Phenotypic and genotypic methods can be used complementarily to ensure the highest possible sensitivity and specificity [55].

1.7 Tuberculosis treatment

Following the discovery of streptomycin in 1947 combination antibiotic therapy against TB rapidly became available in the following years with the discovery of other anti-TB medicines [56-58]. This was also necessary because soon after the first drug was applied to humans affected by tuberculosis it became apparent that TB treatment with a single-agent led to drug resistance and relapse [58]. Today, patients with drug-susceptible (DS)-TB are treated with a four-drug combination of rifampicin, ethambutol, pyrazinamide and isoniazid for two months, followed by a four-month treatment phase consisting of rifampicin and isoniazid. This combination was already introduced in the 1970 [58, 59].

In the recent past, more effort has been put into the re-design of optimal treatment regimens, which may also result in shortened treatment durations [58]. Furthermore, new methods allow for the analysis and optimisation of standard doses [60]. The treatment of extra-pulmonary TB is similar to the treatment of pulmonary TB. In case of DS-TB that affects bone tissue treatment should be prolonged to 9 instead 6 months [59]. Also, TB meningitis requires longer therapy duration and should be treated for at least 12 months [19].

In contrast to standardized therapy for DS-TB patients, the therapy regimen for patients with MDR-TB is adapted to the *M. tuberculosis* drug resistance present in each case [2]. However, as the necessary equipment for drug susceptibility testing is not universally available, the choice of drugs is often based on the drug resistance patterns typical of the respective geographical location, previously taken drugs for a previous TB, comorbidities and the expected adverse effects of the therapy [58]. This standardized treatment approach might be one reason, why therapy outcomes show differences between both patients affected by DS-TB and MDR-TB, although there is only a limited difference in disease severity [61, 62].

The standard therapy duration for MDR-TB is 18-20 months depending on the disease severity [2, 8]. According to the WHO treatment guide for MDR-TB (recently expanded also to rifampicin-resistant TB without resistance to isoniazid; RR-TB), the available drugs have been divided into groups according to which the treatment regimen is composed. Group A contains fluoroquinolones, bedaquiline and linezolid. It is recommended to use all three drugs. In addition, one or both drugs of group B (clofazimine and cycloserine or terizidone) should be added to the therapy regime. Drug group C includes drugs that should be used if there is resistance to antibiotics from groups A and B. Classified in group C are ethambutol, pyrazinamide, delamanid, imipenem or meropenem, amikacin, ethionamide or prothionamide, and p-amino salicylic acid. Other ungrouped medications are also available, such as kanamycin [2].

Baseline smear grade, cavities, adherence and HIV status are associated with treatment outcome. These risk factors were used to construct a patient classification system for shortening therapy in DS-TB patients, which has so far proved to be non-inferior to the classic six-month approach [63]. Due to immunological and infection-susceptible host variability and pharmacokinetic aspects, an individually required duration of therapy can be assumed for relapse-free cure in TB treatment [64]. This could be achieved with a biological, objective and validated surrogate marker, which ideally has diagnostic, prognostic and therapy-accompanying properties [65]. The diagnostic biomarkers like sputum and radiology described in Section 1.5. do not have the necessary accuracy for personalized therapy duration [65]. Currently, no biomarker-guided therapy durations are available [64].

1.8 Biomarker for tuberculosis

A biomarker is a quantifiable, measurable and replicable clinical parameter that provides information on biological, pathological or pharmacological processes in the body and characterizes a diagnostic or prognostic statement or influence on the course of disease and treatment [66]. The quality of a biomarker is often described by the area under the curve (AUC) of a receiver operating characteristic (ROC) curve. The AUC of a ROC curve shows the proportion of cases correctly classified according to a clinical parameter or model [67]. This analysis also allows conclusions to be drawn about sensitivity and specificity. While sensitivity describes the proportion of individuals who are correctly classified by the biomarker as Y=1 (e.g., diseased), specificity measures the proportion of individuals who are correctly identified as Y=0 (e.g., not diseased) [68].

Supplement Table 1 provides an overview about promising biomarkers for monitoring of TB treatment. The gold standard of therapy monitoring and outcome prediction in TB is culture data and microscopy results. A combination of microscopy and culture data is recommended because the outcome prediction for both methods show sensitivity below 81 % and specificity below 60%, respectively. While the sensitivity of both procedures rises to over 90% as the therapy progresses, the specificity decreases continuously [69]. Even though the sputum smear microscopy grade has a lower sensitivity and specificity, this marker is still well suited as a control instrument after culture conversion has been achieved. A persistently high sputum smear microscopy grade and positive cultures weeks after initiation of therapy indicate a reduced response to antibiotic treatment [31, 69, 70]. A faster alternative for culture positivity could be the Molecular Bacterial Load assay (MBLA). Its result is available after few hours, but is also only available if positive cultures are present [71]. Radiological changes can also provide information about the response

to therapy. Lesions become smaller or disappear completely in the course of therapy. Newly appeared lesions may give an indication of deterioration or TB reactivation. Unchanged radiological findings beyond six months indicate inactive disease [50]. PET/CT imaging also provides a better correlation with therapy response than sputum smear microscopy grade, especially in the first two weeks of therapy. It is also suitable for extra-pulmonary TB and HIV infected TB patients [72], however 85% of patients with DS-TB who are successfully treated still have enhanced PET signals at the end of therapy [73] and the method is unlikely to be implemented universally. Sputum smear microscopy grade, culture conversion and radiological findings show a correlation with each other [70, 74].

In addition to identification in sputum, bacterial DNA can also be detected in urine [75]. However, previous studies indicated that while DNA is readily detectable, particularly in the first two weeks, no specific patterns were found during this time. Accordingly, early bacterial activity and drug efficacy by detection of transrenal DNA is less readily visualized by this marker than by microscopy, but offers an alternative when sputum is not available, such as in children or extra-pulmonary TB [76].

Another possible biomarker for TB treatment monitoring is lipoarabinomannan (LAM). LAM is a lipoglycan measured from urine samples and allows direct conclusions to be drawn about the virulence of mycobacteria [77]. However, it does provide further information about the bacterial load. However, it must be remembered that LAM concentration varies with different body fluids and continues to be elevated in co-infections such as HIV [78]. For this reason performance of LAM is much better for immunocompromised patients [79]. The LAM assay for sputum examination was shown to correlate with early bacterial activity [80].

Furthermore, there is the possibility to assess the treatment success by means of clinical scores. The TBscore II is based on TB symptoms and clinical parameters such as BMI. It has been shown that an insufficient reduction of the score during the first four months of therapy is associated with a poor therapy outcome [81]. The DZIF clinical score also showed a significant decrease, especially in the first two months of therapy, compared to measurements at therapy initiation, both in DS-TB and MDR-TB patients. The Ralph score, which refers to radiological changes, showed a significant decrease in the six months of therapy [82].

The concentration of C-reactive protein (CRP) can also provide information about inflammatory processes, but is not sufficiently sensitive or specific [83]. But it is in the early treatment phase that conclusions can be drawn about treatment response and adverse outcomes [84]. Insufficient sensitivity and specificity also apply to interferon γ-induced protein 10 (IP-10). Alpha-1-acid glycoprotein (AGP) has a good sensitive and specific value, at least in diagnostics [83]. Especially in fully drug-susceptible TB, Cytokines might be promising as monitoring tool [85]. Cytokines also

indicate increased disease severity and show a correlation with bacterial load [86]. Overall, cytokines may also play an important role in monitoring and treatment outcome in extra-pulmonary TB [87]. Chemokines can also serve as biomarkers of disease severity and bacterial load and also correlate with delayed culture conversion [88]. The TAM TB assay represents an immunological monitoring tool. During the course of therapy, the patient's pattern of T cell activation shifts more and more towards LTBI/healthy controls [89].

With respect to biomarkers at the OMICS level, proteomics, metabolomics, and transcriptomics should be mentioned above all. Proteomics can be host-related as well as bacterium-related. *M. tuberculosis*-specific peptides also paint a picture of treatment response and could be used as clearance marker, but the methodology using mass spectrometry is costly and cannot be implemented comprehensively [90, 91]. Furthermore, host serum proteomic signature may correlate with 8-week culture status with high sensitivity and specificity [92]. As example for metabolomics biomarker, SLC1G level in urine correlate with different treatment response outcomes. But the sample size of this study was low and an external evaluation is missing so far [93]. There is also evidence that the host metabolomics profile changes, especially during the acute TB treatment phase for *M. tuberculosis* but also for other mycobacteria [94-96].

Total DNA sequencing has given an important impetus to RNA-based research. In recent years, several RNA profiles and signatures have emerged for different diseases. Several RNA-signatures have been identified that have therapeutic and diagnostic significance [97]. RNA enables not only the understanding of the genome, but also an in-depth comprehension of anatomical and pathological processes due to the different regulations and signaling pathway cascades [97].

The cell nucleus contains all the necessary genetic information with the DNA. A gene is a DNA section that contains specific information. RNA is transcribed in the nucleus. In contrast to DNA, RNA is single-stranded and significantly shorter and reduced to only one gene. Not all cells express RNA in the same way, but each cell has a specific expressing repertoire depending on its function [98]. RNA has several roles; for example, as messenger RNA (mRNA), it serves as a building block for protein biosynthesis; is necessary as transfer RNA (tRNA) to generate the amino acid sequence from nucleic acid segments. In addition, ribosomal RNA (rRNA) as a component of ribosomes is essential for peptide binding and also has catalytic functions [98]. In addition, there are a large number of RNA variants that are not coding but have regulatory functions. In addition, there are still a large number of RNA variants that are non-coding but have regulatory roles; microRNA and long noncoding RNA (lncRNA) have recently been studied more frequently in relation to disease patterns [99, 100]. RNA can be measured using a variety of methods - the most common are microarrays and RNAseq for a comprehensive screen or RT-qPCR for the measurement of specific gene expressions [101]. Microarrays have been more and more replaced

by RNAseq technology in recent years due to better accuracy after the method became less expensive than in the past [102]. The idea behind both methods is to quantify RNA concentrations. Microarray measurements are based on nucleic acid probes, typically 60-mers, covalently bound to glass slides. There, they are scanned by fluorescently labelled target sequences and hybridized. The resulting images are translated into signal intensities, which can then be bioinformatically processed [102]. The processing of blood samples in the context of RNA measurement is relatively inexpensive and efficient if not total RNA, but only a few genes determined within a signature are measured [103].

A transcriptome is defined as the whole set of transcribed RNA by genome of a specific cell type or a specific tissue with regard to specific physiological condition [104]. Gene expressions or transcriptomics are commonly used to generate profiles based on momentary snapshots of biological processes and establish a bridge from genotype to phenotype [98, 105]. These genetic profiles are then compared with respect to treatment measures or specific phenotypic features [105]. Transcriptome data are semi-quantitative. For this reason, a normalization procedure must be applied [106]. A distinction can also be made between pathogen and host transcriptomes at the transcriptome level. At the pathogen level, bacterial rRNA in particular was shown to correlate with the number of live mycobacteria and may be suitable as a therapy monitoring system [71, 107, 108].

Supplement table 2 gives an overview about host transcriptomic signatures and models that were described in literature for diagnostic, monitoring and outcome prediction purposes.

In a systematic review, eight RNA signatures with sensitivities ranging from 47.1 to 81.0% and specificities above 90%, showed the onset of active TB 0-3 months prior to disease. This fulfils the WHO target product profiles (TPP) criteria at least in part [109]. Goals and needs for biomarkers in clinical usage were defined by WHO in 2014 as TPP. These include triage/ systematic screening to identify active and latent cases; rapid TB-diagnostic sputum-based / non-sputum based; Next generation drug-susceptibility; Treatment-monitoring and distinction between LTBI and active cases [79]. However, the longer the signatures were applied before disease onset, the lower the accuracy became [109]. An 11-gene version derived from a 16-gene signature [110] showed a specificity of 84% and a sensitivity of 71% for predicting TB up to one year before disease onset [111]. In a prospective observational study, four transcriptional signatures were able to distinguish individuals with and without TB (sensitivity: 83.3 to 90.7%) [112]. A score consisting of three genes was described to predict progression from latent to active disease and could also monitor treatment progress and correlated with pathological patterns exhibited by patients at high risk for treatment failure [113, 114]. Another study was able to diagnose patients with early TB from TB-free individuals with an A of 0.86 even before symptoms developed, but contains 86

genes [115]. Recently, a meta-analysis described a 22-gene signature with an AUC of 0.86 to distinguish active TB cases with and without culture backup from patients with other lung diseases [116]. In comparison with other signatures tested there, it achieves the best performance. Nevertheless, it did not meet the WHO requirements for TPP either [116].The RISK6 score, based on a six-gene signature, was able to identify individuals with latent TB who were at risk of disease progression (AUC 0.87 within 0-12 months and AUC 0.74 within 12-24 months before diagnosis of TB) [117]. RISK6 also showed a continuous decrease in the gene signature-based score during treatment [117].

The signatures and scores described earlier are primarily concerned with the risk of relapse in patients with and without HIV-infection and continue to show a difference especially in the early treatment phase [111, 117]. A previous described signature that was also used for diagnostic purpose and contains 16 genes [110] focused in its validation phase primarily on outcome prediction, which became more accurate as therapy progressed. After four weeks of therapy in DS-TB patients, an AUC of 0.72 was given for the discriminative ability between cure and failure [118]. Interestingly, it was shown that a five gene signature out of the previous 16 was related to the inflammatory status, which could be derived from the results of PET/CT measurements in the study [118]. Recently, a model was also described that also consists of transcriptomes and clinical parameters and can predict outcome as early as four weeks and correlates with inflammatory status, which could also allow conclusions to be drawn about treatment progression [119]. Another signature also combining clinical and transcriptomic parameters to get an AUC of 0.72 for identifying relapse patients even at the beginning of therapy[120].

In principle, the end of therapy as such was not considered a variable of interest according to the current state of knowledge, which means that a therapy end biomarker has not yet been described. Furthermore, there was a lack of suitable data sets with continuous measurements to identify a biomarker that could predict the relapse-free end of therapy and serve as a real time treatment monitoring tool [113, 114].

1.9 Bioinformatic and mathematical conditions

In addition to the experimental work to process blood samples and to obtain RNA-expression data, bioinformatics and biostatistical analysis form the second major pillar of transcriptomic evaluation. Data quality, analysis strategies are increasingly critical to analysis success for multidimensional data, such as transcriptomics data [121]. Two main difficulties can be identified as the large amount of data and the high complexity of a transcriptome [104]. Not all statistical

programs can evaluate such data. *Point and click* software is unsuitable in this context. In addition to the methodological deficits, it is often unable to process large amounts of data [122].

Programming language-based analysis programs are necessary for this task. The advantage of the programming language R in the field of bioinformatic analysis is that users can develop packages with functions themselves and make them available to the community [122]. This is especially helpful because professionals know the needs of their work area best [122]. As a result, a broad spectrum of methods has been established around the Bioconductor, which is a platform for tools for analysing high-throughput OMICS data [121, 123]. Through the package documentation, there is still an analysis pipeline, although these are rarely standardized [121, 124].

In addition, *R* is suitable for direct statistical processing, extension possibilities can also be found through the possibility of integrating components from *Python* and *C++* [123, 125].

The disadvantage for the programming language lies in the limitation to be able to work on very large data sets only with high main memory. The technical progress of the last years created this limitation slightly remedy, a high-performance computer and an efficient working package are nevertheless necessary [126].

In **Figure 1**, the bioinformatic and statistical analysis of transcriptome data is shown in compressed schematic form. First, the data is imputed and then a background correction follows. While the foreground measurement measures the intensity, the background measurement determines the noise that must be corrected. The noises result from the technical replication of the genes, which cannot be associated with biological variability [127]. Furthermore, fluorescent signals are meant here, which do not result from any binding and are non-spatial [128]. For this purpose, there are different methods with different degrees of correction [128]. After that, the foreground intensities are corrected so that all arrays have similar distribution bandwidths. This is only necessary for two-colour microarrays. Here, too, there are different normalization methods [105]. Other types of bias, such as the *batch effect*, the amount of RNA, or the *block effect* must also be considered and minimized if necessary [129, 130]. This is possible with the help of further standardization models [131]. To get an initial impression of the data set, it is useful to apply an unsupervised analysis. Firstly, this serves feature selection, secondly, outliers can be identified, and thirdly, this offers the possibility of determining multicollinearity [132-134]. After this, further data sets can be created, which either contain only selected genetic features, but can also contain all genes. Additional clinical, biological or socio-demographic characteristics of the patients to whom samples belong can be entered here for further analysis. It is then possible to perform directed analyses. This can involve the application of statistical methods to test hypotheses, or algorithms can be built that have predictive power on specific clinical or biological questions. However, these algorithms require validation in external data sets [135-137]. While pathway

analyses are important for gain an overview of the data structure, they main related to underlying biological structures and relationships and are less suitable as a data quality control compared to unsupervised statistical analysis [138].

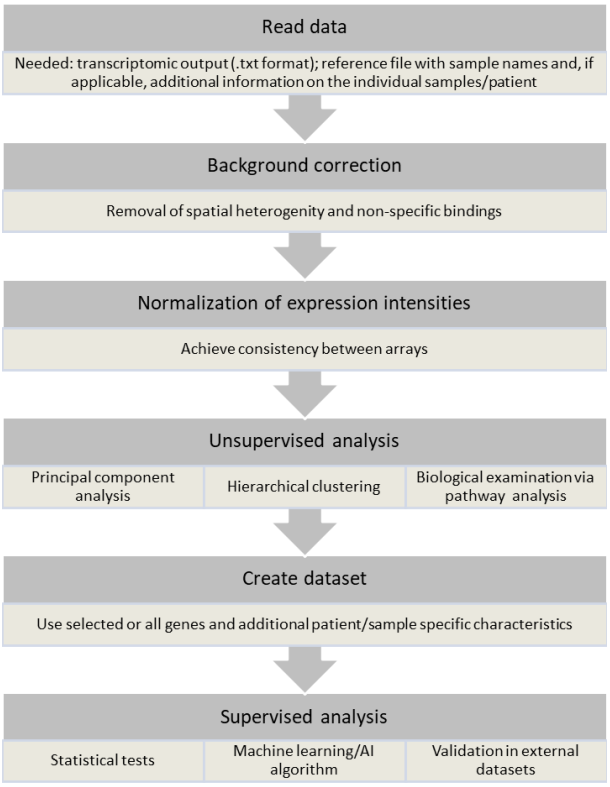

Figure 1 Schematic representation of the bioinformatics and statistical analysis pipeline.

While the preparatory steps follow a relatively strict procedure, the supervised analysis approach offers a variety of possibilities depending on the data set and the research question. On the one hand, statistical tests can be applied that compare the log-fold change of genes with respect to two groups for significant differences. In this way, it can be determined whether certain genes are up- or down-regulated in certain biological or pathological processes [139]. Particularly challenging is that tens of thousands of genes must be examined in a relatively small number of samples. This poses a challenge to classical statistical testing with respect to first and second kind errors and must be countered with the help of correction methods for multiple testing [131, 140].

Longitudinal data, such as those available in cohort studies, continue to pose a particular challenge. On the one hand, the temporal aspect as such increases the dimensions and on the other hand, it makes it difficult to select genes that play a role with regard to both the disease and the changes in disease status over time. Further, cohorts usually consist of a relatively small number of individuals followed over time [131]. If transcriptome analyses are not only used as a

descriptive basis, but have to fulfill a predictive claim, it is necessary that the claim of application on external data sets is taken into account in the model development.

This includes, first and foremost, the existence of independent test data sets to which the model developed in the training data set is applied in order to test performance. Performance is largely dependent on a meaningful choice of variables. In addition to minimizing overfitting, feature selection based on data mining processes has the advantage of reducing the data set, which improves computational power and thus enables more efficient work [131]. Prediction models are based on the principle of parsimony. This means that only as many variables should be included in a model as are necessary to best describing and predicting the dependent variable. Overfitting violates this principle by creating a complexity of the model by unnecessarily many variables. Estimator parameters are assigned to all predictors in a model. The more variables are used to describe the model, the less the influence of each parameter is. The artificial valorization of the influence of random variables without actual effect consequently also leads to the devaluation of the effect of actually relevant variables [141].

1.10 Thesis aim

In this thesis, bioinformatics and biostatistical methods were applied to identify a human transcriptional signature from peripheral blood to distinguish between TB patients and healthy controls. This signature was then applied to develop a score using statistical models to represent treatment responses and to identify individual endpoints of anti-TB therapy. To achieve this, patients with DS-TB and MDR-TB were prospectively recruited in a multi-centre trial with six different cohorts, two of which served as the training data set and four as the test data set for validation. Study visits with blood collection took place at individual and fixed time points, which were used after laboratory and bioinformatic preparation to find a transcriptome signature suitable for the above objectives.

2 Methods

2.1 Study design and participants

Between March 2013 and March 2016, patients with culture-confirmed DS pulmonary tuberculosis and MDR-TB identified by detection of *M. tuberculosis* DNA from sputum by the Xpert MTB/RIF test (Cepheid, Sunnyvale, USA) were prospectively enrolled into the drug-susceptible German identification cohort (DS-GIC) and the multidrug-resistant German identification cohort (MDR-GIC), at the Medical Clinic, Research Centre Borstel; Karl-Hansen-Klinik, Bad Lippspringe; Sankt Katharinen-Krankenhaus, Frankfurt; Thoraxklinik-Heidelberg, Heidelberg; Asklepios Fachkliniken München-Gauting, Munich, all in Germany, as previously described [142]. Between March 2015 and April 2018, patients with DS-TB and MDR-TB were prospectively enrolled into the DS-GVC and MDR-GVC at the same centres and additionally at the Klinikum Dortmund, Germany, and at the University Clinic of Cologne, Germany. Between May 2015 and March 2017, patients with MDR-TB were prospectively enrolled into the MDR-RVC at the Marius-Nasta-Institute (MNI) in Bucharest, Romania. Another cohort of MD-TB patients were enrolled between June 2017 and June 2019 at Research Centre Borstel (MDR-SGVC). Blood sampling followed the same scheme than for GVC. To validate the signature and the hypothetical therapy endpoints calculated retrospectively, additional clinical data were taken into account. A model that can indicate the end of therapy should not at any time confirm the successful end of therapy to a patient with a positive culture [143].

Additional data sources were used to address clinical differences in patients. First, the clinical database for patients hospitalized with TB at the Medical Clinic Borstel that contains information about comorbidities, behavioural aspects like smoking or problematic alcohol consumption, TB-related information like imaging results and microbiological results like time to positivity and time to culture conversion for GICs and GVCs besides others. This database contains also patient characteristic information about enrolled patients from other German study sites. For patients from the RVC, patient characteristic information as well as microbiological data was given [143].

Individuals were not included to the study, if they were less than 18 years of age, under legal supervision, or living with human immunodeficiency virus (HIV).

Between June 2015 and December 2015, adult HCs with no history of previous tuberculosis and without any known concurrent illnesses at the timepoint of blood sampling were enrolled at the Medical Clinic of the Research Centre Borstel (Germany).

Study visits included clinical assessment and blood sampling for whole-blood RNA measurements from PAXgene tubes (Qiagen®, Venlo, Netherlands). Study visits were performed at ideally before

treatment initiation, at 14 days of therapy, at the times of smear conversion and following culture conversion (not available in the MDR-RVC), at 6 months and/or therapy end in patients with DS-TB, and additionally, at 10, 15, 20 months of therapy in patients with MDR-TB. After completion of 4 weeks of therapy, an additional study visit was performed in patients from the MDR-RVC (see **Figure 2**). All patients completed 12 months of evaluation following end-of-therapy to capture disease recurrence. A subset of DS/MDR-GVC participants also provided specimens during this follow-up period. Sputum samples provided by German study participants were evaluated via smear microscopy and culture at the National Reference Centre for Mycobacteria at the Research Centre Borstel. Samples provided by study participants at the MNI were analysed at the Romanian national reference centre for mycobacteria in Bucharest. Anti-tuberculosis therapy regimens were based on comprehensive drug-susceptibility testing and consistent with current therapy recommendations [59, 143-145].

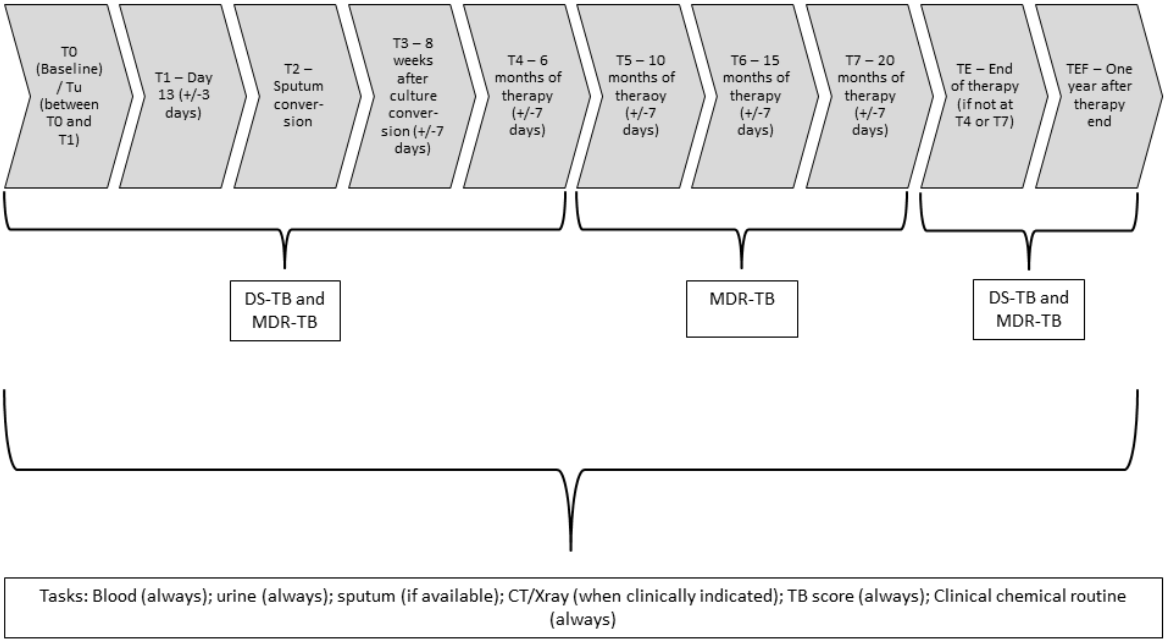

Figure 2 Study visits and tasks for drug-susceptible TB (DS-TB) patients and multidrug-resistant TB (MDR-TB) patients.

2.2 Definition of treatment outcomes

Treatment outcomes were assessed following the TBNET definitions, where relapse-free cure is defined by having a negative M. tuberculosis culture status at 6 months after treatment initiation without positive cultures thereafter and no disease recurrence during the follow-up period of one year after therapy end [146]. For this study TBNET outcome criteria were preferred since the WHO outcome definitions do not include one-year follow-up post treatment completion to

exclude for recurrent disease (**Table 1**). In addition, the WHO's definition for treatment success involves certain items that cannot be predicted by a biomarker since they depend on the patient's behaviour or clinical decisions in the course of therapy (i.e. treatment completion or change of drugs during the course of treatment) [147]. Furthermore, the WHO uses different definitions for patients with and without multidrug-resistant TB [147, 148]. Therefore, WHO treatment outcome definitions could not be applied in this study. The TBNET defines treatment failure by at least one positive *M. tuberculosis* culture 6 months after treatment initiation or thereafter, or a relapse within one year after treatment completion [146].

Table 1 Comparison of the outcome criteria for TB therapy by the WHO and TBnet. Not shown: Undeclared/Lost to follow-up/Not evaluated

Outcome	WHO DS-TB Criteria	WHO MDR-TB Criteria	TBnet Criteria
Cure	Bacteriologically confirmed TB and culture- or smear negativity during the last month of therapy / last sampling [147]	Therapy terminated according to national guidelines, no signs of failure, at least 3 negative culture results at least 30 days after completion of the intensive phase [147]	Negative culture after 6 months of therapy and no relapse within 1 ear after therapy end [148]
Completed Treatment	Completion of therapy without failure evidence but no bacteriological available data during the last month of therapy / last sampling [147]	Therapy terminated according to national guidelines, no signs of failure without three negative culture samples after at least 30 days of therapy [147]	No criteria [148]
Treatment success	Completed therapy + Cure [147]	Completed therapy + Cure [147]	Cure[148]
Failure	Smear or culture positivity for more than 5 months of therapy [147]	Therapy discontinuation or permanent change of at least two antibiotics due to lack of conversion at the end of the intensive phase, renewed positive culture data after conversion, additional development of resistance or adverse events [147]	Culture positivity 6 months after therapy start or relapse within one year after therapy end [148]
Death	Death due to any reason during treatment [147]	Death due to any reason during treatment [147]	Death due to any reason during observation time [148]

Microarray data have been deposited at Gene Expression Omnibus database (GSE147690, GSE147689, GSE147691).

2.3 Ethics

Study approval was granted by the Ethics Committee of the University of Lübeck, Germany (AZ 12-233), which was then approved by the corresponding local Ethic Committees of all participating centres in Germany, and by the Ethics Committee of the MNI (3181/25.03.2015; Bucharest, Romania) [143].

2.4 Data generation

2.4.1 RNA processing and data analysis

Whole blood RNA isolation from PAXgene (Qiagen®, Venlo, the Netherlands) was handled according to the manufacturer´s instructions and stored at -80°C until RNA isolation using the PAXgene blood RNA isolation kit (Qiagen®, Venlo, the Netherlands). Aliquots of isolated RNA were used for quality control to analyze the RNA integrity with the RNA Nano 6000 Kit on an Agilent Bioanalyzer (Agilent®, Böblingen, Germany) according to the manufacturer´s instructions. In case of an insufficient RNA Integrity Number (RIN) as a measure of sample quality and number or signs of degradation, samples were excluded from further analysis [143].

2.4.2 Labelling, hybridization and scanning of microarrays

Total RNA was used for reverse transcription and subsequent Cy3-labelling with the Low-Input Quick Amp Labeling Kit (Agilent®, Böblingen, Germany) according to the One-Color Microarray-Based Gene Expression Analysis Protocol version 6.9.1 (Agilent®, Böblingen, Germany) with RNA Spike-In controls. 1650ng of Cy3-labelled cRNA was hybridized to human 4x44K V2 gene expression microarrays according to manufacturer´s instructions. Arrays were scanned on a SureScan microarray scanner (Agilent®, Böblingen, Germany) at a resolution of 5 µm [143].

2.4.3 Data extraction and normalization

Raw expression data from scanned microarray slides were extracted from tiff files using the Feature Extraction Software version 11 (Agilent®, Böblingen, Germany). Raw data files were imported into Agilent GeneSpring software version 13 (Agilent®, Böblingen, Germany). Percentile Shift was used as normalization method with the 75th percentile as a target and baseline transformation was applied to the median of all samples. Prior to data analysis, quality control

was performed and compromised probes removed from further analysis. “Normalized expression” will refer in the manuscript to Log2 transformed expression values. [143]

2.5 Development of a multistep-model to monitor the course of therapy and to determine the individual therapy endpoint

2.5.1 General conditions

This work consists of several steps. The rough procedure was comparable for all intermediate steps and goes along the following scheme:

1. Identification of a research question for the respective intermediate step including group definition.
2. First variable reduction
3. First model building
4. Further variable reduction
5. Validation

Except for step 1 and step 5, the intermediate steps have different procedures, which will be described in more detail in the following chapters. A mixture of several methods within a multistep approach improves accuracy and makes it feasible to better elaborate specific features for individual questions [149].

The data set used to create the model described in this work was cross-validated in the identification cohort. Furthermore, two additional independent cohorts were established in accordance with the recommendations to verify the results. In the following sections, both the general procedures and the three cohorts are explained in more detail.

Comparisons between cohorts were performed by Kruskal-Wallis Test. The data analyses were performed with the software *R* (versions 3.4.2 to 4.00). The general transcriptomic data preparation was performed with the packages *limma*, *reshape2*, and *tidyr*. The packages *ggplot2* and *vcd* were used for graphical data presentation of model’s results. For pathway and module analysis, *limma* was used and the *pheatmap* package for heatmap visualisation of KEGG pathways. Therapy outcome estimations and the transformation of this factor to a numerical score for the development of the therapy outcome score were performed using the *R* packages *caTools*, *tidyr*, *reshape2*, *randomForest*, *dplyr* and *pROC*. The development of the therapy progress score was executed by using the *R* packages *MASS*, *tidyr*, *reshape2*, *dplyr*, *glmnet* and *caTools*. All available time points were included for the therapy progression score and the therapy end model regardless of therapy outcome status. The *R* packages *car*, *caTools*, *tidyr*, *dplyr*, *glmnet*,

randomForest and *pROC* were used to develop the therapy end model. In general, the outcome "lost to follow-up" was not an exclusion criterion to calculate hypothetical end-of-therapy scores.
The *limma* package of the Bioconductor in *R* was used for preparation of raw text files [105]. *limma* is a Bioconductor package that can be used to analyze gene expression data. Gene expression experiments have a high complexity on the one hand, but on the other hand there are often only a small number of replicates. This makes special preparatory steps necessary, as well as statistical procedures. These are summarized in the *limma* package [105]. In this work, however, in addition to the functions of the *limma* package, other statistical methods are used (see chapter 2.4.2 – 2.4.5) [143].
As shown in **figure 3**, only data from GIC cohort were used for model training, while GVC and RVC data were used for validation. It is important to note that only sub-model 1 and 2 data of both MDR and DS-TB patient data were used. The final model was trained only in DS-TB patients. All models were cross-validated with a split ratio of 0.7. **Figure 3** shows the Flow Chart for the model building procedure.

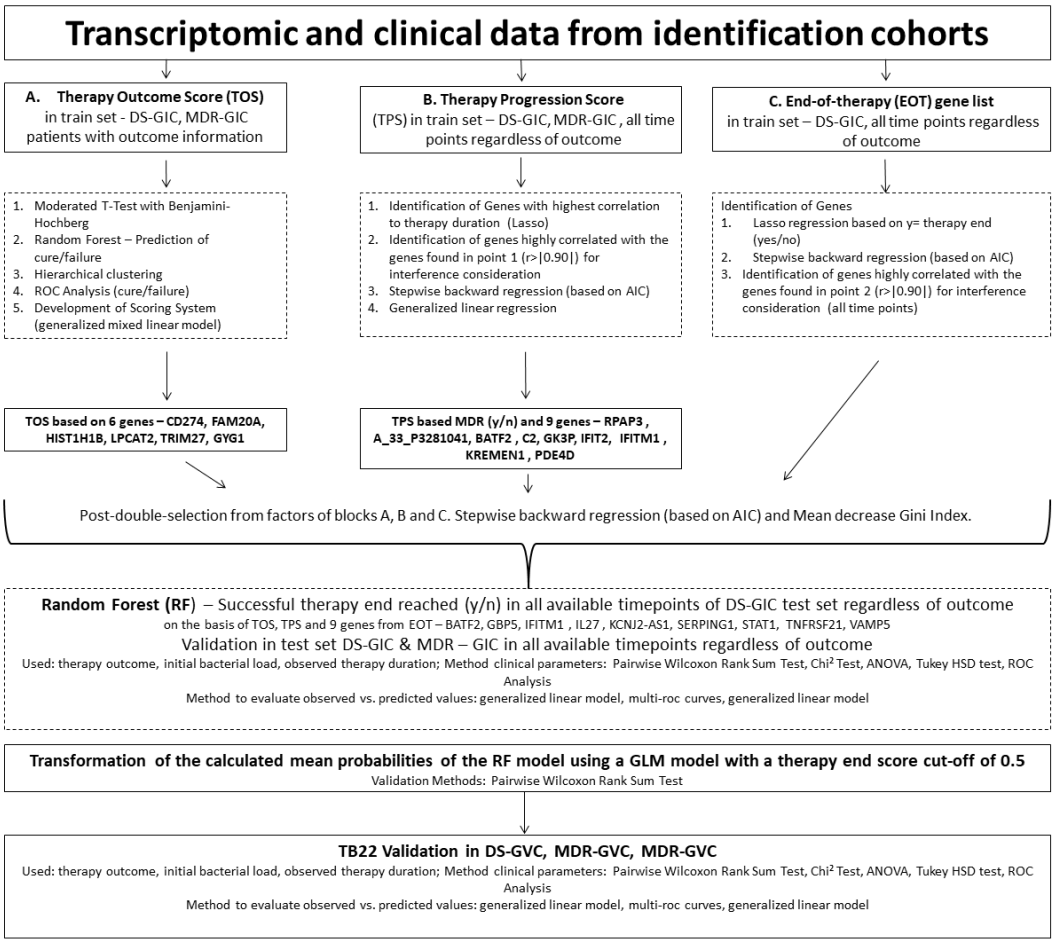

Figure 3 Flow Chart for model development. A) Therapy outcome score (TOS). Identification of genes corresponding to the outcome (Cure/Failure/Death) by cross-validation in all patients of the identification cohort. Conversion of outcomes into numerical values, which act as scores. B) Therapy progress score (TPS). Identification of genes that correlate with remaining therapy duration by cross-validation in all patients of the identification cohort. Subsequent identification of moderating genes and creation of a model that predicts the remaining days to the end of therapy. C) End of therapy gene list identification. Identification of genes that are expressed differently after successful completion of therapy than before therapy end and adoption of TOS and TPS values from A and B [143].

2.5.2 Sub-model 1 – Therapy outcome score (TOS)

In order to find genes that could make a statement about the therapy outcome; however, genes were first identified that were expressed differently between diseased and healthy controls. The idea was to discover TB-associated genes and later to reduce the genes to such an extent that they could not only reflect diagnostic characteristics but much more the status of the diseased person. This was to avoid that those genes that were randomly expressed differently and targeted

a specific aspect of the diseased and thus were not globally usable were given too much weight and that only disease-associated genes were included in the model. TB-related genes should not only differ in their biological plausibility between healthy and TB patients, but should also be able to provide information on the individual disease status that could be used for the prediction of therapy outcome.

To find suitable genes for this purpose, healthy controls and therapy-naïve patients from the training set from GICs were analyzed by a moderated T-test with a fold-change cut-off of ≥2 and a Benjamini-Hochberg multiple testing correction of p ≤0.01 [143]. The advantage of the moderated T-test with Benjamini-Hochberg correction is that for small p-values the misclassification rate is close to the optimum and comparable to Bayesian methods[150]. Hierarchical clustering was run on samples and entities with a Pearson Centered similarity rule and Ward´s Linkage rule [151].

Correlation was calculated according to Pearson´s correlation coefficient [152]. Due to the previously performed normalization, also reduces the variable variance, extreme outliers are not to be expected in the data set, which means that Pearson's method can be used [153]. Analysis of gene lists for enriched pathways was performed via the web-interface of *g:Profiler* [154]. Here, KEGG and Reactome pathways were chosen from the curated gene sets with a p-value cut-off for multiple corrections below 0.05. With this procedure more than 200 genes were identified, which were expressed differently between healthy and TB patients. In addition, stepwise model reduction based on Akaike information criterion (AIC) and recombination using hierarchical clustering was used to identify a gene set of minimal size and optimal fitting [143].

AIC is defined as

2.5.2. −1.

$$AIC = 2k - 2\,ln(L)$$

Where k is the number of estimated parameters and the maximum of likelihood function is *L* [155, 156].

For this purpose, the *randomForest* package for *R* was used, since it offered the possibility to use Breiman's random forest algorithm to create regression models for predicting therapy outcome [157]. For tree building V is the number of all relevant variables. For each node a number v<V is determined, which is the best splitting ratio. Each node receives a random combination of V of size v and remains constant for all trees throughout the operation. The error rate is significantly dependent on the correlation of two trees. The higher the degree of correlation is, the greater the error rate. Variable reduction leads to a decrease in correlation and so using as less variables as necessary not only reduces overfitting but also error rates [158]. In summary, an additional level

of bagging randomness is given in random forest algorithms. A different bootstrap of sample and data is used for each tree, which lead to different classifications in each tree [143, 157].

Single models and sub-models from different gene combinations were chosen to select the most promising gene candidates. The models were adjusted using the Mean Decrease Gini coefficients of random forest analysis [159]. Mean decrease Gini is defined as:

2.5.2. −2.

$$\Delta\, i(\tau) = \; n\, i(\tau) - \; n_l\, i(\tau) - \; n_r i(\tau)$$

Where Kendall's τ coefficient described the rang correlation between variables and Gini impurity $i(\tau)$ is defined as

2.5.2. −2.1.

$$Gini\ Impurity = 1 - Gini$$

with

2.5.2. −2. 2.

$$Gini = \sum_{i=1}^{n} p_i^2$$

Where p_i^2 is probability of all labels within a node.

This resulted in a number of possible models. These selected models were tested for their ability to discriminate between outcome status (see **supplement table 4**) at therapy start via ROC analysis by comparing AUCs [160, 161]. The AUC for ROC curves is a performance parameter that allows conclusion about test validity [161].

The ability to differentiate between patients with therapy failure and cure at different time points after the start of therapy was evaluated. The model that offered the best possible predictive potential for outcome prediction was identified by repeated ROC-curve analysis. This should be seen as a first step for the sub-model. Following the identification of genes that could be classified as having excellent performance, the therapy outcomes cure, failure and death were transformed to numerical values in DS-GIC and MDR-GIC dataset. This numerical translation acted as a dependent variable in the therapy outcome score model (TOS). The previously identified genes of the six gene signature were used as independent variables using a generalized linear model (GLM) [162-166]. It was applied in therapy naïve patients of GICs regardless of their resistance status and resulted in effect estimators for each gene. This TOS model was then applied to the whole GICs datasets, so each patient receives an individual score at each measurement point [143].

Final formula for TOS is defined as:

2.5.2. −3.

$$TOS = -\propto + \beta_1 * CD274 + \beta_2 * FAM20A + \beta_3 * GYG1 + \beta_4 * HIST1H1B + \beta_5 * LPCAT2 + \beta_6 * TRIM27 + \varepsilon_i$$

Where β_i is the estimator for the variables and ε_i is the residual standard error.

2.5.3 Sub-model 2 – Therapy progress score (TPS)

Block B, shown in **Figure 3**, was created independently of those in the previous chapter on TOS. Accordingly, all genes were taken as a starting point here again. However, for this step only therapy naïve patients of the GICs were included in the analysis. In order to be able to map the patient's temporal status of therapy, the genes that had the highest correlation to the observed left duration of therapy (in days) were pre-selected by using least absolute shrinkage and selection operator (LASSO) regression [143]. Here, this works with a penalty term that truncates the estimator of each variable. In lasso regression, the estimator can be reduced to 0, whereby de facto no influence is attributed to a variable. Their β estimator is pushed to 0. The optimal ʎ is determined in a cross-validation procedure (80:20). The best ʎ value leads to the smallest root mean square error (RMSE) and thus the smallest prediction error [119, 122, 167]. Lasso is defined as:

2.5.3 − 1.

$$LASSO = \sum_{i=1}^{n} (y_i - x_i'\beta)^2 + n\lambda \sum_{j=1}^{p} |\beta_j|$$

Where $x_i=(x_i1,...,x_ip)$ is the p-dimensional regression co-variate, $\beta=(\beta1,...,\beta p)$ are the associated regression coefficients, and $\beta_j=0$ *for* $j \leq p0$ and $\beta_j=0$ for $j>p0$ for $p0 \geq 0$[168].

Lasso regression was used here to preselect from the total data set of ~44,000 genes [119, 137, 168-170].

Resistance status (DS-TB / MDR) was included as an additional variable in the initial model of this block [119]. Using stepwise regression based on AIC, a reduced regression model was developed. Other variables, which were not significantly regressed after the AIC-based stepwise regression, were further removed from the model. Using GLM again, the effect estimators for the respective genes enables the generation a numerical value, which allows a conclusion to be drawn about the therapy progression.

Final formula for TPS is defined as:

$$2.5.3 \; - \; 2.$$

$$TPS = -\propto \; + \beta_1 * A_{33_{P3281041}} + \; \beta_2 * BATF2 - \; \beta_3 * C2 - \; \beta_4 * GK3P + \; \beta_5 * IFIT2 - \beta_6 * IFITM1 \; + \; \beta_7 * KREMEN1 \; + \; \beta_8 * PDE4D \; + \; \beta_9 * RPAP3 + \beta_9 * Resistance_{\,MDR,yes} + \; \varepsilon_i$$

Where β_i is the estimator for the variables and ε_i is the residual standard error.

2.5.4 End of therapy (EOT) gene list

The identification of genes that had a significant effect on the end of therapy was performed in the DS-GIC cohort.

Lasso regression was applied to the binary variable "end of therapy" (1= yes/ 0=no, ongoing) in order to identify from the microarray data set those variables by which the dependent variable can be described. The model calculated from the selected variables was further reduced using stepwise regression based on AIC. This resulted in a list of genes that had a significant effect on the probability of reaching the end of therapy. These genes, together with the TOS and the TPS, formed the basis for the final model, which is explained in the following section.

$$2.5.4 \; - \; 1.$$

$$End - of - therapy = \sum_{i=1}^{44000} (y_i - \; Genes_i' * \beta)^2 + 44000 * \Lambda \sum_{j=1}^{44000} |\beta_j|$$

Where $y_i$the probability of therapy end, β is the estimator for each gene and Λ is the amount of shrinkage for β.

2.5.5 Signature identification and final TB22 model

TOS, TPS and EOT list from the previous steps have now been merged into a common model with the binary variable end of therapy. This model was created using only data from DS-TB patients of the identification cohort and cross-validated by train- and test-set in this group. First, a post-double selection was performed to better test the multiplicative effects [171-173]. Subsequently, the stepwise backward regression based on the AIC was performed for the purpose of variable reduction. The model was then further reduced by removing non-significant variables and testing model performance with an ANOVA and AIC [143].

The remaining 22 genes were used to test diagnostic performance in an external dataset (GSE144127). This was used because this data set is the data basis for a meta-analysis comparing the diagnostic performance of different signatures [116]. Using the same methodology as stated

in the article, comparability with other signatures has been achieved. For data analysis, the dataset provided by Hoang et al at GEO repository was used (GSE144127) [116, 174].

First, binary variables were created from the given dataset in terms of diagnosis, which corresponded to the research question. These were active culture confirmed pulmonary TB vs. other diseases; highly probable active pulmonary TB vs. other diseases; all active pulmonary TB vs other disease; active extra-pulmonary TB vs. other diseases; latent vs. active TB and latent TB vs. other diseases. For each analysis, the inapplicable cases in comparison-specific datasets were removed. Then, the remaining datasets were split into a respective training set (80%) and a test set (20%) using the sample function of the *R* package *caTools* with a split ration of 80:20.

The random forest algorithm of the *randomForest* package was used to train the TB22 signature in the training datasets.

The random forest models were then applied to the test set using *predict* function. The roc function of the *pROC* package determined the match rate of observed and predicted diagnostic results by means of the area under the curve (AUC).

For final setting, the random forest algorithm was used. The results from this final model can be summarized in two ways. On the one hand there is the simplified, binary statement whether the therapy can be terminated or not, on the other hand a metric variable arises with the probability, which also permits a progress check during the therapy and thus a monitoring.

Final TB22 is defined as:

$2.5.5 - 1.$

$$TB22 = -\propto + \beta_1 BATF2 - \beta_2 GBP5 * TOS - \beta_3\, IFITM1, -\beta_4 IL27 + \beta_5 KCNJ2AS1, + \beta_6 SERPING, + \beta_7 STAT1 + \beta_8\, TNFRSF21, + \beta_9 VAMP5, + \beta_{10} TOS, + \beta_{11} TPS + \beta_{12} TOS * TPS + \varepsilon_i$$

Where β_i is the estimator for the variables and ε_i is the residual standard error and

$2.5.5 - 1.2.$

$$TB22(mr\ (X,Y)) = P_{\emptyset}(h(X,\emptyset) = Y) - max_{j \neq Y} P_{\emptyset}(h(X,\emptyset) = j)$$

Where $\mathrm{h_k(X)}$are an esemble of classifiers and *X,Y* is the distribution of random vectors. In random forest, *h(X, Ø)* =$h_k(X)$.

The remaining 22 genes were used to test diagnostic performance in an external dataset (GSE144127). This was used because this data set is the data basis for a meta-analysis comparing the diagnostic performance of different signatures [116]. Using the same methodology as stated in the article, comparability with other described signatures could be achieved.

First, binary variables were created from the given dataset in terms of diagnosis, which

corresponded to the research question. These were culture confirmed TB vs other diseases; highly probable TB vs other disease; TB (confirmed and not confirmed) vs. other diseases; latent vs. active TB; latent TB vs. other diseases. The cases that were not needed in the respective analyses were removed. Then, the respective remaining dataset was split into a training set (80%) and a test set (20%) using the sample function of the *R* package *caTools* with a split ration of 80:20.

The random forest algorithm of the *randomForest* package was used to train the respective models for the signature in the training dataset.

Random forest models were then applied to the test set using predict function. The roc function of the *pROC* package determined the match rate of observed and predicted diagnostic results by means of the area under the curve (AUC).

2.6 Model Validation

2.6.1 Comparison to existing signature

As shown **supplement table 2**, previous studies have already presented RNA-based biomarker signatures, which were developed for prognostic purposes of various clinical endpoints. Although a point-of-care signature has not yet been described with regard to shortening the duration of therapy in TB, it would be theoretically possible to rewrite different progression signatures to the final therapeutic classification as described in this work. In order to compare the performance of the signature described in this thesis with previous achievements in TB biomarker research and to determine the added value of this work's model, biomarker signatures have been identified in the literature [143].

In order to further validate the discriminative performance of the model, published RNA signatures and scores were also assessed for their capability to identify end-of-therapy timepoints and compared with the model [114, 117, 118, 175-184]. For this purpose, TB associated signatures or scores with less than 100 genes were identified. The gene sets or scores were trained in the training set of the DS-GICs (split ratio 0.7) using random forest algorithms to classify between for the probability of cure associated end-of-therapy time points. Duplicate genes in single signatures and unspecific sample names were not considered for the models. The resulting models were then applied to the independent DS-GVC data set to test for their discriminative performance in an external data set (**Supplement Table 2** and **Figure 3**). In a subsequent step, the application of the models was extended to both DS and MDR-GVC data sets to test for plausibility.

In comparison to the model development and verification in the previous section, the entire DS-GIC was used as the model training data set in this section, since the previously fixed

compartments meant that model optimization via cross-validation did not have to take place. At the same time, GVC, RVC and SGVC were merged, as a simple validation was sufficient for the purpose of this section and thus to have a larger validation data set. Since all samples were processed in the same laboratory by the same staff and under the same conditions, and since statistical analysis and data preparation were also performed by a single person, a negligible batch effect, if any, was assumed for which no special countermeasures were taken [185].

For TOS and TB22, AUC was used to determine how well they indicate the end of therapy. Again, a random forest algorithm was trained with the respective genes and applied to the validation cohort. For the model, which is described in this thesis, the respective preceding sub-models for TOS and TPS were applied as well. No other signature described a comparable data splitting for modelling, but when gene signature was described as a replicable score, the respective calculated score was used for performance testing.

In brief, the signatures used for the comparison are described in **Supplement table 3**

2.6.2 Systems biology validation

After the bioinformatic and biostatistical identification of the genes, the check for clinical plausibility and the performance comparison with existing TB-related signatures, the biological functionality was examined. For this purpose, two approaches were used. First, the respective genes that were part of the model, where described for their overall biological function, their related systemic pathways and their previous description with regard to TB in literature. Second, an unbiased approach was used to identify related pathways that showed significantly up- or downregulated genes in this dataset.

An enrichment analysis identifies genes that are overrepresented in the dataset. While each gene is present only once in DNA analyses, the expression level in RNA data is more differentiated and a conclusion can be drawn depending on whether they are up- or down-regulated in a group, meaning that they are over- or under-expressed. Since in the context of the several 10,000 transcriptomes an unsorted result representation took place, significant transcriptomes must be summarized here, in order to be able to make a concrete conclusion[186, 187]. Genes and their up- and downregulation have been grouped a priori into different pathways that describe specific functionalities or are characteristic of specific disease phenotypes.

These gene sets and the resulting pathways are the basis for gene enrichment analyses. In this work, two different enrichment analyses are performed. First, the pathways involved are identified based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) [187]. Those were identified by using *kegga* function of *limma* package after applying Empirical Bayes Statistics for

Differential Expression (*ebayes*) to all genes with regard to the specific question (see next paragraph). With *ebayes* a microarray linear model fit is built, which also compute moderated F- and T-statistics and log-odds of differential expression by empirical Bayes moderation of the standard errors [105]. After identification of KEGG pathways, analysis of transcriptional modules (*tmod*) is further performed [186]. These transcriptional modules were already described in 2008 and differ from pathways in that they were classified using a data-mining algorithm and do not follow biological pathways alone. The constructed modules are based on biological variability. A cluster analysis was performed to identify the associated genes to specific disease and functional classifications [188]. The genes are first sorted with respect to their coefficient from the previous ebayes analysis and then an enrichment analysis is performed. Then, based on coefficients and significance value, the number of up- and down-regulated genes of each module is calculated and subsequently plotted [186].

Here, several disease-related categorizations were made to identify KEGG pathways and *tmods* that are relevant for those comparisons. For the analyses the data from all cohorts were used to be able to identify also smaller effects in an enlarged sample size [189].

The categories tested were the following:

- TB patients vs. healthy controls
- DS-TB patients vs. MDR-TB patients
- Therapy naïve patients vs. patients with ongoing therapy
- Therapy naïve patients vs. patients after two weeks of treatment
- Patients with ongoing therapy vs. patients with clinical therapy end
- Patients with clinical therapy end vs. patients with calculated therapy end
- Patients with cavities vs. patients without cavities
- Patients before culture conversion vs. patients after culture conversion
- Patients with therapy outcome cure vs. patients with therapy outcome death and failure at baseline
- Calculated therapy end vs. healthy controls

Module and pathway results were clustered to identify related questions. Furthermore, the KEGG Pathways were assigned to head categories and further clustering was also done for this purpose. The background is the representation of the multidimensional involvement of different pathways, since the disease cannot be reduced to direct immunological aspects only.

The KEGG pathways were also subjected to an analysis with regard to their concordance rates of the individual questions.

3 Results

3.1 Patient characteristics

Seventy-nine patients were included in the GIC, 60 in GVC, 52 in RVC and 20 in SGVC. **Table 2** shows the results from study enrolment. While the GICs and the GVCs had a multi-centric character, only patients from MNI where enrolled for the RVC and the SGVC included patients from Research Centre Borstel only [143].

Table 2 also gives an overview of the individual cohorts in terms of age and gender structure as well as TB-related clinical data; namely, time to positivity (TTP+) at therapy initiation, time to culture conversion (TTC), and therapy outcome according to TBnet criteria [148].

Table 2 Characteristics of TB patients, observed and predicted therapy durations in patients from the German identification cohort (GIC), the German validation cohort (GVC), and the Romanian validation cohort (RVC).

	DS-TB (n=81)		**MDR (n=100)**			
	GIC	**GVC**	**GIC**	**GVC**	**SGVC**	**RVC**
	n= 50	**n= 28**	**n= 30**	**n= 31**	**n=20**	**n= 52**
Baseline Age (mean, SD)	48.2 (40.0-60.2)	34.6 (22.1-49.3)	36.2 (32.0-41.6)	33.2 (24.5-44.7)	29.2 (25.1 – 39.4)	37.0 (28.3-46.7)
Baseline TTC+ in days (mean, SD)	21.0 (8.0-29.0)	10 (7.0-12.0)	22.0 (11.8-32.5)	22.0 (11.8-32.5)	15.0 (13.0 – 20.0)	40.0 (27.5-56.0)
Time to culture conversion in days (mean, SD)	47.5 (25.8-75.0)	46.0 (24.5-55.0)	38.0 (33.0-215.5)	50.0 (30.5-59.8)	21.0 (17.0 – 87.0)	32.0 (27.0-60.0)
Therapy outcome						
Cure	29 (58.0%)	20 (71.4%)	17 (56.7)	20 (62.5%)	2 (9.5%)	34 (65.4%)
Failure	7 (14.0%)	1 (3.6%)	3 (10.0%)	1 (3.1%)	0 (0.0%)	4 (7.7%)
Death	1 (2.0%)	1 (3.6%)	2 (6.6%)	1 (3.1%)	0 (0.0%)	-
Lost to follow-up/ undeclared/pending	13 (26.0%)	6 (21.4%)	8 (26.7%)	10 (31.3%)	18 (90.5%)	14 (26.9%)

SD= standard deviation, TTC+=time to culture positivity, MDR-TB= multidrug-resistant and extensively drug-resistant tuberculosis.

3.2 Signature performances

In this chapter the results and performance of the individual models are presented. The array data went through three individual steps as shown in **Figure 4**. First, transcriptomes were identified that were significantly up- or down-regulated in comparison between therapy naïve patients and HCs, from which the model for the TOS consisting of 6 genes - CD274, GYG1, FAM20A, HIST1H1B, TRIM27 and LPCAT- was generated. The second step identified the genes that had a significant influence on the remaining therapy time even after estimator penalizing by lasso regression. After model reduction steps the genes RPAP3, A_33_P3281041, BATF2, C2, GK, IFIT2, IFITM1, KREMEN1 and PDE4D remained for the TPS. As a last step, genes were identified for the EOT 64 by lasso regression, which after estimator penalizing showed significant influences on the binary determination of the end of therapy. The EOT, TPS and TOS were then used in the fourth step to create a model that can distinguish between ongoing therapy and cure-associated therapy end [143].

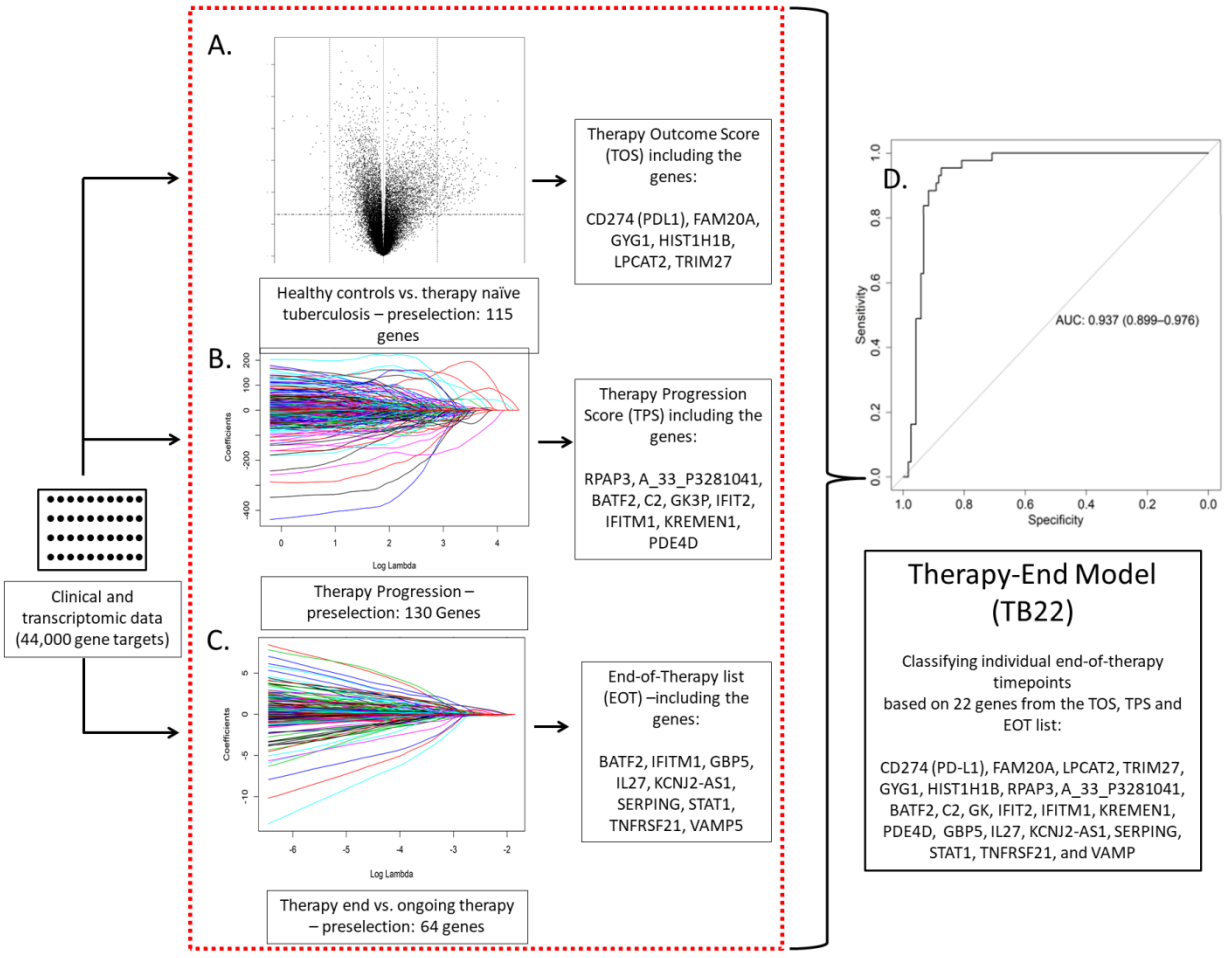

Figure 4 Multi-step development of the TB22 for tuberculosis treatment Simplified flow chart showing the multi-step approach of transcriptomic and clinical data analysis to develop TB22 that identifies the optimal

timepoint to stop anti-tuberculosis therapy. A. Development of therapy outcome score (TOS). Showing the volcano plot representing differentially expressed genes in healthy controls vs. therapy naïve drug-susceptible (DS-) and multidrug-resistant (MDR) tuberculosis patients from the German identification cohorts (GICs). Genes that were significantly up- or down-regulated (significant 2-fold change after Benjamini-Hochberg correction) form the basis for the TOS development. B. Therapy progression score (TPS) development. Depiction of penalizing regression coefficient adjustment (y-axis) and the explained deviation as a function of Log-λ (x-axis) for variable selection to identify genes that predict the remaining days of therapy that has been conducted in reality in all sample measurements from DS- and MDR-GIC tuberculosis patients. Each line represents one gene of interest and the genes shown in the plot were pre-selected by the initial Lasso regression step. The initial data selection was carried out on the entire data set with 44,000 gene targets. C. End-of-therapy list (EOT list). Showing penalizing regression coefficient adjustment (y-axis) and the explained deviation as a function of Log-λ (x-axis) for variable selection to identify genes that classify between sample measurements in DS-GIC tuberculosis patients under therapy vs. timepoints at the end of relapse-free therapy in DS-GIC tuberculosis patients. Each line represents one gene of interest and the gene targets shown in the plot were pre-selected by the initial Lasso regression to reduce the number of genes of interest. D. Therapy end model (TB22). Implementing the gene scores (TOS and TPS) and the EOT list into a machine learning algorithm model (Random Forest), a final simplified TB22 for the calculation of end-of-therapy timepoints was developed via Generalized Linear Model (GLM). The initial TB22 evaluation was carried out on data from DS-GIC tuberculosis patients. The ROC-curve shows TB22 classification accuracy in the independent data set of DS German validation cohort (GVC) tuberculosis patients (Area under the curve (AUC) 0.937, confidence interval (CI) 0.899-0.976)). TB22 was further applied to patients with multidrug-resistant tuberculosis from the GIC, GVC, and from the Romanian validation cohort (MDR-RVC) [143].

3.2.1 Diagnostic performance

In the article by Hoang et al., based on the dataset used here to review diagnostic performance, different signatures are described to distinguish between culture confirmed and/or very likely pulmonary and extra-pulmonary TB and other diseases [116]. TB22 signature showed similar performances to discriminate between culture-confirmed tuberculosis and other diseases compared with best signature provided by Hoang et al (Hoang et al. area under the curve [AUC]: 0.88 (95%-confidence interval (CI): 0.77-0.96) [116], TB22 AUC: 0.89 (CI: 0.82-0.96)); **Figure 5**). For non-culture confirmed pulmonary TB, the AUC was 0.80 (95% CI: 0.66 - 0.93) in Hoang et al. [116] and also (95% CI: 0.68 - 0.93) for TB22 signature (**Figure 5**). The overall performance for the diagnosis of pulmonary TB was thus AUC=0.87 (95%-CI: 0.81 - 0.92) for Hoang et al. [116] and AUC=0.84 (95%-CI: 0.77 - 0.91) for the TB22 signature (**Figure 5**). For the diagnosis of extra-pulmonary TB, the best signature of Hoang et al. shows an AUC of 0.86 (95%-CI: 0.68 - 0.94) [116],

while the TB22 signature has an AUC of 0.89 (95%-CI: 0.82 - 0.96) (**Figure 5**). In addition to the comparisons described in the article, we tested how well TB22 can distinguish between active and latent TB. Here, TB22 shows a performance of AUC=0.86 (95%-CI: 0.63 – 1.00) (**Figure 5**).

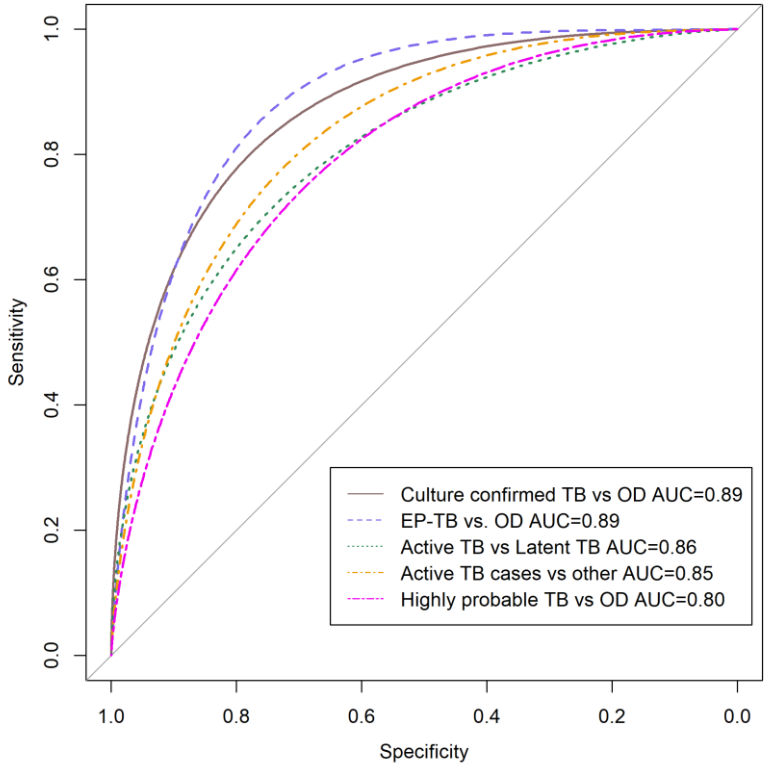

Figure 5 ROC curves for TB22 Signature diagnostic performance for different discriminations. TB=tuberculosis; OD= other diseases; AUC= area under the curve.

3.2.2 Performance of TOS and TPS

The best predictive property for discriminating between healthy controls and therapy-naive TB patients was given by a signature with 6 genes - CD274, FAM20A, GYG1, HIST1H1B, LPCAT2 and TRIM27.The discriminative performance of the individual genes of the TOS patients show AUCs between 0.74 and 0.93 at different time points. In the next step, all genes were processed into a signature using an RF machine learning algorithm. The performance of the six-gene signature in terms of its ability to discriminate between outcome groups (cure/failure/death) at different points in time can be rated as outstanding, with AUCs between 0.85 and 1[190]. With regard to the TOS resulting from the genes, it can be seen that patients who die during the course of therapy have a higher TOS at the beginning of therapy (**Figure 6**). Patients with failure have a higher TOS on average at the beginning of therapy [143].

Further validation of this finding was performed by using the patients' culture conversion status, the drug-resistance status, and the radiographic extent of disease at baseline to re-assure the plausibility of the therapy outcome score. Kruskal-Wallis test showed highly significant differences of the therapy outcome score between outcome groups at baseline in the GICs ($p<0.001$). MDR- and DS-GIC patients with prolonged times of culture conversion (2-month culture conversion status) also showed higher baseline score values (median: 3.4, IQR: 1.6–5.0) when compared to those with early culture conversion (median: 1.8, IQR: 0.5–3.4, $p=0.111$). Therapy naïve MDR- and DS-GIC patients experiencing cure had a median score of 0.86 (IQR: -0.37–2.3) while patients with therapy failure exhibited a median score of 1.3 (IQR: 0.34–4.61) and deceased patients a value of 5.83 (IQR: 5.08–10.04; **Figure 6**). The performance of this outcome score to predict therapy outcome in therapy naïve patients was 0.85 (95% CI 0.78–0.92) [143].

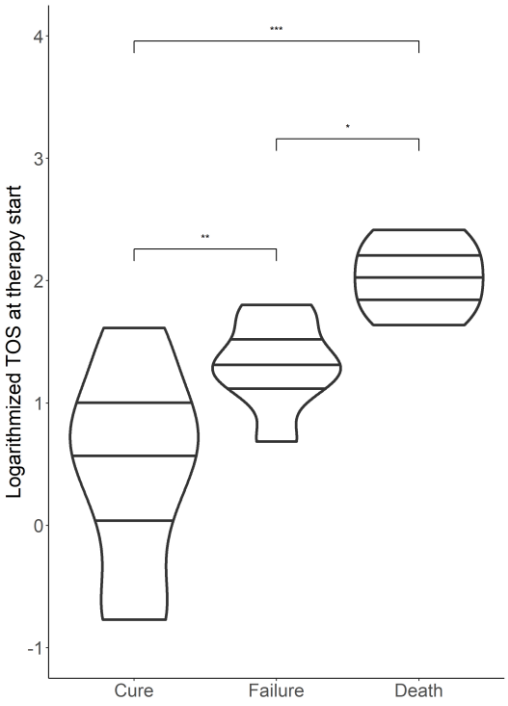

Figure 6 Violin plots showing the comparison of logarithmic therapy outcome score (TOS) values in therapy naïve tuberculosis patients, irrespectively of drug-resistance status, from the German identification cohorts (GIC) with regard to their therapy outcome (cure, failure, or death according to the TBNET-criteria) [148]. Figure adapted from the original publication [143].

Figure 7 shows the course of the TPS as function of the days since the start of therapy stratified by cohort and resistance level. It becomes apparent that the model shows similar courses for DS-TB patients independent of cohort. The same is the case for MDR-TB patients but patients that are

infected by an MDR strain of *M. tuberculosis* showed a higher level in the beginning. The model claim was not to predict the upcoming or remaining therapy duration.

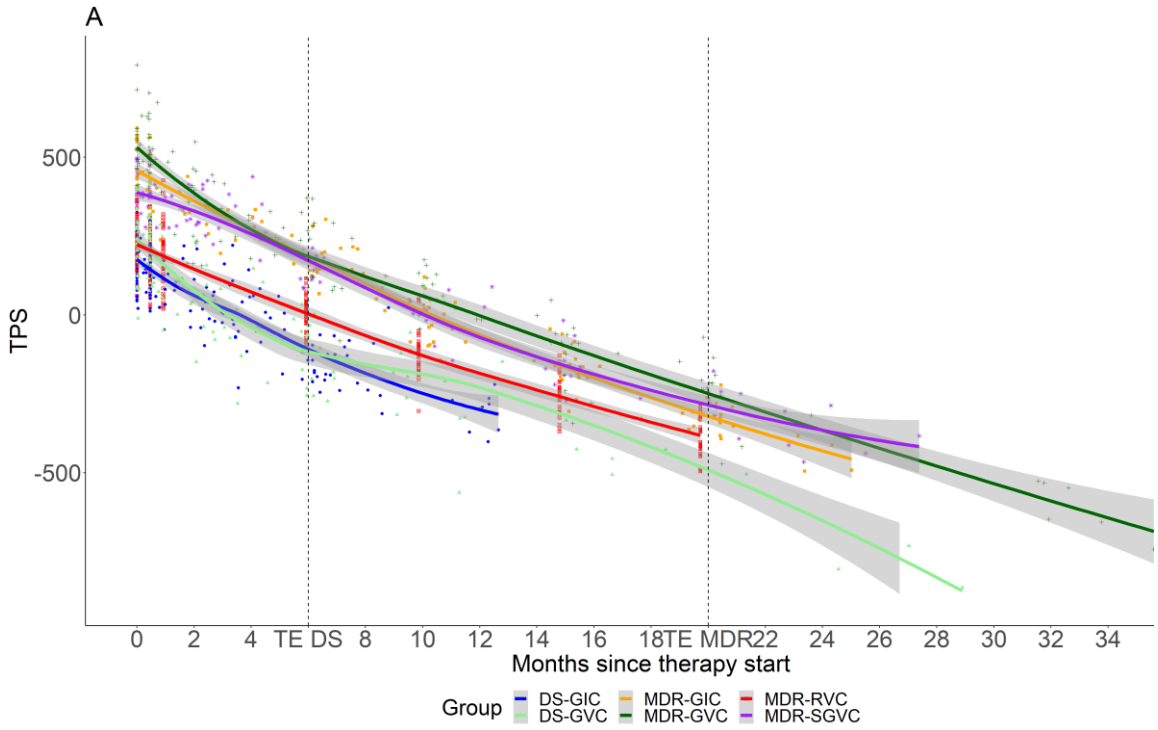

Figure 7 Y-axis: Therapy progression score (TPS) as a function of days since therapy start (x-axis) stratified by cohorts.

Supplement table 4 gives an overview about the model´s parameter. The table summarizes the methodology once again, but also shows the genes used in each case. In addition, it also shows the respective performance parameters of the individual models, including the explanation of the variance, the mean decrease Gini index and the discriminatory ability in the form of the AUC [143].

3.2.3 TB22 performance in the cohorts

The TB22 model finding and validation was completed in three steps. First, the model was trained in the training data set of DS-TB GIC patients and cross-validated in the test data set of the same cohort and patient group. Second, was then applied to the group of DS-TB patients of the GVC to test the performance of the model in an independent data set. Third, model was then applied to the MDR-TB patients in all three cohorts to determine the likelihood that the therapy could be terminated for all patients at all measurement time points. Afterwards, a simplification was made as a GLM model to ensure complete and reliable reproducibility of results, which is not given for random forest Models [143].

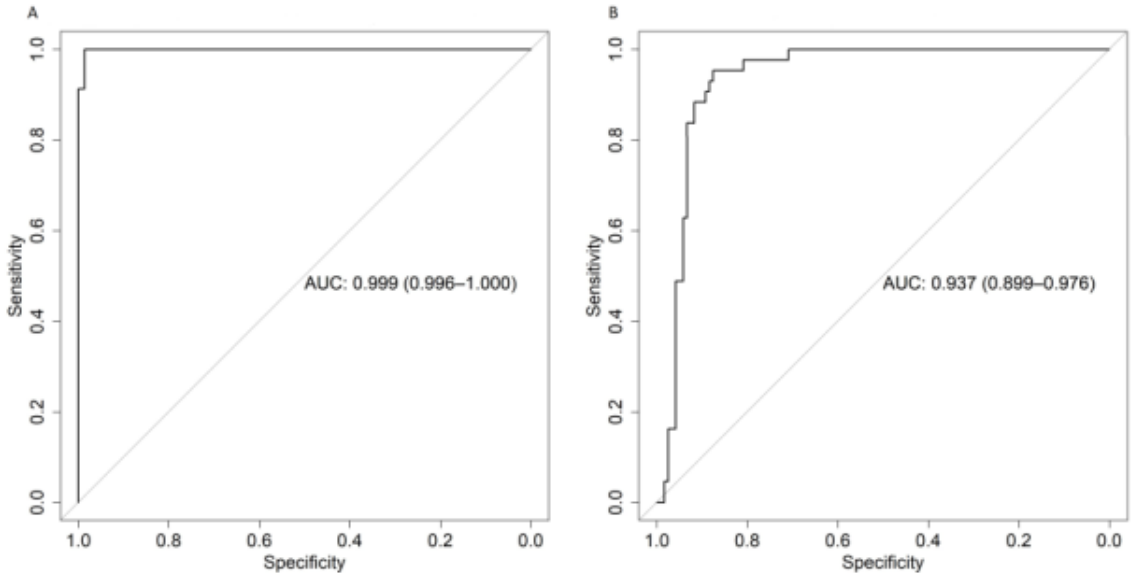

Figure 8 ROC curve analysis to evaluate the biomarker model's performance to predict DS-TB patient therapy end. Clinically conducted therapy end time points are correlated with biomarker defined therapy end timepoints. Figure 8A: Biomarker model performance in DS-TB patients from the German identification cohort (GIC), Area under the curve (AUC) = 0.999 (confidence interval (CI) 0.996 – 1). Figure 8B: Biomarker model performance in DS-TB patients from the German validation cohort (GVC), AUC = 0.912 (CI 0.868 – 0.956) [143].

Each measurement resembles an independent end-of-therapy calculation for a TB patient under therapy. All calculation results above the cut-off ≥0.5 indicate for hypothetical end-of-therapy timepoints with cure as final treatment outcome. The model identified end-of-therapy timepoints with high accuracy in DS-GVC patients (area under the curve [AUC] 0.94; 95% confidence interval [CI]:0.90-0.98; **Table 4** and **Figure 8**). It was applied to MDR-GIC, and to patients from the independent DS- and MDR-GVCs, and patients from the MDR-RVC to calculate hypothetical therapy durations. **Figure 9A-D** shows the end-of-therapy probabilities of the different cohorts as a function of time under therapy [143].

The proportion of patients who reached the model´s threshold for the calculated end-of-therapy at the end of clinical anti-tuberculosis treatment was 100% in the DS-GIC and 97.4% in the DS-GVC. Patients who did not reach the threshold indicating a relapse-free end-of-therapy at month 6 showed an increased time to sputum culture conversion when compared to those who did (median of 68 days, IQR: 50.0–126.0 days vs. median of 46.0 days, IQR: 30.0–63.0 days; p=0.041). None of the patients in the MDR-GIC, MDR-GVC and MDR-SGVC reached the threshold for cure at 6 months. In MDR-RVC only one patient (1.9%; culture conversion within 2 weeks) was above the threshold after 6 months of therapy. Following 15 months of therapy, the overall proportions of

MDR-TB patients having reached cure according to the model were 84.6% in the MDR-GIC, 40% in the MDR-GVC. 60.0% in MDR-SGVC and 88.5% in the MDR-RVC [143].

The calculated therapy durations did not differ significantly from observed durations for the DS-GIC patients (median calculated 175.0 days vs. observed 184.0 days, p=0.104) but they do for the DS-GVC group (median calculated 225 vs. observed 273.0 days, p=0.001), which could be explained by the larger gaps between sampling timepoints, or higher bacillary burden at baseline. Calculated therapy durations were significantly shorter compared to those observed in patients of the MDR-GIC (median calculated 420.0 days vs. observed 638 days, p=0.001), the MDR-GVC group (median calculated 430.0 days vs. observed 641 days, p<0.001) and MDR-SGVC (median calculated 355.0 days vs. observed 641 days, p=0.001). Calculated therapy durations in MDR-RVC patients were also significantly shorter than the observed durations (median calculated 450.0 vs. observed 609.0 days, p=0.001). For patients in the MDR-GIC, this would have resulted in a median reduction of therapy duration by 218 days. In the MDR-GVC, therapy would have been reduced by a median of 211 days, and for a median of 161 days in the MDR-RVC. In the MDR-SGVC, the median reduction from observed therapy to calculated end point is 254 days. According to the therapy end model, 32.9% of patients with MDR-TB who had a negative *M. tuberculosis* culture status at 6 months of therapy reached the therapy end model threshold after 10 months, 69.5% were above the therapy end model threshold after 15 months and 97.6% were above the therapy end threshold after 20 months of treatment, respectively. Also, in patients from the DS-GVC, 52.0% of patients who had a negative culture status at 2 months of anti-TB therapy had a therapy end model status at month 6 of anti-TB treatment indicating cure. In contrast, 10.0% of patients who had a positive culture status at 2 months of anti-TB therapy had a TB22 status at months 6 of anti-TB treatment indicating cure [143].

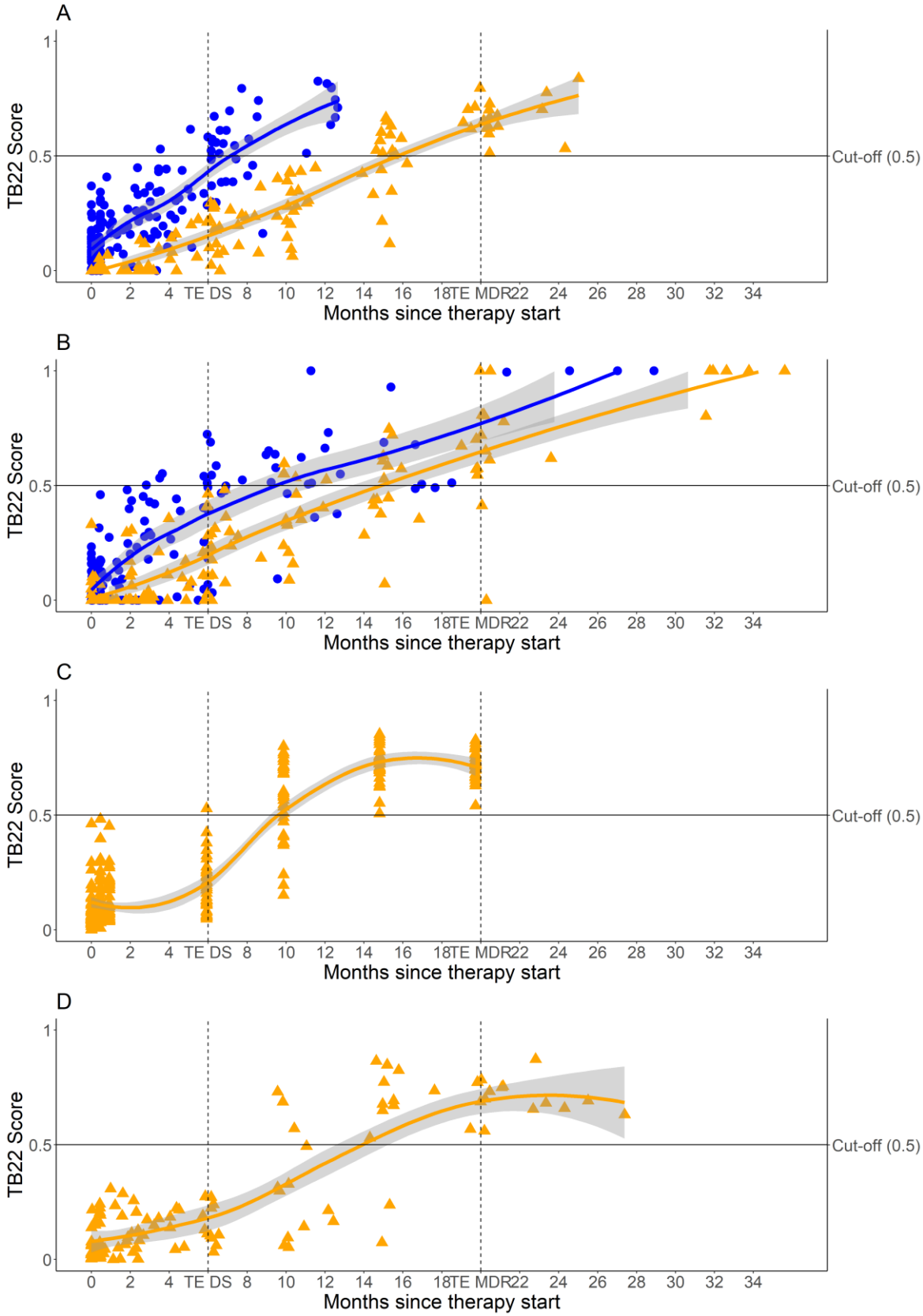

Figure 9 TB22 scores for end-of-therapy by the therapy end model (TB22) over the time of anti-TB treatment for the five cohorts of drug-susceptible (DS) tuberculosis (TB) and multidrug-resistant (MDR) TB patients of German Identification cohort (GIC), German Validation cohort (GVC) and Romanian Validation

cohort (RVC) following the therapy end model. Y-axis: TB22 scores for end-of-therapy, horizontal line: probability threshold (P≥0.5) for relapse-free end-of-therapy; X-axis: time under treatment (months), first vertical dotted line indicates 6 months of therapy, the common timepoint of therapy end (TE) in drug-susceptible TB, second vertical dotted line indicates the usual timepoint for TE in multidrug-resistant TB after 20 months of therapy. Figure 9A: TB22 scores in DS-GIC and MDR-GIC over time (blue: DS-GIC, orange: MDR-GIC) Figure 9B: TB22 probabilities in DS-GVC and MDR-GVC patients over time (blue DS-GVC, orange: MDR-GVC) Figure 9C: TB22 scores in MDR-RVC patients over time (orange). Figure 9D: TB22 scores in MDR-SGVC patients over time (orange) [143].

Probabilities for end-of therapy for all patients from the different cohorts were below the threshold at baseline (**Figure 10A**). The majority of patients with DS-TB reached the end-of-therapy threshold at 6 months while drug-resistant TB patients did not (**Figure 10B**). Nearly all TB patients from the different cohorts reached the end-of-therapy threshold at the end of clinical therapy (**Figure 10C**). The model probabilities for end-of-therapy were also compared between patients with DS-TB and with MDR-TB at relevant bacteriologically defined endpoints such as the individual time of sputum culture and smear microscopy conversion. End-of-therapy probabilities were well below the threshold for both DS-TB and MDR-TB at these timepoints, but probability values were significantly lower for patients with MDR-TB when compared to DS-TB patients in the GICs (median probability at smear conversion: DS-GIC P=0.21 vs. MDR-GIC P=0.06, p=0.038; median probability at culture conversion: DS-GIC P=0.29 vs. MDR-GIC P=0.04, p=0.007) and the GVCs (median probability at smear conversion: DS-GVC P=0.09 vs. MDR-GVC P=0.01, p=0.040; median probability at culture conversion: DS-GVC P=0.29 vs. MDR-GVC P=0.04, p=0.007). Of note, no patient with positive sputum culture result reached the threshold for end-of-therapy as classified by the model. When the model probabilities for therapy end were stratified for drug-resistance status in pooled data from the different cohorts, they showed low probabilities for therapy end at baseline, after 2 weeks of therapy, but probabilities above the threshold at clinical therapy end timepoints [143].

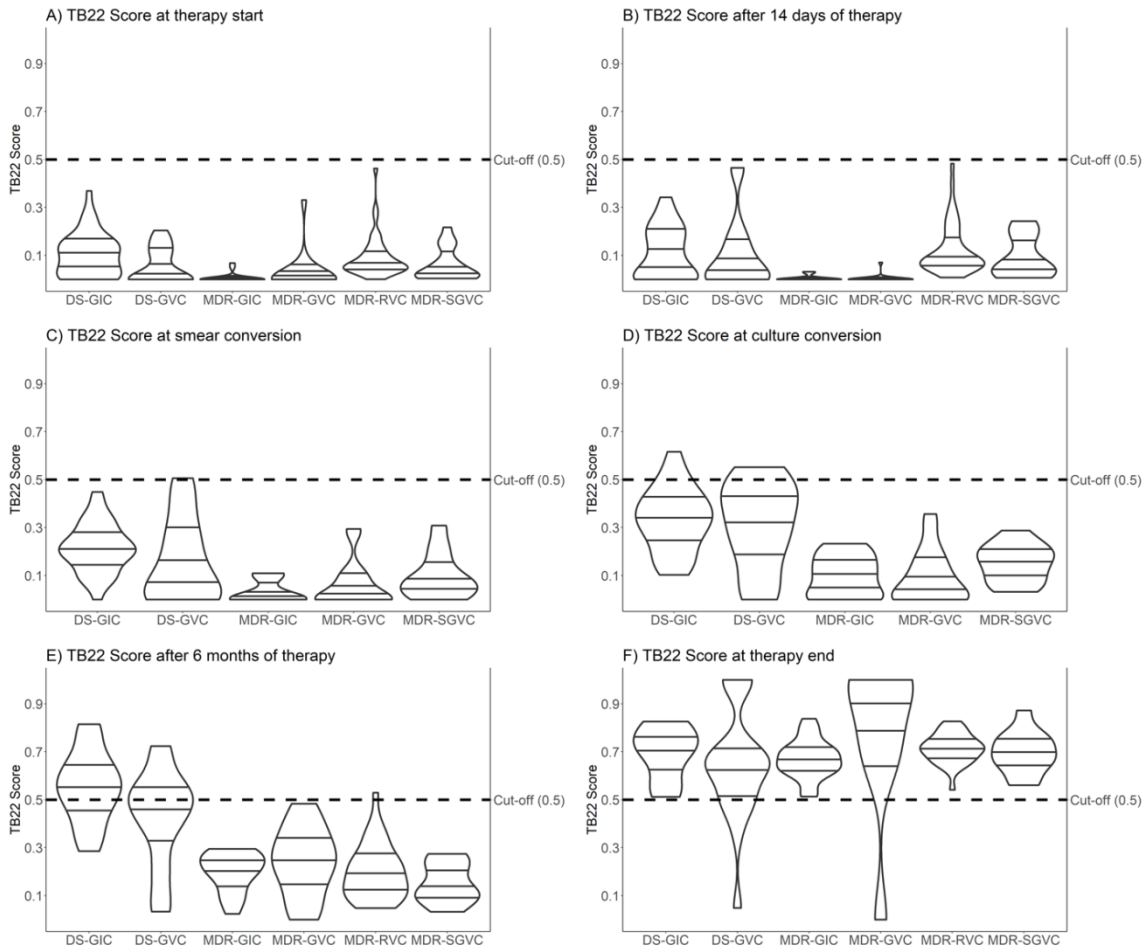

Figure 10 Violin plot with comparison of the calculated therapy end probabilities between the cohorts at different times. Y-axis: probabilities for therapy end, dotted line: threshold for cure-associated therapy; X-axis showing the different cohorts. 10A: at therapy beginning, 10B: after 14 days of therapy 10C: at smear conversion, 10D: at culture conversion 10E: at first measurement after six months 10F: after successful completed therapy. MDR= multidrug- TB, DS-TB= susceptible and mono-resistant TB. Drug susceptible German Identification cohort (DS-GIC), multidrug resistant German Identification cohort (MDR-GIC), drug susceptible German Validation cohort (DS-GVC), multidrug resistant German Validation cohort (MDR-GVC), multidrug resistant Romanian Validation cohort (MDR-RVC), multidrug resistant second German Validation cohort (MDR-SGVC) [143].

3.2.4 Clinical validation

In this chapter, therapy durations and model results are examined using clinical and sociodemographic parameters. Table 8 shows the comparison between observed and calculated therapy duration with regard to their influence by clinical parameters. It becomes clear that the clinical parameters have a significantly lower influence on the observed therapy duration of MDR-TB patients than on the observed therapy duration of DS-TB patients. Cavities have a similar

impact on observed and calculated duration of therapy in DS-TB patients; however, the impact on calculated duration is significant. The effect of cavities is most pronounced on the calculated duration of therapy in MDR-TB patients. In contrast, culture conversion at two months had a much stronger and significant impact on DS-TB therapy durations (calculated and observed), which is not the case for MDR-TB patients.

The baseline smear grade shows no significance on any observed or calculated duration of therapy.

Table 3 Additional days on therapy in the presence of cavities and a high baseline smear grade, and failure to achieve culture conversion (CC) at 2 months for observed duration of therapy and calculated duration of therapy. DS-TB patients: drug-susceptible tuberculosis patients from DS-GIC and DS-GVC. MDR-TB patients: multi-drug resistant TB patients from MDR-GIC, MDR-GVC, MDR-RVC and MDR-SGVC.

	DS-TB patients		**MDR-TB patients**	
	Observed (p-value)	**Calculated (p-value)**	**Observed (p-value)**	**Calculated (p-value)**
Cavities	+40.2 (0.302)	+47.7 (0.008*)	+7.9 (0.523	+76.7 (<0.001*)
No CC after two months	+182.8 (<0.001*)	+32.7 (0.044*)	+23.2 (0.006*)	+19.4 (0.089)
Baseline smear grade >2	-54.3 (0.134)	+9.5 (0.551)	+9.2 (0.282)	+2.3 (0.831)

DS-TB= Drug-susceptible tuberculosis patients, MDR-TB=Multidrug-resistant tuberculosis, CC=Culture conversion

A further validation was carried out by analysing, which classification patterns exist regarding the cultural status in the course of therapy. The mosaic plot shows results for patients of the GIC, GVC and SGVC. Here, all measurement points were selected for which information was still available, whether the culture status was positive (1) or negative (0). The classification between successful therapeutic and ongoing therapy and the cultural status shows a highly significant correlation with $p<0.001$ (**Figure 10**). Here, it is shown how high the proportions of the final therapy classifications (0= ongoing, 1= calculated end of therapy) were depending on the culture status. No measurement time of patients with positive culture was classified as TB22= 1. Altogether, patients at 105 measurement (20.9%) had a negative culture and were classified as successfully completed therapy. At 193 time points (38.4%), patients had a negative culture but did not reach

the threshold to be classified as successfully completed therapy. At 204 measurement points (40.6%) patients showed a positive culture status and a classification for continuation of therapy.

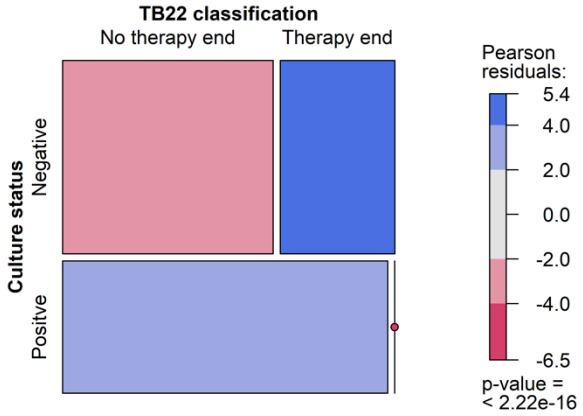

Figure 11 Mosaic plot shows relationship between Therapy End Model (TB22) classification (x-axis) and culture status (y-axis; positive/negative), p-value <0.001 shows a highly significant correlation.

3.3 Model's added value validation and biological plausibility

In this chapter, the model will be examined once again with regard to its added value and biological plausibility. For this purpose, the performance comparison with other published signatures is presented. This is followed by a critical evaluation of the genes taking into account, their pathways, and previous TB-related mentioning in the literature in order to discuss their biological plausibility.

3.3.1 Performance comparison with previously in literature described signatures for clinically relevant endpoints in tuberculosis

16 Signatures that were described previously dealing with TB were used to compare their ability to reflect therapy course [110, 114, 118, 175-179, 181-183, 191-194]. If one authors described several signatures, the distinction was made by numbers of genes in signature, respectively.

Table 4 shows the AUCs of the signatures for the classifications that the end of therapy has been reached as well as the mean monthly changes for the therapy end probability for each tested signature. With regard to the identification of the end of therapy for DS-TB patients, TB22 shows the best performance (**Table 4**).

Table 4 Comparison of the area under the curve (AUC), in drug-susceptible (DS-TB) patients of German validation cohort (GVC), generalized linear model (GLM)-Estimator for monthly changes of probabilities & scores for DS-TB GVC patients and multidrug-resistant (MDR-TB) GVC and Romanian Validation cohort (RVC) patients

	Identification of therapy end in GVC DS-Patients	**p-Value ROC compared to TB22**	**Mean monthly change in therapy end Probability in DS-GVC patients and MDR-TB patients from GVC and RVC patients** ΔP(TE)
TB22	0.94 (0.90 – 0.98)	-	0.023
Anderson44	0.77 (0.68 – 0.85)	0.007*	0.003
Berry83	0.68 (0.57 – 0.78)	<0.011*	0.004
Blankley4	0.77 (0.69 – 0.86)	0.007*	0.013
Kaforou27	0.81 (0.74 – 0.89=	0.047*	0.007
Kaforou44	0.74 (0.64 – 0.84)	0.023*	0.003
Kaforou53	0.79 (0.70 – 0.88)	0.043 *	0.005
Laux da Costa5	0.79 (0.70 – 0.88)	0.045*	0.007
Maertzdorf3	0.76 (0.66 – 0.86)	0.008*	0.009
RISK6	0.70 (0.60 – 0.82)	0.001*	<0.001
Sambaray10	0.75 (0.65 – 0.84)	0.006*	0.003
Singhania20	0.71 (0.61 – 0.80=	<0.001*	0.004
Suliman4	0.65 (0.54 – 0.76)	<0.001*	0.007
Sutherland4	0.56 (0.45 – 0.67)	<0.001*	0.003
Sweeney3	0.75 (0.65 – 0.84)	0.002*	0.007
Thompson9	0.81 (0.71 – 0.89)	0.048*	<0.001
Thompson16	0.62 (0.51 – 0.73)	<0.001*	0.001
Thompson32	0.65 (0.54 – 0.76)	<0.001*	0.004
Zak16	0.78 (0.70 – 0.87)	0.041*	0.004

DS-GVC= Drug-susceptible German validation cohort, MDR-TB= Multidrug-resistant tuberculosis, RVC= Romanian validation cohort, ΔP(TE) = Mean monthly changes for probability of having therapy end

Except for TB22, most of the signatures showed an insufficient increase in the mean probabilities, so that the end of therapy was not correctly detected. This also becomes clear when one looks at

the average monthly probability changes. With an increase of ΔP(TE)=0.023 per month, the TB22 shows the clearest mean increase, while the probability change of the other signatures is clearly below this performance with between ΔP(TE)= 0.003 and ΔP(TE)=0.013. This is also reflected by looking at **Figure 12**. Here, the mean probability for the end of therapy is shown for the six signatures with the highest probability change per month for DS-TB and MDR-TB patients. Table adapted from original publication [143].

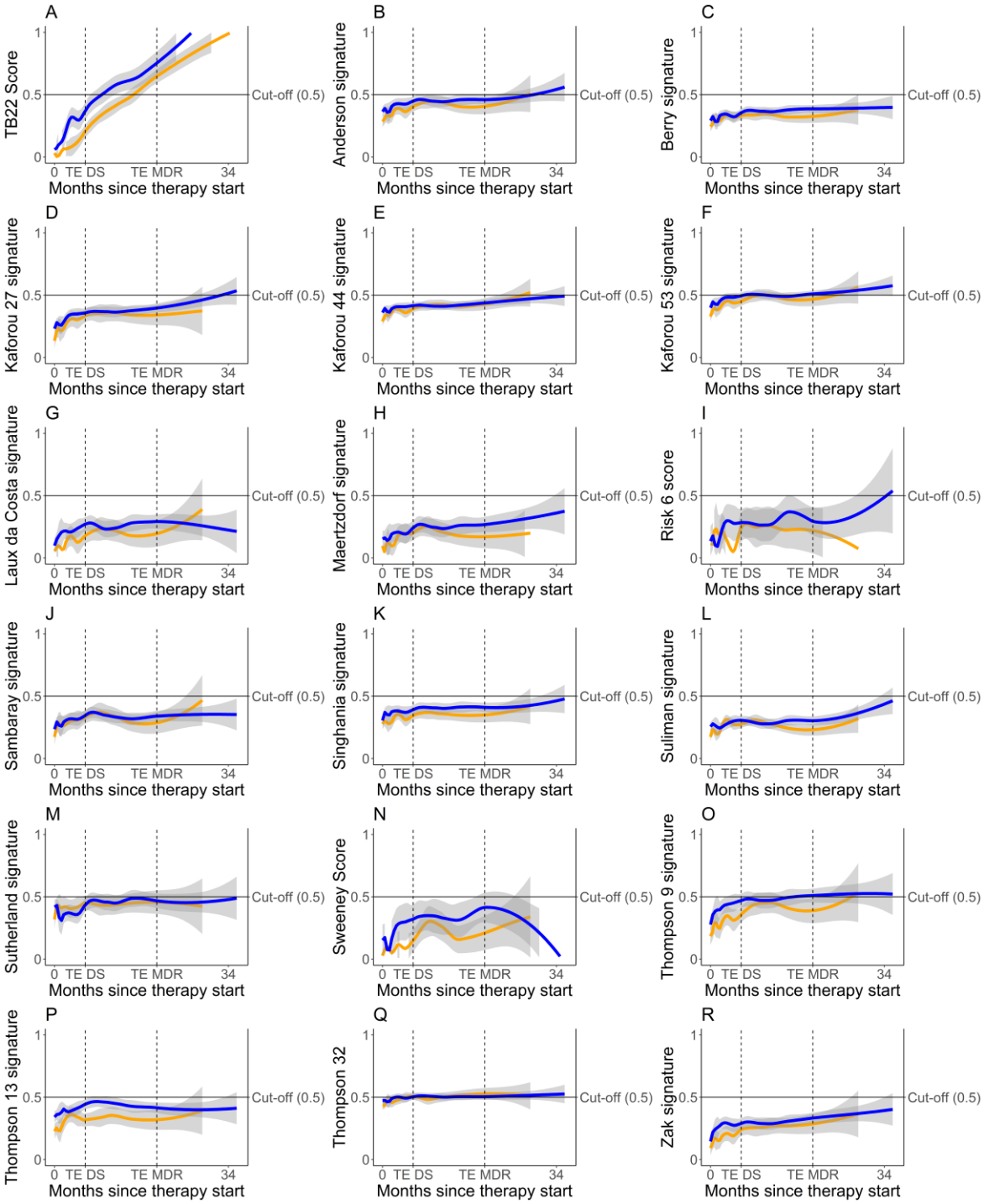

Figure 12 Comparison of the 22-RNA gene signature model with published RNA-signatures and scores to identify end-of-therapy timepoints in drug-susceptible and multidrug-resistant tuberculosis patients from the German Validation cohort (GVC). Calculated therapy end timepoints in drug-susceptible (DS)-German Validation cohort (GVC) patients (blue) and multidrug-resistant (MDR)-GVC patients (orange) for different RNA signatures and scores. Y-axis: probabilities (P[TE]) from random forest models for the classification of the end-of-therapy with a cut-off of ≥0.11, X-axis: time under therapy in months. Figure 12A: 22-RNA gene therapy end model (TB22). Figure 12B: Anderson et al., 43 genes [175]. Figure 12C: Berry et al., 87 genes [176]. Figure 12D: Kaforou et al., 27 genes [177]. Figure 12E: Kaforou et al., 44 genes [177]. Figure 12F:

Kaforou et al., 143 genes [177]. Figure 12G: Laux da Costa et al., 3 genes [178]. Figure 12H: Maertzdorf et al., 3 genes [179]. Figure 12I: Penn-Nicholson et al., 6 genes [117]. Figure 12J: Sambaray et al., 10 genes [194]. Figure 12K: Singhania et al., 20 genes [181]. Figure 12L: Suliman et al., 4 genes [182]. Figure 12M: Sutherland et al., 4 genes [183]. Figure 12N: Sweeney et al., 3 genes [114]. Figure 12O: Thompson et al., 9 genes [118]. Figure 12P: Thompson et al., 13 genes [118]. Figure 12Q: Thompson et al., 32 genes [118]. Figure 12R: Zak et al., 16 genes [110] [143].

3.3.2 Immunological examination

The main aim of this work was to find a gene signature model to calculate the probability that the therapy was successful. 22 genes were identified in total. In this chapter, these RNA targets will be discussed in more detail. **Supplement table 4** shows the fundamental function of the genes as well as their previous TB-related description in literature. Furthermore, respective pathways will be shown. Except A_33_P33271041, KCNJ2-AS1 and TNFRSF21, all genes where formerly descripted at least in being up- or downregulated in correlation with TB-related scientific issues. The numbers and areas of activities vary a lot between the 22 genes. Most of them were placed in at least one of the following subcategories: immune signaling; metabolism; DNA and RNA regulation and repair and/or neurotransmitters. Two genes, STAT1 and TRIM 27, also show an affiliation to tuberculosis related pathways.

After considering the individual genes with respect to their involvement in the respective pathways in **Supplement table 4**, in a second step enrichment analyses were performed on the complete gene set in order to have an unbiased insight into what is important with respect to relevant TB-related questions.

Supplement table 5 shows all involved pathways for respective comparisons. No comparison showed an association with the tuberculosis pathway. However, the results of the *topkegg* function showing the 20 most important pathways of each comparison are shown here. Nevertheless, genes associated with the TB pathway are significantly up- and down-expressed [195]. Furthermore, a total of 10 other pathways have been shown to be significant, assigned to other infectious diseases such as hepatitis B, influenza, SARS-COV19, shigellosis, or Chagas. Those occurred in the comparison of patients with cavitary lung disease vs patients without cavitary disease at baseline, favorable outcomes vs unfavorable outcomes at baseline, before culture conversion vs. after culture conversion therapy naive patients vs. patients after two weeks of therapy, discordance between observed and calculated therapy end, DS-TB patients vs. MDR-TB patients at baseline and healthy controls vs. TB patients at baseline. In addition, 10 pathways associated with the immune system were named. These are Phagosome, mTOR signaling pathway, Neutrophil extracellular trap formation, Toll-like receptor signaling pathway, RIG-I-like

receptor signaling pathway, Natural killer cell mediated cytotoxicity, Th17 cell differentiation, T cell receptor signaling pathway, Fc epsilon RI signaling pathway and Leukocyte transendothelial migration, which were significantly up- or downregulated for therapy naive patients vs. patients after two weeks of therapy, therapy naive patients vs. patients under therapy, patients under therapy vs. patients after therapy end, healthy controls vs. TB patients at baseline, patients with cavitary disease vs patients without cavitary disease at baseline, favorable outcomes vs unfavourable outcomes at baseline and DS-TB patients vs. MDR-TB patients at baseline. For the comparison between healthy controls and patients at the calculated end of therapy, the 20 most important pathways are pentose phosphate pathway, central carbon metabolism in cancer, purine metabolism, glycolysis/gluconeogenesis, glutamatergic synapse, melanogenesis, relaxin signaling pathway, olfactory transduction, ECM-receptor interaction, hedgehog signaling pathway, apelin signaling pathway, autophagy – animal, basal transcription factors, protein digestion and absorption, nucleotide excision repair, DNA replication, autophagy – other, primary bile acid biosynthesis, biosynthesis of unsaturated fatty acids, fatty acid elongation. The last one shows the highest percentage with an upregulation of 25% of all genes involved in the pathway.

In order to provide a better overview, the pathways were summarized with respect to their known function. **Figure 13** shows the mean percentage of genes that were up- or downregulated for the different comparisons with respect to their compressed functionality. Zero (dark blue) means that the pathway was not implicated. It is also clear here that the up- or down-regulation for the discordance comparison between observed and calculated end of therapy in percentage terms is very low overall for all functions. For the comparison between healthy controls and the calculated end of therapy, differences are mainly found with regard to the cell life cycle, DNA and RNA regulation and metabolism. However, hormonal regulation and neuronal signal transmission also stand out.

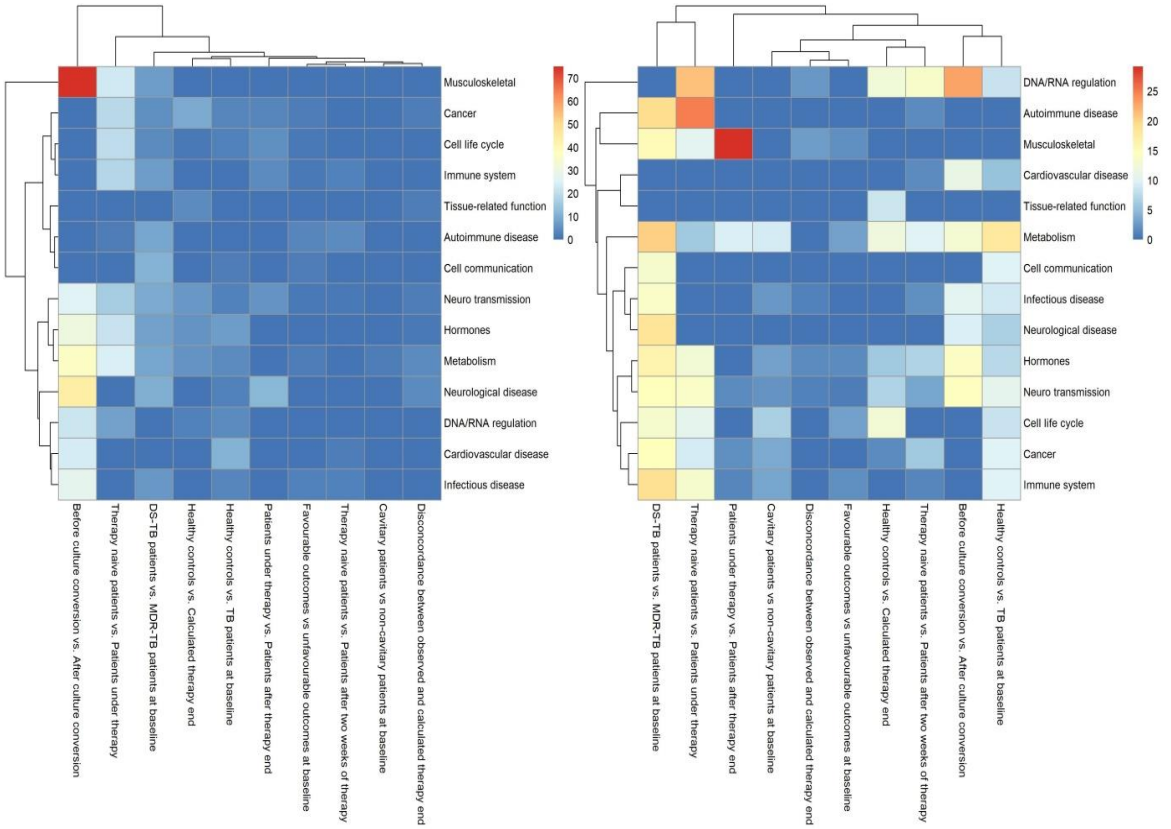

Figure 13 Summary of KEGG Pathway participations and mean percentage of down- or up-regulated genes for various TB-related comparisons.

Both, **Supplement table 4 and Figure 13**, show that the involvement of metabolic, immunological and DNA- and RNA-regulating pathways also playing an important role in disease processes. The comparison of patients with and without cavities; before and after culture conversion and the comparison of ongoing therapy, successfully completed therapies and before the start of therapy show significant up- and down-regulations of genes.

The next step was to further summarize the results. The agreement rate of the pathways between the individual comparison questions was determined. This is shown in **Figure 14**. The pathway agreement of the individual questions with respect to the *toppkegg* results shows a very low agreement between the individual comparisons. The highest agreement was 20% and was found in the comparison between patients before and after culture conversion and therapy-naive patients vs. patients under therapy, as well as in the comparison between cavities and the difference between calculated and calculated therapy end. For the latter, the same pathways are carbon metabolism, type II diabetes mellitus, shigellosis, and thyroid cancer.

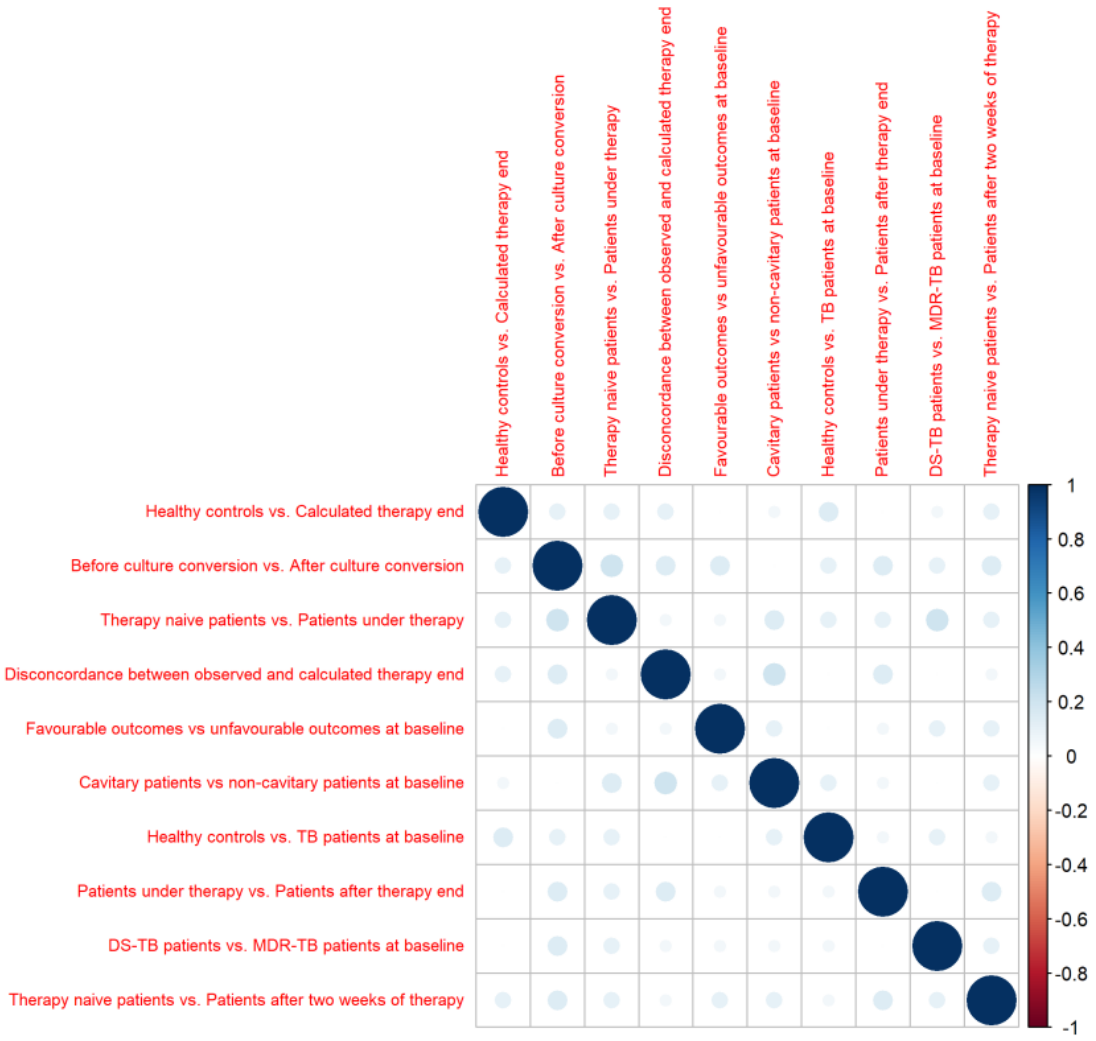

Figure 14 Pathway concordance rates.

Analysis of *tmods* shows compressed results from an algorithmic approach of gene clustering in **Figure 15**. Here, it is shown that in the disease process, modules from the area of T cells, B cells, NK cells, monocytes and other immune system-related modules are predominantly addressed. However, cell cycle regulation and DNA/RNA regulatory aspects are also addressed here. Overall, the results found here are congruent with the KEGG Pathway analysis, but this is a more compact presentation. Whereas in the KEGG analysis the 16 most represented pathways were used in each case, in the *tmods* only the significant modules are used, resulting in a different number of gene clusters involved. It is striking that the discordance between calculated and observed end of therapy shows a difference only in one module. This is the cell cycle. Even more remarkable, however, is that in the comparison between healthy controls and patients from the time of the calculated end of therapy, no module with significantly up- or down-regulated genes could be found. For this reason, this comparison could not be shown in **Figure 15**.

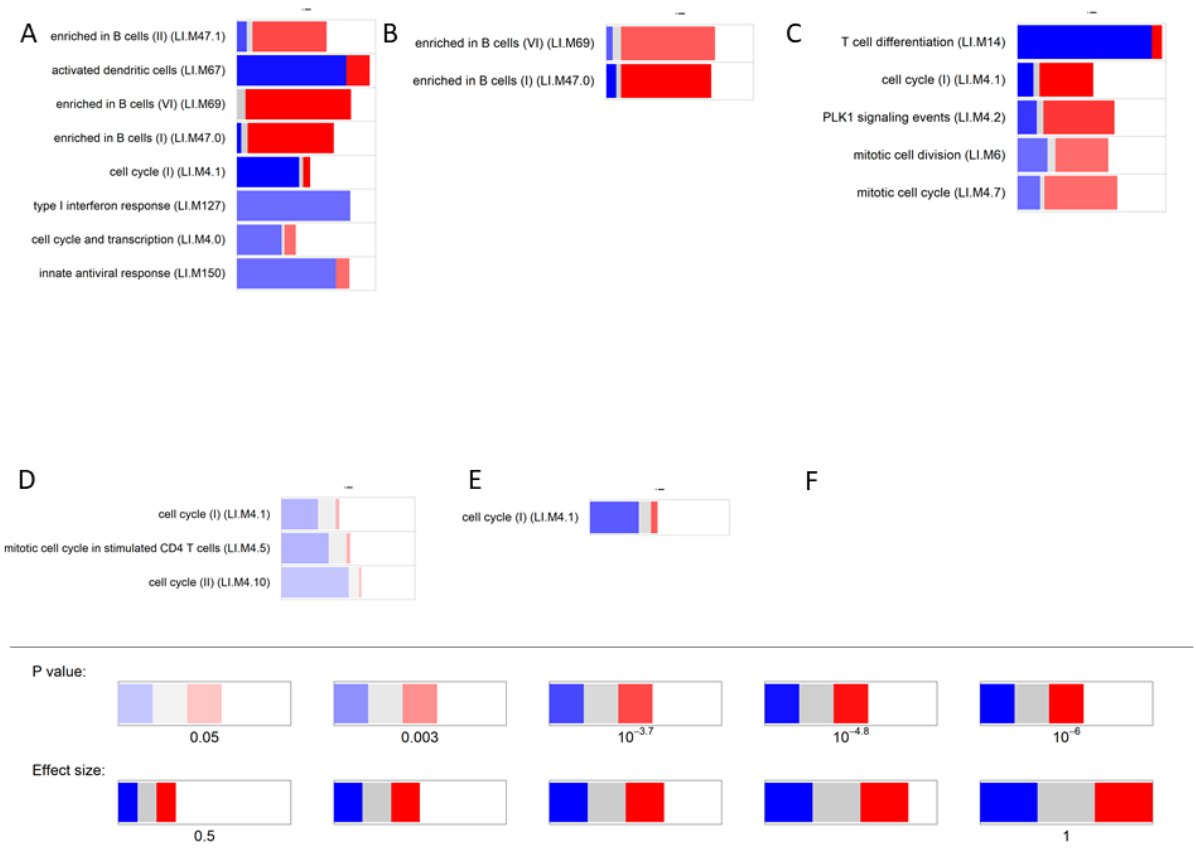

Figure 15 *tmods* analysis 21 A) Healthy controls vs. TB-patients at baseline. B) DS-TB vs. MDR-TB patients at baseline. C) Favourable vs unfavourable outcomes at baseline. D) Patients at observed therapy end vs healthy controls. E) Discordance of observed and calculated therapy end. F) Calculated therapy end vs. healthy controls.

4 Discussion

4.1 Evaluation of results

In contrast to most other infectious diseases TB is characterized by a very long duration of therapy to achieve relapse-free cure. Until now, the recommended duration of antibiotic therapy for patients who developed TB is highly standardized depending on the presence or absence of antibiotic drug-resistance and the extend of the disease. However, in theory, each patient affected by TB needs an individual duration of anti-TB therapy until the disease is cured.

The aim of this thesis is the identification and validation of a biomarker for the bacterial-free diagnosis of TB, for treatment monitoring and to individualize endpoints of therapy where relapse-free cure has occurred in a step-wise approach. First, bioinformatics and biostatistical methods were used to identify a human transcriptional (RNA) signature to distinguish between TB patients from healthy controls. Second, this signature was applied to develop a score using statistical models to represent treatment responses. Subsequently, this score was used to identify individual time points in the course of anti-TB treatment suggesting that a status of cure was reached. To achieve this, two cohorts of TB patients were recruited in Germany, one with DS-TB patients, one with MDR-TB patients (identification cohorts). Whole-blood transcriptome analysis resulted in a 22-gene RNA signature (TB22) to distinguish between TB patients and healthy controls and individuals with other diseases in an external dataset (GSE144127, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144127). This 22 gene signature was also the basis for a multistep statistical to calculate the TB22 score that indicates individual therapy durations in two cohorts, one with patients with DS-TB and one with patients with MDR-TB. Subsequently TB22 was validated in four independent cohorts, one with patients with DS-TB and three with patients with MDR-TB.

The model provides individual probabilities for cure-associated end-of-therapy time points at any given moment throughout therapy, therefore providing unique data for therapy monitoring. The performance to detect the successful end of therapy described by sensitivity/1-specificity is AUC=0.93 in the external validation cohort. In comparison to presently published RNA signatures or scores TB22 shows superiority in identifying end-of-therapy timepoints [143].

In addition, TB22 shows good diagnostic performance for the diagnosis of active TB in differentiation from other diseases and/or latent infection with *M. tuberculosis* [174]. TB22 is non-inferior to the best described RNA signatures for diagnosis of active TB [174]. Additionally, TB22 also shows changes in the course of therapy that correlate with the response to therapy. By exceeding the threshold value of 0.5, the end of therapy in DS-TB patients can be detected with

high accuracy. By using the TB22 model, individual therapy end points can also be determined for MDR-TB. The kinetics in the course of therapy and the accuracy of the end-of-therapy determination are superior to other TB-related gene signatures tested [143].

The gold standard for TB-diagnosis is culture data [196]. Approximately 15-17% of all newly diagnosed pulmonary TB patients have culture negative TB and the diagnosis in these patients is made by clinical considerations [197]. This is especially the case in extra-pulmonary cases and in children [197-199]. Comparing the diagnostic performance of TB22 with the existing clinically used pathogen-based diagnostic tools, the signature identified here is non-inferior. The diagnostic performance of TB22, which in this work is AUC=0.85, is superior to the AUC of the Ziehl-Neelsen microscopy (AUC=0.80) technique [200].

Xpert MTB/RIF and Xpert Ultra are widely used screening tools for diagnosing TB [201]. GeneXpert shows a sensitivity of 90.9% and a specificity of 95.6% for the latest model (Xpert Ultra) and a sensitivity of 8.7% and a specificity of 98.4% for the previous model (Xpert MTB/RIF) [202]. In people living with HIV, sensitivity is 61.8% and specificity 98.8% [48] and shows even better results in diagnosing TB in children [203].

LAM shows sensitivity for pulmonary TB between 34 – 60% in a pooled meta-analysis [222] and differs between TB forms of extra-pulmonary TB between 47 and 94% in people living with HIV [223]. The diagnostic performance of TB22 does not seem to be inferior compared to currently used clinical tests. This can be particularly helpful for the diagnosis of patients with culture-negative TB.

In the recent past, efforts were already made to shorten the standardized therapy durations. Standardized short-course treatments of patients with DS-TB were evaluated for non-inferiority in three clinical trials: REMOX [204], OFLOTUB [205], and RIFAQUIN [206]. While standardized shorter therapies were not non-inferior to standard of care in these trials treatment of pulmonary TB over 4- months with a four-drug regimen of rifapentin, isoniazid, moxifloxacin and pyrazinamide recently showed to achieve comparable cure rates to the standard 6-month treatment regimen and may soon become a new gold standard [207]. For MDR-TB patient, as a result of an observational cohort study performed in Bangladesh [208], shorter durations of therapy for patients with MDR-TB with a fixed combination of medicines over 9-12 months were adopted by the WHO in 2019, providing that patients did not have TB before, that there was no extra-pulmonary involvement, women were not pregnant and there was a limited extend of disease and pattern of *M. tuberculosis* drug resistance [2, 209]. Although promising results with success rates up to 85-90%, less than 5% of all MDR-TB patients worldwide would have the co-morbidity and resistance profile that is expected from treatment according to the Bangladesh regime [210-212]. Furthermore, there is the possibility to administer a shortened all-oral

bedaquiline-based therapy regime for about 9-12 months [2]. Currently, further non-inferiority trials are underway to test the efficacy of new approaches with shortened treatment durations [209, 210]. Therapy reduction has also been recommended by the WHO for the use of delamanid and bedaquiline. However, this is only recommended if there is no resistance to fluoroquinolones, and no second-line medication may have been administered in advance [2]. There have been efforts in the past to reduce the duration of therapy for DS-TB patients with culturally and smear negative, drug susceptible TB to four months [69]. However, a number of further requirements are necessary for this; most importantly, the absence of any co-morbidity. Although some of the requirements for therapy shortening correspond with the study characteristics of this work were not tested here - namely the absence of HIV infection, the limitation to pulmonary TB and the age of majority of the patients, the TB22 offers a potential tool for condition-free individualized therapy duration compared to generalize therapy duration shortening [36, 157-160]. Therefore, the standardized short-term treatment for MDR-TB patients is only suitable for a marginal group of patients, but does not provide a solution for the majority of patients who still need to be treated for 18-20 months. The "one-fits-all" therapy duration of the current standard therapy is shortened according to certain criteria and then applied in a standardized manner to the appropriate group of patients [2].

In comparison, TB22 score offers individualized therapy duration instead of a standardized therapy shortening, which, at least in the cohorts reviewed here, is accompanied by a therapy shortening compared to the standard therapy for most patients.

Based on the procedure used here to determine the individual end of therapy, there is a kinetic in the course of TB22 that indicates a response to therapy in patients. Kinetics due to changes in the course of therapy were also found in other immune-based test procedures, which were initially investigated primarily for their diagnostic capability [213, 214]. The IGRA test used for the diagnosis of LTBI also shows some potential for use as a biomarker in therapy monitoring. However, the results found in a systematic review on the changes during therapy were controversial [215]. While some studies found a decline in IGRA over the course of therapy, others found no changes; a few studies even reported an increase [215]. However, if one compares the individual values of the IGRA during therapy, arbitrary courses become apparent [216].

The Tam-TB test, which was developed for culture-free diagnosis, especially in children and extra-pulmonary therapy, shows a dynamic course during therapy and could serve as therapy monitoring [217]. However, the time points studied show relatively large intersections in the TAM-TB results [217]. Further studies, possibly including individual courses, would be desirable to assess the suitability of the TAM-TB as a therapy monitoring tool.

Another possible immunological approach to therapy monitoring is fluorescence-activated flow cytometry (FACS). Here, cell populations and surface markers of the immune cells are measured. With regard to TB, a FACS therapy monitoring system that usually consists of a combination of different populations and markers, could distinguishes between LTBI and active TB as well as between active TB and treated TB [218, 219]. Despite significant differences, there are still relatively large overlaps between the groups and continuous measurements for concluding performance assessment is missing [218, 219].

The TB22, on the other hand, reflects the disease severity and the patient's response to therapy on an individual level. There is a clear cut-off of TB22=0.5 to distinguish between treatment-naive and treatment-responsive TB, with no overlap of numerical values between the two conditions. Also, directional changes clearly indicate a worsening or improvement of the disease state and do not show random movements. The TB22 score could therefore represent a therapy monitoring tool that is similarly easy and unambiguous to interpret as culture data, but can be used in the case of culture negativity additionally.

Transcriptomes are well suited for real-time therapy monitoring and the determination of individual therapy duration for two reasons: On the one hand, RNA is particularly suitable as a biomarker because they are relatively easy to detect and quantify [104]. It can provide a very dynamic insight into cell status and process flows, which is not the case with DNA [101]. Also, the fact that multiple copies of RNA are present in cells provides a higher level of information content than DNA, which is present in a single version in the cell [101]. Furthermore, RNA offers higher sensitivity and specificity compared to other systems biology data, such as proteomics [101, 220]. Transcriptomes are used to identify expression profiles that allow conclusions to be drawn about specific processes. Mutations or epigenetic changes that allow an increased basic risk for certain diseases are not considered here [101]. It is above all the potential of having a pathogen-free tool that makes the analysis of transcriptomes so attractive for clinical use. Therapies can be initiated more quickly and the M. tuberculosis transmission rate could be reduced [221].

The TB22 as a score based on transcriptomic data was developed using a multistep approach that accommodated three different foci that were merged into a single score in the end. The severity of the disease, the progress of the therapy and the successful completion of the therapy are the cornerstones of the score, which was created using machine learning algorithms.

The data preparation in this work was done using *limma*, a validated package of the program *R*. It includes tools that allow stable analyses and is well suited for both complex and small data sets [105, 222].

A mix of methods was used; while the genes of the TOS were identified by moderated t-test and Benjamini-Hochberg correction, the identification of the relevant genes for the TPS and the TB22 was based on Lasso regression. The t-test describes the expression levels of the genes in isolation from each other and the usage of Benjamini-Hochberg shows a low misclassification rate [150]. Statistical filtering methods such as the identification of statistically significant genes by corrected T-test, partly including a log-fold threshold, have already been described frequently for transcriptome studies in tuberculosis [176, 179, 180]. Although the method is well suited for variable reduction, no variable dependencies are examined, which is why in this work the important variables are identified for the last model step using lasso regression, which places the genes in context with each other and also takes multicollinearity into account [223]. Genes that have no effect will be excluded from further analysis, which is particularly helpful in large datasets where qualitative variable selection by hypothesis is not feasible [168-171, 223]. The method mix opens up the possibility of identifying a large number of genes by different methods. By using the T-test, important genes with high differences in expression levels were identified. In addition, lasso regression allowed the inclusion of genes that would not have been identified by the T-test but are important as part of a highly multi-correlated data set, since gene expressions in signal cascades in transcriptomic processes are highly dependent from each other [224, 225].

A general problem in modelling is overfitting. First, from a qualitative point of view, it is better to look only at the variables that make up a real effect in order to understand the - in this case biological - correlations. If there are too many variables in the model, the estimation parameters reflect an arbitrary distribution rather than the actual influence. Second, economic considerations also play a role in the further application - the analysis of fewer genes is much less expensive, and the model runs much faster and results can be determined quicker. Last but not least, accuracy is significantly increased if only variables of actual relevance are included in the model. With many unnecessary variables, genes that have no actual influence are given unnecessarily much influence, while the effect of actually significant genes would be massively underestimated in this case. This would have the consequence that the model would show excellent results in the training data set, but that these results could not be replicated on other data sets and would therefore become useless for the actual research question [135, 226, 227]. In the field of transcriptome research, different methods can be used to reduce the variables reduction. These can be unsupervised clustering, the application of multiple corrected tests or penalized regression methods [228-230]. In this work, LASSO regression was used as a method of penalized regression. The idea behind penalized regression models is that the estimator of each variable is shrunk, leaving only those that have a real effect and accounting for the multicollinearity of the variables [137, 230]. Genes that exhibit high multicollinearity are thus dropped from the data set [231]. The

LASSO regression can shrink the estimator down to 0, which would make the respective variable insignificant for the model and is a well suited approach for genetic high-dimensional data [232]. Ridge regression as another form of penalized regression would not have been appropriate for this work, as with this method the estimator always remains >0 and variables are not eliminated [233]. This method would therefore not be suitable for pre-selecting important variables. Elastic net regression as another form of penalized regression, which is a mixture of LASSO and ridge regression, is also not as suitable as LASSO regression due to the reduced variable reduction rate [233]. Elastic net was already previously used for transcriptomic variable selection [175], however, the resulting signature contains 44 genes and is thus significantly larger [175]. Furthermore, the elastic net regression was not applied there as a first step for variable reduction, but for model stabilization [175]. In Anderson et al. the dimensionality reduction took place with a principal component analysis (PCA) [175], in which the covariance matrix is generally considered difficult to interpret [234]. Furthermore, the detection of invariances is fraud with risk and requires a larger training data set with explicit information, which means that the PCA method would not have made sense for the present work [234]. Singhania et al. also used a network analysis, which resulted in a correlation matrix of the 5000 most important genes determined by log2 fold change and could take into account biological connections in the form of formed modules from individual genes [181]. However, the actual performance and identifiability of module clusters is not clear [235], which is why module-based analyses were not used as a preselection tool in this work, but rather for validation.

It was also shown that the LASSO regression has the highest accuracy especially in genetic variable selection [236]. LASSO regression is a widely used method in transcriptome research [237-241] and other genetic studies, such as GWAS [242]. Sambaray et al. used a different approach to pre-select by setting up their model as a Human Protein-Protein Interaction Network [194]. This incorporates interactions of the individual proteins into the model and gives the model an additional component of biological significance instead of just statistical significance only [194]. But this kind of network modelling requires a huge amount of very specific data, which is not the case when using classic statistical and machine learning algorithms like t-test or LASSO regression [194].

Stepwise regression, which was used for further variable reduction, primarily provides a quick procedure for selecting the final model. Variables are removed as long as model performance is maintained. If the performance drops significantly with further exclusion of variables, the final model was identified. The stepwise regression used here is based on AIC. Although it only takes statistical significance into account and ignores biological aspects, it is particularly useful for

complex causal variables such as those in this work [243]. Stepwise regression was also used for only 2 of the 3 models in this paper. Variable selection for the final model was done using random forest. Here, variable importance is determined in different sub-models to have an accurate representation of the importance of individual variables. The importance offers a very sensitive measurement for variable selection, which is particularly limited when different variable types are used in the model [244]. The random forest algorithm of Breiman and Cutler was introduced in 2001 and is based on the classification tree approach [158, 245]. It is a black box classification test method, where each classification consists of an arbitrary set of disjoint expressions of parameters. The parameters considered relevant are used to create the classifications [158, 245]. Random forest represents a combination of different classification trees, all interacting independently. The advantage of the random forest algorithm is that there is no need to make any prior statistical assumptions about the distributions of the individual variables [246]. This provides much greater flexibility with regard to the weighting of individual variables and the combination of the individual variables with each other [247]. Although random forest algorithms require a high computing capacity, which increases exponentially with the number of variables, this limitation is countered by the technical development, which allows computing capacity for increasingly complex algorithms [247]. The quality of random forest algorithms is strongly dependent on the input quality of the data. While this is generally the case for all machine learning algorithms, the recombinant nature of random forest trees makes data quality even more important for this algorithm [246]. However, this is not the case in this work. Interestingly, previous TB-related transcriptomic signatures were also identified using random forest [178, 179]. Compared to other binomial regression methods, this shows high accuracy [248], but the reproducibility of the algorithm is severely limited [249]. For this reason, the results of the random forest algorithm were translated into a model that offers fixed estimators. The use of logistic regression as alternative to random forest to identify a TB signature has also been described [183]. However, this algorithm offers significantly poorer performance compared to the random forest and was therefore not used in this work [250].

Even after dimension reduction, probability of finding random results with statistical significance in these high-dimensional data sets is very high given the number of variables. For this reason, there is a need to validate the findings from the original dataset. For this purpose, the original data set is divided into a training set and a test set. The model is trained using the first data set and then tested in the test set. However, there is also a risk that the results in the test set may be randomly replicable. This may be due to similar patient structures or similar stitching of the observations in a data set to each other, although this was not done for some TB-related

signatures described in other studies; instead, an existing data set was split into training and test set only without further external validation [175, 176, 178]. In summary, it is helpful to plan biomarker studies or "OMICS" studies for a multi-centre study design in order to provide a certain variance in the data set. However, this is not sufficient to eliminate the risk of randomly significant results that are subject to the error of second kind. For this reason, results from "omics" studies should without exception be replicated in independent validation cohorts before they are published [135]. A good transcriptome signature is defined by being as small as possible, i.e. containing as few genes as possible, and having a high discriminatory power between categories of interest [251].

Due to the standardized, protocol-based blood sample collection and sample processing, as well as the use of a standardized preparation pipeline through the use of *limma*, this work met the requirements of the random forest algorithm to the greatest extent possible [252].

In some studies, the methodology was not described in enough detail to compare the procedure with the methods used here [110, 114, 117, 118, 182, 253, 254].

It was found that the calculated and observed durations of therapy in DS-TB patients showed a high concordance. Furthermore, it became clear that the observed and calculated durations of therapy were similarly influenced by clinical parameters, such as the presence of cavities. Although the smear grade showed no significant effects here, a trend in the courses was nevertheless discernible.

The genes included in the therapy end model are involved in different functional signaling pathways and cannot be connected to a single functional background. All genes that are part of the therapy end model, except KCNJ1-AS1, PDE4D, TMFRSF21 and A_33_P33271041, were previously described as part of host responses to TB [176, 177, 181, 193, 251, 254-259]. The genes that were identified for the therapy end model belong to several signaling cascades (e.g., related to metabolism, cell signaling, DNA repair and RNA transport), which reflects the complexity of individual treatment responses for the host. Supplement table 6 displays a list of all genes of the TB22 score with an overview of which TB- or immune-related studies they have already been described in and a list of all pathways in which the respective genes are involved.

4.2 Limitations and strengths

This thesis has several limitations. The number of cases in the identification cohort used for model training is relatively small, which implies a risk of overfitting [260] although efforts were made to reduce this risk as much as possible and to validate the findings. Furthermore, there is a whole series of patients in this cohort for whom no measurement time is available. This and the fact

that, as expected, later time points are missing for deceased patients in whom a severe disease manifestation can be suspected, can lead to a model bias in which patients with a complete data series are given a higher weight than persons with missing measurement time points [261, 262]. The signature identified and verified in this work contains 22 genes. Compared to other signatures that already use single-digit gene signatures, this number is relatively high. However, genetic information was aggregated into sub-models. With this approach, numbers of variables in the final were reduced by compressing, which also reduces the risk of severe overfitting. In addition, a mathematical strengths of this work lay in the presence of multiple validation cohorts [136]. This can reduce the risk of severe overfitting as well, as the application to the validation cohorts with poor fitting could have already given indications of this. Furthermore, one cohort was enrolled in another European country, which reduces the risk of an unidentified centre bias, since patients from one related to one centre might be more similar to each other than patients from another centre [263].

Furthermore, the translation of outcomes - cure, failure, death - into numerical expressions as a basis for TOS modeling is an artificial approach. With regard to the creation of the TOS, low number of cases of patients who died and patients with treatment failure must also be mentioned. But TOS is also to be regarded as a variable in the TB22, which was only designed as an intermediate step. This reduces the problem of the artificial approach and the low number of cases due to the application restriction. Moreover, the model described here was created and evaluated using microarray transcriptome analysis, while the RNAseq procedure is actually state of the art. In this work, microarrays were used because at the beginning of the patient recruitments in 2013, the RNAseq method was not yet so widespread and microarrays were used more frequently. The method has not been adapted over the years to have a consistent method for evaluation. In the current non-inferiority study, mainly RNAseq data are evaluated in addition to microarrays as reference.

An advantage of this approach is that measurements can be taken at any time without any interference between previous measurements. Each measurement is completely independent of previous results. Although the inclusion of previous results and progressions could increase the precision and would also be able to make future-oriented prognoses, it seems impractical for daily clinical practice to work with a model that is dependent on strict measurement schedules. However, making predictions using a prediction model based on probability calculations of previous measurement dates is in principle possible, regardless of the model premise of independence of probability for the original calculation. This can be a complementary aspect when validating the model in a non-inferiority study.

One of the advantages of the work presented here is the different validation cohorts that are available. [175, 176, 178]. However, the disadvantage here is that the test set is too similar to the training data set and thus the performance is overestimated. In this work, however, the performance in an actual external data set was decisive, which makes the estimation of the actual performance more robust. Although the fitting was not 100% consistent in the data set of DS-GVC patients, the value shows a clinical plausibility. The DS-GVC included many cases of patients with drug susceptible TB but severe disease progression who were treated for more than 6 months. The additional time was estimated on the basis of expertise. For these patients, too, there is no biomarker that could reliably determine individual end of therapy time points being associated with relapse free cure. With regard to the difference between observed and predicted therapy duration, the result of the model nevertheless appears conclusive. Compared to most other published marker combinations, the findings were affirmed by considering various established clinical endpoints such as smear and culture status, radiological findings, and strict outcome criteria [142, 146, 264], which include a follow-up period of one year after completion of therapy to capture disease recurrence. In contrast to the other signatures included in the comparison, the model was specifically trained to identify end-of-therapy time points, this distinction also explains the model's superior performance when compared to other published signatures that were mainly developed to predict the future onset of disease and to diagnose active TB rather than for the conduction of individualized therapy durations [110, 114, 117, 118, 175-183].

A limitation of this work is that clinical data are not fully available for all patients. Furthermore, not all clinical aspects could be tested extensively, as the group size was too small for comorbidities or certain sociodemographic and behavioral aspects (e.g., intravenous drug use, problematic alcohol use, diabetes) or did not meet the study inclusion criteria from the outset (HIV-positive patients, children). Furthermore, except TCC, clinical parameters that were available at the beginning of the therapy were used for plausibility examination without follow-up of those clinical aspects. The signature presented here mirrors the therapy response. This can be explained not only by the occurrence of individual risk factors, but also by the presence of several risk aspects and also by events during therapy, such as interruptions due to lack of compliance or side effects. Changes in dosage or medication can also have effects that are more serious than individual risk factors. The creation of complete risk profiles and the determination in multiple regressions was not possible due to the small number of cases.

Another aspect that has to be addressed is that mortality in both German cohorts is higher than the average [33]. One reason for this could be that most patients of the German cohorts were recruited in the Research Centre Borstel. The hospital of the Research Centre Borstel is specialized in the treatment of TB. As a result, mainly severe and/or multidrug-resistant cases are referred

there, which explains the increased mortality. There were no data on mortality among patients of the RVC. But relapse cases are missing in the cohorts to verify whether the calculated end of therapy would start later than the observed end of therapy before relapse. Furthermore, the number of failure cases due to delayed culture conversion is relatively small. The TB22 performance for this accordingly requires further validation. The cohort composition of patients with DS-TB and MDR-TB patients also does not correspond to reality in its ratio [6, 32]. The lost to follow up rates in our cohorts are much higher than in WHO TB report [6]. The reason for this is that the TBnet criteria for the treatment outcome in TB were applied in this study, which includes a follow-up period of one year after the successful completion of therapy. Also, the lost to follow up in this study is not synonymous with an overall lost to follow up with respect to the disease in general, but rather implies that the study was terminated by the patient or that no further samples were collected [3, 4]. Reasons for this include the fact that many patients did not come from Germany and (had to) leave the country in the course of therapy or in the subsequent follow-up period. In addition, many patients left the area surrounding the hospital after the hospital stay. Follow-up general practitioner care was not always stringent or contact with the treating general practitioner could not be established.

Nevertheless, the close integration of transcriptome data with extensive patient data from the database must be considered an unique strength of this work. The model could be validated with respect to culture data, which is currently used as a standard therapy monitoring tool. This includes treatment response (TCC) and initial bacterial load and disease severity (TTP+, sputum smear grade). In addition, imaging results could be used as a clinical plausibility tool.

Furthermore, all cohorts seem to be representative for TB patients without HIV co-infection. In Germany as well as worldwide, TB mainly affects the age group of 20-40 years (except in the Western Pacific region, where mainly over-65 years of age are affected) [6]. The mean age of the cohorts considered here corresponds to this relatively young average age of the disease. The fact that there are more men than women in the cohorts also reflects well on the total population of patients [6].

5 Summary and outlook

In conclusion, a host 22-gene RNA-based model was prospectively identified in two cohorts and subsequently validated in four additional cohorts with avtive TB that categorizes patients with active TB and that may predict individual treatment durations for patients treated against drug susceptible and MDR-TB. Application of this model may potentially shorten treatment duration in the majority of patients with MDR-TB and may have substantially impact on clinical management. The method of comparing transcriptomic signatures not only in a descriptive way but also testing their ability to discriminate between two outcomes is a previously unused approach for transcriptomics in translational medicine. The presented data are plausible and the validation of the model is based on clinical and microbiological findings. Nethertheless, the model's translation into clinical practice will require further evaluation in larger studies and diverse patient populations and the development of an implementable platform to support feasibility in resource limited settings.

To ascertain superiority of a biomarker-guided approach to individualize the duration of anti-TB therapy, a prospective, non-inferiority trial has been initiated (ClinicalTrials.gov Identifier: NCT04783727). This study includes centres in Germany, Ukraine, Moldova and Romania for enrollment of MDR-TB patients. The primary aim of this two-armed clinical trial is to compared he rate of relapse in patients with MDR-TB treated as long as the TB22 model suggested vs standardized treatment duration as recommend by national treatment guidelines.

In addition, it will be evaluated whether the TB22 signature model may also have a role to guide the decision for duration of treatment in TB-patients living with HIV or immune-compromising conditions, children with TB and patients with extrapulmonary TB. Furthermore, it needs o be evaluated, whether the TB22 signature model is also applicable to non-caucasian TB-patient population. [174].

Ausführliche Zusammenfassung als deutsche Beilage zur Doktorarbeit

Tuberkulose ist die häufigste zum Tod führende bakterielle Infektionskrankheit weltweit. Die Krankheit wird durch eine Infektion mit einem Bakterium, *Mycobacterium tuberculosis*, verursacht. Die Weltgesundheitsorganisation (WHO) schätzt, dass etwa ein Viertel der Weltbevölkerung mit *M. tuberculosis* infiziert ist, wobei nur ca. 5 - 10% der Infizierten eine Tuberkulose entwickelt. Jährlich erkranken etwa 10 Millionen Menschen an einer Tuberkulose. In ca. 5 % der Fälle sind die Bakterien gegen die beiden best-wirksamen Medikamente, Rifampicin und Isoniazid, resistent (multiresistente Tuberkulose [MDR-TB]).

Wenn die Tuberkulosebakterien gegen alle Medikamente empfindlich sind beträgt die Mindestdauer der Behandlung mit einer Kombinationstherapie aktuell 6 Monate. Für die Standardtherapie der MDR-TB empfehlen die Leitlinien aktuell 18 Monate einer Kombinationstherapie mit Zweitlinienmedikamenten. Die Behandlung der MDR-TB ist nicht nur langwierig, sie ist auch mit hohen Kosten und einer hohen Rate von medikamentenbedingten Nebenwirkungen verbunden. Die Dauer der Therapie folgt aktuell einem Einheitsansatz, da es keine Biomarker gibt, um die Dauer der Therapie zu individualisieren. Obgleich es kulturelle Nachweismethoden gibt, welche auch den Goldstandard in der Diagnostik darstellen, stellt die Bakterienkultur im späteren Therapieverlauf keinen geeigneten Marker für die Gesamtdauer der Therapie dar, da ein Bakterienwachstum in der Regel nur zu Beginn der Therapie nachweisbar ist. Der Zeitpunkt, an dem dieses Wachstum nicht mehr nachweisbar ist, die sogenannte Kulturkonversion, bedeutet nicht das Ende einer erfolgreichen Therapie.

Die lange Dauer der MDR-TB wird empfohlen, um die Rate der Patienten, die erfolgreich behandelt werden, optimal hoch zu halten. Sicherlich werden die meisten Patienten länger behandelt, als es notwendig wäre, um eine Heilung zu erzielen. Unter bestimmten Voraussetzungen empfiehlt die WHO aktuell auch kürzere MDR-TB Therapieregime, allerdings kommen diese aufgrund von Ausschlußkriterien kaum für Patienten in Westeuropa in Betracht. In diesem Kontext ist es wichtig zu betonen, dass diese Empfehlungen der WHO auch eine standardisierte Behandlungsdauer bedeuten. Biomarker zur Individualisierung der Behandlungsdauer der MDR-TB stehen aktuell nicht zur Verfügung.

In den letzten Jahren mehren sich die wissenschaftlichen Erkenntnisse, dass RNA-Profile aus dem Blut von Patienten dazu geeignet sind, einen Progress von einer latenten Infektion mit *M. tuberculosis* zur einer aktiven Tuberkulose vorherzusagen. RNA-Signaturen eignen sich außerdem zur bakterien-freien Diagnose einer Tuberkulose und könnten geeignete Marker für ein Therapiemonitoring darstellen.

Ziel dieser Arbeit ist die Entwicklung eines RNA-Signatur Modells für eine individualisierte Therapiedauer der MDR-TB. Hierfür wurden Patienten in sechs unabhängigen Kohorten rekrutiert: Patienten mit sensibler Tuberkulose (drug-sensitive German identification cohort DS-GIC, n=49; drug-sensitive German validation cohort DS-GVC, n=32), Patienten mit MDR-TB (multidrug-resistant German identification cohort MDR-GIC, n=30; multidrug-resistant German validation cohort MDR-GVC, n=21; multidrug-resistant Romanian validation cohort MDR-RVC, n=52; multidrug-resistant second German validation cohort MDR-SGVC, n=20). Nach Prüfung von Einschluß und Ausschlußkriterien und Einwilligung der Patienten wurde ihnen vom Zeitpunkt der Diagnose bis ein Jahr nach Beendigung der Therapie Blut zu RNA Microarray Analysen abgenommen. Die Zeitpunkte umfassen für die Patienten mit sensibler Tuberkulose den Therapiestart, 14 Tage nach Therapiestart, den Zeitpunkt der mikroskopischen Konversion, den Zeitpunkt der Kulturkonversion, 6 Monate nach Therapiestart und – falls die Therapie nicht nach 6 Monaten beendet werden konnte, eine zusätzliche Visite zum Therapieende. Bei der Validierungskohorte wurde weiterhin ein Jahr nach Therapieende eine Visite initiiert. Die Kohorten mit Patienten mit MDR-TB hatten weiterhin planmäßige Visiten nach 10 Monaten, 15 Monaten und 20 Monaten nach Therapiebeginn. Auch hier wurden zusätzliche Visiten initiiert, wenn die Therapie nach 20 Monaten nicht beendet werden konnte und die deutschen Validierungskohorten wiesen ebenfalls zusätzlich eine Visite ein Jahr nach Therapieende auf. Daten aus der Identifizierungskohorten dienten als Trainingsdatensätze, in welchen die Algorithmen entwickelt wurden, während aus den Validierungskohorten unabhängige Datensätze generiert wurden, um die entwickelten Algorithmen in externen Datensätzen zu überprüfen. Außerdem wurde ein weiterer Datensatz generiert, um Patienten mit einer Tuberkulose von Patienten mit anderen Krankheiten und gesunden Kontrollen zu unterscheiden (Hoang-Datensatz). Weiterhin wurde die diagnostische Genauigkeit der hier entwickelten RNA Signatur zu den in der wissenschaftlichen Literatur beschriebenen Signaturen verglichen.

Der Algorithmus zur Identifizierung einer RNA Signatur wurde in einem mehrstufigen mathematischen Ansatz mithilfe der Programmiersprache R entwickelt. Im ersten Schritt wurden Gene identifiziert, welche zu Therapiebeginn unterschiedlich zwischen Patienten mit erwünschtem (Heilung) und unerwünschtem (Therapieversagen und Tod) Therapieausgang unterschieden und nach Variablenreduktionsverfahren verblieben sechs Gene, welche die Basis für einen Punktwert zur Erfassung des Erkrankungsschweregrades gemäß dem Therapieausgang bildeten. Im zweiten Schritt wurden Gene identifiziert, welche mit zunehmender Therapiedauer unterschiedliche Expressionslevel aufwiesen. Hier wurden die Anzahl der Gene auf acht reduziert, die den Therapiefortschritt als Punktwert auf Basis der Therapiedauer angeben. In einem dritten Schritt wurden Gene identifiziert, die bei Patienten mit sensibler Tuberkulose während der Therapie und nach erfolgreicher Beendigung der Therapie unterschiedlich exprimiert sind. Diese Liste von Genen

wurde gemeinsam mit den Punktwerten errechnet aus den beiden oben beschrieben Verfahren zur Krankheitsschwere und dem Therapiefortschritt einem Variablenreduktionsverfahren unterzogen, um die wichtigsten Punktwerte und Gene beizubehalten, um zwischen erfolgreichem Therapieende und laufender Therapie bei Patienten mit sensibler Tuberkulose zu unterscheiden. Neun Gene sowie die Punktwerte zur Krankheitsschwere und dem Therapieverlauf verblieben im finalen Modell. Einige Gene waren sowohl bei der Errechnung des Therapiefortschritts als auch für die Unterscheidung zwischen laufender und erfolgreich beendeter Therapie von Bedeutung, sodass insgesamt 22 Gene für den Algorithmus der vorliegenden Arbeit von Bedeutung sind. Der Algorithmus liefert einen Punktwert, den sogenannten TB22, welcher mittels Schwellenwert von 0.5 einen Rückschluss darauf zulässt, ob die Therapie theoretisch beendet werden könnte oder weitergeführt werden muss.

Bei der Anwendung des TB22 Algorithmus in den unabhängigen Datensätzen der Validierungskohorten zeigte sich zunächst, dass TB22 bei Patienten mit sensibler Tuberkulose mit einer area under the curve (AUC) von 0.937 eine sehr hohe Genauigkeit bei der Unterscheidung für die laufende und erfolgreich beendete Therapie aufweist. Bei der Nutzung von TB22 bei den Kohorten mit einer MDR-TB zeigte sich eine potentielle mittlere Therapiereduktion um 218 Tage in der MDR-GIC, um 211 Tage in der MDR-GVC, 161 Tage in der MDR-RVC und um 254 Tage in der MDR-SGV. In keiner Kohorte wurde der Schwellenwert für ein potentielles Therapieende überschritten, während noch positive Kulturen vorlagen. Die errechnete Therapiedauer zeigte weiterhin Korrelationen zur Krankheitsschwere, welche mittels klinischer Aspekte wie Kavernen, Bakterienlast und Zeit bis zur Kulturkonversion abgebildet wurden.

Um den zusätzlichen Nutzen der identifizierten Gensignatur und des TB22 zu ermitteln, wurden verschiedene bereits in der Literatur beschriebenen Tuberkulose-bezogenen RNA-Signaturen identifiziert und entweder als Genset oder der Beschreibung in der Literatur entsprechend als Punktwert ebenfalls im Trainingsdatensatz der Patienten mit sensibler Tuberkulose trainiert und anschließend auf die anderen Kohorten angewendet. TB22 zeigte dabei die größte Genauigkeit zur Therapieunterscheidung zwischen Patienten mit sensibler Tuberkulose während und nach der Therapie. Außerdem handelte es sich hierbei um das einzige Verfahren, welche einen Anstieg während des Therapieverlaufs aufweist, während andere Signaturen und Punktwerte eher stagnierende Verläufe anzeigten. Auch eine zusätzliche Überprüfung der TB22 Signatur hinsichtlich der diagnostischen Eigenschaften im Hoang-Datensatz zeigte, dass die TB22 Signatur sowohl bei der Identifizierung von Tuberkulosepatienten von anderen Erkrankungen als auch bei der Unterscheidung von latenter Infektion mit *M. tuberculosis* und aktiver Tuberkulose den besten bislang publizierten Signaturen vergleichbar ist.

Systembiologische Analysen zeigten, dass die identifizierten Gene keine Unterscheidung zwischen Patienten mit MDR-TB und sensibler Tuberkulose aufzeigen. Zum Zeitpunkt des mittels TB22 kalkulierten Therapieendes unterscheidet sich das Expressionsniveau der TB22 Gene nicht mehr zwischen den Tuberkulosepatienten und gesunden Probanden.

Als weitere Schritte stehen nun die Untersuchung des Biomarkers TB22 in einer nicht-Unterlegenheitsstudie gegenüber dem Therapiestandard mit festgelegter Behandlungsdauer an (diese Studie ist bereits angelaufen). Es muss außerdem geklärt werden, ob die bisherigen Ergebnisse auch auf andere Patientengruppen, z.B. Menschen mit einer HIV-Infektion, Kindern oder Menschen aus verschiedenen Regionen der Erde übertragbar sind.

Zusammenfassend wurde mit dem TB22 RNA-Signatur Modell in der hier vorliegenden Arbeit erstmalig ein Biomarker entwickelt, um die Dauer der Tuberkulosetherapie zu individualisieren. Nach prospektiver Identifizierung der TB22 Signatur in zwei unabhängigen Kohorten wurde das RNA-Signatur Modell an vier unabhängigen Kohorten prospektiv validiert. Das auf 22 Genen basierendes RNA-Modell ist in der Lage die individuelle Dauer der antimikrobiellen Therapie bei Tuberkulosepatienten vorauszusagen und ist darin anderen bislang publizierten RNA-Signaturen zur Diagnose der Tuberkulose überlegen. Die Anwendung des TB22 Modells könnte die Behandlungsdauer bei der Mehrzahl der Patienten mit MDR-TB substantiell verkürzen[174].

References

1. Cambau E, Drancourt M. Steps towards the discovery of Mycobacterium tuberculosis by Robert Koch, 1882. Clin Microbiol Infect 2014: 20(3): 196-201.

2. World Health Organization. WHO Guidelines Approved by the Guidelines Review Committee. WHO consolidated guidelines on tuberculosis: Module 4: Treatment - Drug-resistant tuberculosis treatment. Geneva: WHO; 2020.

3. World Health Organization. Global Tuberculosis Report 2018. Geneva: WHO; 2018.

4. World Health Organization. Global Tuberculosis Report 2019. Geneva: WHO; 2019.

5. Centers for Disease Control. Nosocomial transmission of multidrug-resistant tuberculosis to health-care workers and HIV-infected patients in an urban hospital--Florida. MMWR Morb Mortal Wkly Rep 1990: 39(40): 718-722.

6. World Health Organization. Global Tuberculosis Report 2020. Geneva: WHO; 2020.

7. Knight GM, McQuaid CF, Dodd PJ, Houben R. Global burden of latent multidrug-resistant tuberculosis: trends and estimates based on mathematical modelling. Lancet Infect Dis 2019: 19(8): 903-912.

8. Ahmad N, Ahuja SD, Akkerman OW, Alffenaar JC, Anderson LF, Baghaei P, Bang D, Barry PM, Bastos ML, Behera D, Benedetti A, Bisson GP, Boeree MJ, Bonnet M, Brode SK, Brust JCM, Cai Y, Caumes E, Cegielski JP, Centis R, Chan PC, Chan ED, Chang KC, Charles M, Cirule A, Dalcolmo MP, D'Ambrosio L, de Vries G, Dheda K, Esmail A, Flood J, Fox GJ, Fréchet-Jachym M, Fregona G, Gayoso R, Gegia M, Gler MT, Gu S, Guglielmetti L, Holtz TH, Hughes J, Isaakidis P, Jarlsberg L, Kempker RR, Keshavjee S, Khan FA, Kipiani M, Koenig SP, Koh WJ, Kritski A, Kuksa L, Kvasnovsky CL, Kwak N, Lan Z, Lange C, Laniado-Laborín R, Lee M, Leimane V, Leung CC, Leung EC, Li PZ, Lowenthal P, Maciel EL, Marks SM, Mase S, Mbuagbaw L, Migliori GB, Milanov V, Miller AC, Mitnick CD, Modongo C, Mohr E, Monedero I, Nahid P, Ndjeka N, O'Donnell MR, Padayatchi N, Palmero D, Pape JW, Podewils LJ, Reynolds I, Riekstina V, Robert J, Rodriguez M, Seaworth B, Seung KJ, Schnippel K, Shim TS, Singla R, Smith SE, Sotgiu G, Sukhbaatar G, Tabarsi P, Tiberi S, Trajman A, Trieu L, Udwadia ZF, van der Werf TS, Veziris N, Viiklepp P, Vilbrun SC, Walsh K, Westenhouse J, Yew WW, Yim JJ, Zetola NM, Zignol M, Menzies D. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. Lancet 2018: 392(10150): 821-834.

9. Conradie F, Diacon AH, Ngubane N, Howell P, Everitt D, Crook AM, Mendel CM, Egizi E, Moreira J, Timm J, McHugh TD, Wills GH, Bateson A, Hunt R, Van Niekerk C, Li M, Olugbosi M, Spigelman M. Treatment of Highly Drug-Resistant Pulmonary Tuberculosis. N Engl J Med 2020: 382(10): 893-902.

10. Burki T. BPaL approved for multidrug-resistant tuberculosis. Lancet Infect Dis 2019: 19(10): 1063-1064.

11. Abidi S, Achar J, Assao Neino MM, Bang D, Benedetti A, Brode S, Campbell JR, Casas EC, Conradie F, Dravniece G, du Cros P, Falzon D, Jaramillo E, Kuaban C, Lan Z, Lange C, Li PZ, Makhmudova M, Maug AKJ, Menzies D, Migliori GB, Miller A, Myrzaliev B, Ndjeka N, Noeske J, Viney K, Weyer K, Zhang BJ, Ahmad Khan F. Standardised shorter regimens versus individualised longer regimens for rifampin- or multidrug-resistant tuberculosis. Eur Respir J 2020: 55(3). 1-13.

12. Wahid A, Ahmad N, Ghafoor A, Latif A, Saleem F, Khan S, Atif M, Iqbal Q. Effectiveness of Shorter Treatment Regimen in Multidrug-Resistant Tuberculosis Patients in Pakistan: A Multicenter Retrospective Record Review. Am J Trop Med Hyg 2021: 104(5): 1784-1791.

13. Anh LTN, A MVK, Ramaswamy G, Htun T, Thanh Hoang Thi T, Hoai Nguyen G, Quelapio M, Gebhard A, Nguyen HB, Nguyen NV. High Levels of Treatment Success and Zero Relapse in Multidrug-Resistant Tuberculosis Patients Receiving a Levofloxacin-Based Shorter Treatment Regimen in Vietnam. Trop Med Infect Dis 2020: 5(1). 1-12.

14. Günther G, Gomez GB, Lange C, Rupert S, van Leth F. Availability, price and affordability of anti-tuberculosis drugs in Europe: a TBNET survey. Eur Respir J 2015: 45(4): 1081-1088.

15. Lan Z, Ahmad N, Baghaei P, Barkane L, Benedetti A, Brode SK, Brust JCM, Campbell JR, Chang VWL, Falzon D, Guglielmetti L, Isaakidis P, Kempker RR, Kipiani M, Kuksa L, Lange C, Laniado-Laborín R, Nahid P, Rodrigues D, Singla R, Udwadia ZF, Menzies D. Drug-associated adverse events in the treatment of multidrug-resistant tuberculosis: an individual patient data meta-analysis. Lancet Respir Med 2020: 8(4): 383-394.

16. Lange C, Alghamdi WA, Al-Shaer MH, Brighenti S, Diacon AH, DiNardo AR, Grobbel HP, Gröschel MI, von Groote-Bidlingmaier F, Hauptmann M, Heyckendorf J, Köhler N, Kohl TA, Merker M, Niemann S, Peloquin CA, Reimann M, Schaible UE, Schaub D, Schleusener V, Thye T, Schön T. Perspectives for personalized therapy for patients with multidrug-resistant tuberculosis. J Intern Med 2018. 163-188.

17. Pontali E, Raviglione MC, Migliori GB. Regimens to treat multidrug-resistant tuberculosis: past, present and future perspectives. Eur Respir Rev 2019: 28(152). 1-7.

18. Gröschel MI, Walker TM, van der Werf TS, Lange C, Niemann S, Merker M. Pathogen-based precision medicine for drug-resistant tuberculosis. PLoS Pathog 2018: 14(10): e1007297.

19. Heyckendorf J, Olaru ID, Ruhwald M, Lange C. Getting personal perspectives on individualized treatment duration in multidrug-resistant and extensively drug-resistant tuberculosis. Am J Respir Crit Care Med 2014: 190(4): 374-383.

20. Wanner R, Springer B. Die Tuberkulose, vom Altertum in die Neuzeit. Bulletin des médecins suisses 2008: 89: 37.

21. Bange F, Hahn H, Kaufmann S, Lange C, Ulrichs T. Mykobakterien. Medizinische Mikrobiologie und Infektiologie. Springer, Berlin, Heidelberg, 2020; pp. 447-465.

22. Koch R. Die Ätiologie der Tuberkulose (Nach, einem in der Physiologischen Gesellschaft zu Berlin am 24. März 1882 gehaltenen Vortrage). Physiologische Gesellschaft zu Berlin, 1882; pp. 428-445.

23. Migliori GB, Eker B, Richardson MD, Sotgiu G, Zellweger JP, Skrahina A, Ortmann J, Girardi E, Hoffmann H, Besozzi G, Bevilacqua N, Kirsten D, Centis R, Lange C. A retrospective TBNET assessment of linezolid safety, tolerability and efficacy in multidrug-resistant tuberculosis. Eur Respir J 2009: 34(2): 387-393.

24. Djelouadji Z, Raoult D, Daffé M, Drancourt M. A single-step sequencing method for the identification of Mycobacterium tuberculosis complex species. PLoS Negl Trop Dis 2008: 2(6): e253.

25. Werner H. Mycobacterium. In: Werner H, ed. Medizinische Mikrobiologie mit Repetitorium. Walter de Gruyter GmbH & Co KG, Berlin, 2020; pp. 240 - 248.

26. Wirth T, Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T, Kremer K, van Soolingen D, Rüsch-Gerdes S, Locht C, Brisse S, Meyer A, Supply P, Niemann S. Origin, spread and demography of the Mycobacterium tuberculosis complex. PLoS Pathog 2008: 4(9): e1000160.

27. Austermann P. Tuberkulose-eine wieder aktuelle Lungenkrankheit. GRIN Verlag, 2005. 15-19.

28. Houben RM, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLoS Med 2016: 13(10): e1002152.

29. Mack U, Migliori GB, Sester M, Rieder HL, Ehlers S, Goletti D, Bossink A, Magdorf K, Hölscher C, Kampmann B, Arend SM, Detjen A, Bothamley G, Zellweger JP, Milburn H, Diel R, Ravn P, Cobelens F, Cardona PJ, Kan B, Solovic I, Duarte R, Cirillo DM. LTBI: latent tuberculosis infection or lasting immune responses to M. tuberculosis? A TBNET consensus statement. Eur Respir J 2009: 33(5): 956-973.

30. European Centre for Disease Prevention and Control WHO Regional Office for Europe. Tuberculosis surveillance and monitoring in Europe 2021 –2019 data. Copenhagen: WHO; 2021 22.03.2021.

31. Suárez I, Fünger SM, Kröger S, Rademacher J, Fätkenheuer G, Rybniker J. The Diagnosis and Treatment of Tuberculosis. Dtsch Arztebl Int 2019: 116(43): 729-735.

32. Bauer T, Häcker B, Otto-Knapp R, Trost C. Welttuberkulosetag 2021: Infektionskontrolle gestern und heute–125 Jahre Deutsches Zentralkomitee zur Bekämpfung der Tuberkulose (DZK). Epid Bull 2021: 11: 3-5.

33. Robert Loch-Institut. Aktuelle Daten und Informationen zu Infektionskrankheiten und Public Health. Epid Bull 2020: 11: 3-22.

34. Heye T, Stoijkovic M, Kauczor HU, Junghanss T, Hosch W. [Extrapulmonary tuberculosis: radiological imaging of an almost forgotten transformation artist]. Rofo 2011: 183(11): 1019-1029.

35. Brodhun B, Altmann D, Hauer B, Fiebig L, Haas W. Bericht zur Epidemiologie der Tuberkulose in Deutschland für 2015. Robert Koch-Institut, 2016. 19-64.

36. Eum SY, Kong JH, Hong MS, Lee YJ, Kim JH, Hwang SH, Cho SN, Via LE, Barry CE, 3rd. Neutrophils are the predominant infected phagocytic cells in the airways of patients with active pulmonary TB. Chest 2010: 137(1): 122-128.

37. Ziegler R, Just HM, Castell S, Diel R, Gastmeier P, Haas W, Hauer B, Loytved G, Mielke M, Moser I, Nienhaus A, Richter E, Rüden H, Rüsch-Gerdes S, Schaberg T, Wischnewski N, Loddenkemper R. [Tuberculosis infection control--recommendations of the DZK]. Pneumologie 2012: 66(5): 269-282.

38. World Health Organization. Guidelines on the management of latent tuberculosis infection. World Health Organization, Geneva, 2015.

39. Gualano G, Mencarini P, Lauria FN, Palmieri F, Mfinanga S, Mwaba P, Chakaya J, Zumla A, Ippolito G. Tuberculin skin test - Outdated or still useful for Latent TB infection screening? Int J Infect Dis 2019: 80s: S20-s22.

40. Boom WH, Schaible UE, Achkar JM. The knowns and unknowns of latent Mycobacterium tuberculosis infection. J Clin Invest 2021: 131(3).

41. Drain PK, Bajema KL, Dowdy D, Dheda K, Naidoo K, Schumacher SG, Ma S, Meermeier E, Lewinsohn DM, Sherman DR. Incipient and Subclinical Tuberculosis: a Clinical Review of Early Stages and Progression of Infection. Clin Microbiol Rev 2018: 31(4).

42. Caws M, Marais B, Heemskerk D, Farrar J. Clinical Manifestations. Tuberculosis in Adults and Children. Springer Nature, 2015; p. 66.

43. Fuentes Ferrer M, Gutiérrez Torres L, Ayala Ramírez O, Rumayor Zarzuelo M, del Prado González N. Tuberculosis of the spine. A systematic review of case series. Int Orthop 2012: 36(2): 221-231.

44. Rock RB, Olin M, Baker CA, Molitor TW, Peterson PK. Central nervous system tuberculosis: pathogenesis and clinical aspects. Clin Microbiol Rev 2008: 21(2): 243-261, table of contents.

45. Ntsekhe M, Mayosi BM. Tuberculous pericarditis with and without HIV. Heart Fail Rev 2013: 18(3): 367-373.

46. Ryu YJ. Diagnosis of pulmonary tuberculosis: recent advances and diagnostic algorithms. Tuberc Respir Dis (Seoul) 2015: 78(2): 64-71.

47. Walzl G, McNerney R, du Plessis N, Bates M, McHugh TD, Chegou NN, Zumla A. Tuberculosis: advances and challenges in development of new diagnostics and biomarkers. Lancet Infect Dis 2018: 18(7): e199-e210.

48. Shapiro AE, Ross JM, Yao M, Schiller I, Kohli M, Dendukuri N, Steingart KR, Horne DJ. Xpert MTB/RIF and Xpert Ultra assays for screening for pulmonary tuberculosis and rifampicin resistance in adults, irrespective of signs or symptoms. Cochrane Database Syst Rev 2021: 3(3): Cd013694.

49. Lee JJ, Chong PY, Lin CB, Hsu AH, Lee CC. High resolution chest CT in patients with pulmonary tuberculosis: characteristic findings before and after antituberculous therapy. Eur J Radiol 2008: 67(1): 100-104.

50. Nachiappan AC, Rahbar K, Shi X, Guy ES, Mortani Barbosa EJ, Jr., Shroff GS, Ocazionez D, Schlesinger AE, Katz SI, Hammer MM. Pulmonary Tuberculosis: Role of Radiology in Diagnosis and Management. Radiographics 2017: 37(1): 52-72.

51. Bhalla AS, Goyal A, Guleria R, Gupta AK. Chest tuberculosis: Radiological review and imaging recommendations. Indian J Radiol Imaging 2015: 25(3): 213-225.

52. Bomanji J, Sharma R, Mittal BR, Gambhir S, Qureshy A, Begum SMF, Paez D, Sathekge M, Vorster M, Sobic Saranovic D, Pusuwan P, Mann V, Vinjamuri S, Zumla A, Pascual TNB. PET/CT features of extrapulmonary tuberculosis at first clinical presentation: a cross-sectional observational^{18} F-FDG imaging study across six countries. Eur Respir J 2020: 55(2).

53. Goletti D, Lee MR, Wang JY, Walter N, Ottenhoff THM. Update on tuberculosis biomarkers: From correlates of risk, to correlates of active disease and of cure from disease. Respirology 2018: 23(5): 455-466.

54. Grobbel HP, Merker M, Köhler N, Andres S, Hoffmann H, Heyckendorf J, Reimann M, Barilar I, Dreyer V, Hillemann D, Kalsdorf B, Kohl TA, Sanchez Carballo P, Schaub D, Todt K, Utpatel C, Maurer FP, Lange C, Niemann S. Design of Multidrug-Resistant Tuberculosis Treatment Regimens Based on DNA Sequencing. Clin Infect Dis 2021: 73(7): 1194-1202.

55. Schön T, Miotto P, Köser CU, Viveiros M, Böttger E, Cambau E. Mycobacterium tuberculosis drug-resistance testing: challenges, recent developments and perspectives. Clin Microbiol Infect 2017: 23(3): 154-160.

56. Bandodkar B, Shandil R, Bhat J, Balganesh T. Two Decades of TB Drug Discovery Efforts—What Have We Learned? Applied Sciences 2020: 10(16): 5704.

57. Barry CE. Lessons from seven decades of antituberculosis drug discovery. Curr Top Med Chem 2011: 11(10): 1216-1225.

58. Zumla A, Nahid P, Cole ST. Advances in the development of new tuberculosis drugs and treatment regimens. Nat Rev Drug Discov 2013: 12(5): 388-404.

59. Schaberg T, Bauer T, Brinkmann F, Diel R, Feiterna-Sperling C, Haas W, Hartmann P, Hauer B, Heyckendorf J, Lange C, Nienhaus A, Otto-Knapp R, Priwitzer M, Richter E, Rumetshofer R, Schenkel K, Schoch OD, Schönfeld N, Stahlmann R. [Tuberculosis Guideline for Adults - Guideline for Diagnosis and Treatment of Tuberculosis including LTBI Testing and Treatment of the German Central Committee (DZK) and the German Respiratory Society (DGP)]. Pneumologie 2017: 71(6): 325-397.

60. Grobbelaar M, Louw GE, Sampson SL, van Helden PD, Donald PR, Warren RM. Evolution of rifampicin treatment for tuberculosis. Infect Genet Evol 2019: 74: 103937.

61. Hedberg-Oldfors C, Mensch A, Visuttijai K, Stoltenburg G, Stoevesandt D, Kraya T, Oldfors A, Zierz S. Polyglucosan myopathy and functional characterization of a novel GYG1 mutation. Acta Neurol Scand 2018: 137(3): 308-315.

62. Jain K, Desai M, Solanki R, Dikshit RK. Treatment outcome of standardized regimen in patients with multidrug resistant tuberculosis. J Pharmacol Pharmacother 2014: 5(2): 145-149.

63. Imperial MZ, Nahid P, Phillips PPJ, Davies GR, Fielding K, Hanna D, Hermann D, Wallis RS, Johnson JL, Lienhardt C, Savic RM. A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis. Nat Med 2018: 24(11): 1708-1715.

64. Lange C, Aarnoutse R, Chesov D, van Crevel R, Gillespie SH, Grobbel HP, Kalsdorf B, Kontsevaya I, van Laarhoven A, Nishiguchi T, Mandalakas A, Merker M, Niemann S, Köhler N, Heyckendorf J, Reimann M, Ruhwald M, Sanchez-Carballo P, Schwudke D, Waldow F, DiNardo AR. Perspective for Precision Medicine for Tuberculosis. Front Immunol 2020: 11: 566608.

65. Nora D, Salluh J, Martin-Loeches I, Póvoa P. Biomarker-guided antibiotic therapy-strengths and limitations. Ann Transl Med 2017: 5(10): 208.

66. Califf RM. Biomarker definitions and their applications. Exp Biol Med (Maywood) 2018: 243(3): 213-221.

67. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982: 143(1): 29-36.

68. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol 2008: 56(1): 45-50.

69. Alene KA, Viney K, Yi H, McBryde ES, Yang K, Bai L, Gray DJ, Xu Z, Clements ACA. Comparison of the validity of smear and culture conversion as a prognostic marker of treatment outcome in patients with multidrug-resistant tuberculosis. PLoS One 2018: 13(5): e0197880.

70. Olaru ID, Heyckendorf J, Grossmann S, Lange C. Time to culture positivity and sputum smear microscopy during tuberculosis therapy. PLoS One 2014: 9(8): e106075.

71. Sabiiti W, Azam K, Farmer ECW, Kuchaka D, Mtafya B, Bowness R, Oravcova K, Honeyborne I, Evangelopoulos D, McHugh TD, Khosa C, Rachow A, Heinrich N, Kampira E, Davies G, Bhatt N, Ntinginya EN, Viegas S, Jani I, Kamdolozi M, Mdolo A, Khonga M, Boeree MJ, Phillips PPJ, Sloan D, Hoelscher M, Kibiki G, Gillespie SH. Tuberculosis bacillary load, an early marker of disease severity: the utility of tuberculosis Molecular Bacterial Load Assay. Thorax 2020: 75(7): 606-608.

72. Martin C, Castaigne C, Vierasu I, Garcia C, Wyndham-Thomas C, de Wit S. Prospective Serial FDG PET/CT During Treatment of Extrapulmonary Tuberculosis in HIV-Infected Patients: An Exploratory Study. Clin Nucl Med 2018: 43(9): 635-640.

73. Esmail H, Lai RP, Lesosky M, Wilkinson KA, Graham CM, Coussens AK, Oni T, Warwick JM, Said-Hartley Q, Koegelenberg CF, Walzl G, Flynn JL, Young DB, Barry Iii CE, O'Garra A, Wilkinson RJ. Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-18F fluoro-D-glucose positron emission and computed tomography. Nat Med 2016: 22(10): 1090-1093.

74. Heo EY, Chun EJ, Lee CH, Kim YW, Han SK, Shim YS, Lee HJ, Yim JJ. Radiographic improvement and its predictors in patients with pulmonary tuberculosis. Int J Infect Dis 2009: 13(6): e371-376.

75. Labugger I, Heyckendorf J, Dees S, Häussinger E, Herzmann C, Kohl TA, Richter E, Rivera-Milla E, Lange C. Detection of transrenal DNA for the diagnosis of pulmonary tuberculosis and treatment monitoring. Infection 2017: 45(3): 269-276.

76. Heyckendorf J, Labugger I, van der Merwe L, Garcia-Basteiro AL, Diacon AH, Lange C. Serial measurements of transrenal mycobacterial DNA as indicators of the early bactericidal activity (EBA) of antituberculosis drugs. Tuberculosis (Edinb) 2017: 102: 31-33.

77. Guérardel Y, Maes E, Briken V, Chirat F, Leroy Y, Locht C, Strecker G, Kremer L. Lipomannan and lipoarabinomannan from a clinical isolate of Mycobacterium kansasii: novel structural features and apoptosis-inducing properties. J Biol Chem 2003: 278(38): 36637-36651.

78. Bulterys MA, Wagner B, Redard-Jacot M, Suresh A, Pollock NR, Moreau E, Denkinger CM, Drain PK, Broger T. Point-Of-Care Urine LAM Tests for Tuberculosis Diagnosis: A Status Update. J Clin Med 2019: 9(1).

79. MacLean E, Broger T, Yerlikaya S, Fernandez-Carballo BL, Pai M, Denkinger CM. A systematic review of biomarkers to detect active tuberculosis. Nat Microbiol 2019: 4(5): 748-758.

80. Kawasaki M, Echiverri C, Raymond L, Cadena E, Reside E, Gler MT, Oda T, Ito R, Higashiyama R, Katsuragi K, Liu Y. Lipoarabinomannan in sputum to detect bacterial load and treatment response in patients with pulmonary tuberculosis: Analytic validation and evaluation in two cohorts. PLoS Med 2019: 16(4): e1002780.

81. Rudolf F, Lemvik G, Abate E, Verkuilen J, Schön T, Gomes VF, Eugen-Olsen J, Østergaard L, Wejse C. TBscore II: refining and validating a simple clinical score for treatment monitoring of patients with pulmonary tuberculosis. Scand J Infect Dis 2013: 45(11): 825-836.

82. Heyckendorf J, van Leth F, Avsar K, Glattki G, Günther G, Kalsdorf B, Müller M, Olaru ID, Rolling T, Salzer HJF, Schuhmann M, Terhalle E, Lange C. Treatment responses in multidrug-resistant tuberculosis in Germany. Int J Tuberc Lung Dis 2018: 22(4): 399-406.

83. Santos VS, Goletti D, Kontogianni K, Adams ER, Molina-Moya B, Dominguez J, Crudu V, Martins-Filho PRS, Ruhwald M, Lawson L, Bimba JS, Garcia-Basteiro AL, Petrone L, Kabeer BS, Reither K, Cuevas LE. Acute phase proteins and IP-10 as triage tests for the diagnosis of tuberculosis: systematic review and meta-analysis. Clin Microbiol Infect 2019: 25(2): 169-177.

84. Wilson D, Moosa MS, Cohen T, Cudahy P, Aldous C, Maartens G. Evaluation of Tuberculosis Treatment Response With Serial C-Reactive Protein Measurements. Open Forum Infect Dis 2018: 5(11): ofy253.

85. Nikolayevskyy V, Balabanova Y, Kontsevaya I, Ignatyeva O, Skenders G, Vasiliauskiene E, Bockel DV, Drobniewski F. Biomarkers of treatment success in fully sensitive pulmonary tuberculosis patients: a multicenter longitudinal study. Biomark Med 2020: 14(15): 1439-1452.

86. Kumar NP, Moideen K, Banurekha VV, Nair D, Babu S. Plasma Proinflammatory Cytokines Are Markers of Disease Severity and Bacterial Burden in Pulmonary Tuberculosis. Open Forum Infect Dis 2019: 6(7): ofz257.

87. Ambreen A, Khaliq A, Naqvi SZH, Tahir A, Mustafa M, Chaudhary SU, Mirza S, Mustafa T. Host biomarkers for monitoring therapeutic response in extrapulmonary tuberculosis. Cytokine 2021: 142: 155499.

88. Kumar NP, Moideen K, Nancy A, Viswanathan V, Shruthi BS, Sivakumar S, Natarajan M, Kornfeld H, Babu S. Plasma chemokines are biomarkers of disease severity, higher bacterial burden and delayed sputum culture conversion in pulmonary tuberculosis. Sci Rep 2019: 9(1): 18217.

89. Ahmed M. Exploring sputum independent host biomarkers for detection of Tuberculosis disease and monitoring treatment. München: Ludwig-Maximilian Univesität München; 2019.

90. Liu C, Zhao Z, Fan J, Lyon CJ, Wu HJ, Nedelkov D, Zelazny AM, Olivier KN, Cazares LH, Holland SM, Graviss EA, Hu Y. Quantification of circulating Mycobacterium tuberculosis antigen peptides allows rapid diagnosis of active disease and treatment monitoring. Proc Natl Acad Sci U S A 2017: 114(15): 3969-3974.

91. Kaewseekhao B, Roytrakul S, Yingchutrakul Y, Salao K, Reechaipichitkul W, Faksri K. Proteomic analysis of infected primary human leucocytes revealed PSTK as potential treatment-monitoring marker for active and latent tuberculosis. PLoS One 2020: 15(4): e0231834.

92. Nahid P, Dorman SE, Alipanah N, Barry PM, Brozek JL, Cattamanchi A, Chaisson LH, Chaisson RE, Daley CL, Grzemska M, Higashi JM, Ho CS, Hopewell PC, Keshavjee SA, Lienhardt C, Menzies R, Merrifield C, Narita M, O'Brien R, Peloquin CA, Raftery A, Saukkonen J, Schaaf HS, Sotgiu G, Starke JR,

Migliori GB, Vernon A. Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. Clin Infect Dis 2016: 63(7): e147-e195.

93. Fitzgerald BL, Islam MN, Graham B, Mahapatra S, Webb K, Boom WH, Malherbe ST, Joloba ML, Johnson JL, Winter J, Walzl G, Belisle JT. Elucidation of a Human Urine Metabolite as a Seryl-Leucine Glycopeptide and as a Biomarker of Effective Anti-Tuberculosis Therapy. ACS Infect Dis 2019: 5(3): 353-364.

94. Combrink M, du Preez I, Ronacher K, Walzl G, Loots DT. Time-Dependent Changes in Urinary Metabolome Before and After Intensive Phase Tuberculosis Therapy: A Pharmacometabolomics Study. Omics 2019: 23(11): 560-572.

95. Tientcheu LD, Maertzdorf J, Weiner J, Adetifa IM, Mollenkopf HJ, Sutherland JS, Donkor S, Kampmann B, Kaufmann SH, Dockrell HM, Ota MO. Differential transcriptomic and metabolic profiles of M. africanum- and M. tuberculosis-infected patients after, but not before, drug treatment. Genes Immun 2015: 16(5): 347-355.

96. du Preez I, Luies L, Loots DT. The application of metabolomics toward pulmonary tuberculosis research. Tuberculosis (Edinb) 2019: 115: 126-139.

97. Casamassimi A, Federico A, Rienzo M, Esposito S, Ciccodicola A. Transcriptome Profiling in Human Diseases: New Advances and Perspectives. Int J Mol Sci 2017: 18(8).

98. Brix J, Koch H-G, Heinrich PC. Transkription und Prozessierung der RNA. Löffler/Petrides Biochemie und Pathobiochemie. Springer, 2014; pp. 567-587.

99. Lekka E, Hall J. Noncoding RNAs in disease. FEBS Lett 2018: 592(17): 2884-2900.

100. Bao Z, Yang Z, Huang Z, Zhou Y, Cui Q, Dong D. LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. Nucleic Acids Res 2019: 47(D1): D1034-d1037.

101. Xi X, Li T, Huang Y, Sun J, Zhu Y, Yang Y, Lu ZJ. RNA Biomarkers: Frontier of Precision Medicine for Cancer. Noncoding RNA 2017: 3(1).

102. Mantione KJ, Kream RM, Kuzelova H, Ptacek R, Raboch J, Samuel JM, Stefano GB. Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq. Med Sci Monit Basic Res 2014: 20: 138-142.

103. Vlahou A, Magni F, Mischak H, Zoidakis J. Integration of omics approaches and systems biology for clinical applications. John Wiley & Sons, 2018.

104. Dong Z, Chen Y. Transcriptomics: advances and approaches. Sci China Life Sci 2013: 56(10): 960-967.

105. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015: 43(7): e47.

106. Quinn TP, Erb I, Gloor G, Notredame C, Richardson MF, Crowley TM. A field guide for the compositional analysis of any-omics data. Gigascience 2019: 8(9).

107. Honeyborne I, McHugh TD, Phillips PP, Bannoo S, Bateson A, Carroll N, Perrin FM, Ronacher K, Wright L, van Helden PD, Walzl G, Gillespie SH. Molecular bacterial load assay, a culture-free biomarker for rapid and accurate quantification of sputum Mycobacterium tuberculosis bacillary load during treatment. J Clin Microbiol 2011: 49(11): 3905-3911.

108. Honeyborne I, Mtafya B, Phillips PP, Hoelscher M, Ntinginya EN, Kohlenberg A, Rachow A, Rojas-Ponce G, McHugh TD, Heinrich N. The molecular bacterial load assay replaces solid culture for measuring early bactericidal response to antituberculosis treatment. J Clin Microbiol 2014: 52(8): 3064-3067.

109. Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. PLoS Med 2019: 16(4): e1002786.

110. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, Mahomed H, Erasmus M, Whatney W, Hussey GD, Abrahams D, Kafaar F, Hawkridge T, Verver S, Hughes EJ, Ota M, Sutherland J, Howe R, Dockrell HM, Boom WH, Thiel B, Ottenhoff THM, Mayanja-Kizza H, Crampin AC, Downing K, Hatherill M, Valvo J, Shankar S, Parida SK, Kaufmann SHE, Walzl G, Aderem A, Hanekom WA. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet 2016: 387(10035): 2312-2322.

111. Darboe F, Mbandi SK, Naidoo K, Yende-Zuma N, Lewis L, Thompson EG, Duffy FJ, Fisher M, Filander E, van Rooyen M, Bilek N, Mabwe S, McKinnon LR, Chegou N, Loxton A, Walzl G, Tromp G, Padayatchi N, Govender D, Hatherill M, Karim SA, Zak DE, Penn-Nicholson A, Scriba TJ. Detection of Tuberculosis Recurrence, Diagnosis and Treatment Response by a Blood Transcriptomic Risk Signature in HIV-Infected Persons on Antiretroviral Therapy. Front Microbiol 2019: 10: 1441.

112. Turner CT, Gupta RK, Tsaliki E, Roe JK, Mondal P, Nyawo GR, Palmer Z, Miller RF, Reeve BW, Theron G, Noursadeghi M. Blood transcriptional biomarkers for active pulmonary tuberculosis in a high-burden setting: a prospective, observational, diagnostic accuracy study. Lancet Respir Med 2020: 8(4): 407-419.

113. Warsinske HC, Rao AM, Moreira FMF, Santos PCP, Liu AB, Scott M, Malherbe ST, Ronacher K, Walzl G, Winter J, Sweeney TE, Croda J, Andrews JR, Khatri P. Assessment of Validity of a Blood-Based 3-Gene Signature Score for Progression and Diagnosis of Tuberculosis, Disease Severity, and Treatment Response. JAMA Netw Open 2018: 1(6): e183779.

114. Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. Lancet Respir Med 2016: 4(3): 213-224.

115. Ho J, Bokil NJ, Nguyen PTB, Nguyen TA, Liu MY, Hare N, Fox GJ, Saunders BM, Marks GB, Britton WJ. A transcriptional blood signature distinguishes early tuberculosis disease from latent tuberculosis infection and uninfected individuals in a Vietnamese cohort. J Infect 2020: 81(1): 72-80.

116. Hoang LT, Jain P, Pillay TD, Tolosa-Wright M, Niazi U, Takwoingi Y, Halliday A, Berrocal-Almanza LC, Deeks JJ, Beverley P, Kon OM, Lalvani A. Transcriptomic signatures for diagnosing tuberculosis in clinical practice: a prospective, multicentre cohort study. Lancet Infect Dis 2021: 21(3): 366-375.

117. Penn-Nicholson A, Mbandi SK, Thompson E, Mendelsohn SC, Suliman S, Chegou NN, Malherbe ST, Darboe F, Erasmus M, Hanekom WA, Bilek N, Fisher M, Kaufmann SHE, Winter J, Murphy M, Wood R, Morrow C, Van Rhijn I, Moody B, Murray M, Andrade BB, Sterling TR, Sutherland J, Naidoo K, Padayatchi N, Walzl G, Hatherill M, Zak D, Scriba TJ. RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. Sci Rep 2020: 10(1): 8629.

118. Thompson EG, Du Y, Malherbe ST, Shankar S, Braun J, Valvo J, Ronacher K, Tromp G, Tabb DL, Alland D, Shenai S, Via LE, Warwick J, Aderem A, Scriba TJ, Winter J, Walzl G, Zak DE. Host blood RNA signatures predict the outcome of tuberculosis treatment. Tuberculosis (Edinb) 2017: 107: 48-58.

119. Sivakumaran D, Jenum S, Vaz M, Selvam S, Ottenhoff THM, Haks MC, Malherbe ST, Doherty TM, Ritz C, Grewal HMS. Combining host-derived biomarkers with patient characteristics improves signature performance in predicting tuberculosis treatment outcomes. Commun Biol 2020: 3(1): 359.

120. Ronacher K, Chegou NN, Kleynhans L, Djoba Siawaya JF, du Plessis N, Loxton AG, Maasdorp E, Tromp G, Kidd M, Stanley K, Kriel M, Menezes A, Gutschmidt A, van der Spuy GD, Warren RM, Dietze R, Okwera A, Thiel B, Belisle JT, Cliff JM, Boom WH, Johnson JL, van Helden PD, Dockrell HM, Walzl G. Distinct serum biosignatures are associated with different tuberculosis treatment outcomes. Tuberculosis (Edinb) 2019: 118: 101859.

121. Sepulveda JL. Using R and Bioconductor in Clinical Genomics and Transcriptomics. J Mol Diagn 2020: 22(1): 3-20.

122. Friedman J. The elements of statistical learning: Data mining, inference, and prediction. Springer open, 2017.

123. Gentleman R. R programming for bioinformatics. CRC Press Taylor & Francis Group, Seattle, 2008.

124. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 2005: 21(16): 3439-3440.

125. Huppmann D, Gidden M, Fricko O, Kolp P, Orthofer C, Pimmer M, Kushin N, Vinca A, Mastrucci A, Riahi K. The MESSAGEix Integrated Assessment Model and the ix modeling platform (ixmp): An open framework for integrated and cross-cutting analysis of energy, climate, the environment, and sustainable development. Environ Model Softw 2019: 112: 143-156.

126. Smirnov P, Safikhani Z, El-Hachem N, Wang D, She A, Olsen C, Freeman M, Selby H, Gendoo DM, Grossmann P, Beck AH, Aerts HJ, Lupien M, Goldenberg A, Haibe-Kains B. PharmacoGx: an R package for analysis of large pharmacogenomic datasets. Bioinformatics 2016: 32(8): 1244-1246.

127. Kim JH, Shin DM, Lee YS. Effect of local background intensities in the normalization of cDNA microarray data with a skewed expression profiles. Exp Mol Med 2002: 34(3): 224-232.

128. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, Smyth GK. A comparison of background correction methods for two-colour microarrays. Bioinformatics 2007: 23(20): 2700-2707.

129. Scherer A. Batch effects and noise in microarray experiments: sources and solutions. John Wiley & Sons, United Kingdom, 2009.

130. Kerr MK, Churchill GA. Statistical design and the analysis of gene expression microarray data. Genet Res 2001: 77(2): 123-128.

131. Fan J, Li R. Statistical challenges with high dimensionality. In: Proceedings of the international Congress of Mathematicians. Madrid, Spain; 2006.

132. Manbari Z, AkhlaghianTab F, Salavati C. Hybrid fast unsupervised feature selection for high-dimensional data. Expert Syst Appl 2019: 124: 97-118.

133. Zimek A, Schubert E, Kriegel H. A survey on unsupervised outlier detection in high-dimensional numerical data. Stat Anal Data Min 2012: 5(5): 363-387.

134. Chum O, Matas J. Unsupervised discovery of co-occurrence in sparse high dimensional data. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2010: IEEE; 2010. p. 3416-3423.

135. Trezzi J, Hiller K, Mollenhauer B. The importance of an independent validation cohort for metabolomics biomarker studies. Mov Disord 2018: 33(5): 856-856.

136. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016: 69: 245-247.

137. Musoro JZ, Zwinderman AH, Puhan MA, ter Riet G, Geskus RB. Validation of prediction models based on lasso regression with multiply imputed data. BMC Med Res Methodol 2014: 14: 116.

138. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 2012: 8(2): e1002375.

139. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014: 15(12): 550.

140. Noble WS. How does multiple testing correction work? Nat Biotechnol 2009: 27(12): 1135-1137.

141. Hawkins DM. The problem of overfitting. J Chem Inf Comput Sci 2004: 44(1): 1-12.

142. Heyckendorf J, van Leth F, Kalsdorf B, Olaru ID, Günther G, Salzer HJF, Terhalle E, Rolling T, Glattki G, Müller M, Schuhmann M, Avsar K, Lange C. Relapse-free cure from multidrug-resistant tuberculosis in Germany. Eur Respir J 2018: 51(2).

143. Heyckendorf J, Marwitz S, Reimann M, Avsar K, DiNardo AR, Günther G, Hoelscher M, Ibraim E, Kalsdorf B, Kaufmann SHE, Kontsevaya I, van Leth F, Mandalakas AM, Maurer FP, Müller M, Nitschkowski D, Olaru ID, Popa C, Rachow A, Rolling T, Rybniker J, Salzer HJF, Sanchez-Carballo P, Schuhmann M, Schaub D, Spinu V, Suárez I, Terhalle E, Unnewehr M, Weiner J, 3rd, Goldmann T, Lange C. Prediction of anti-tuberculosis treatment duration based on a 22-gene transcriptomic model. Eur Respir J 2021: 58(3).

144. Schaberg T, Bauer T, Castell S, Dalhoff K, Detjen A, Diel R, Greinert U, Hauer B, Lange C, Magdorf K, Loddenkemper R. [Recommendations for therapy, chemoprevention and chemoprophylaxis of tuberculosis in adults and children. German Central Committee against Tuberculosis (DZK), German Respiratory Society (DGP)]. Pneumologie 2012: 66(3): 133-171.

145. World Health Organization. WHO treatment guidelines for drug-resistant tuberculosis, 2016 update, Geneva, 2016.

146. Günther G, Lange C, Alexandru S, Altet N, Avsar K, Bang D, Barbuta R, Bothamley G, Ciobanu A, Crudu V, Danilovits M, Dedicoat M, Duarte R, Gualano G, Kunst H, de Lange W, Leimane V, Magis-Escurra C, McLaughlin AM, Muylle I, Polcová V, Popa C, Rumetshofer R, Skrahina A, Solodovnikova V, Spinu V, Tiberi S, Viiklepp P, van Leth F. Treatment Outcomes in Multidrug-Resistant Tuberculosis. N Engl J Med 2016: 375(11): 1103-1105.

147. World Health Organization. Definitions and reporting framework for tuberculosis–2013 revision: updated December 2014 and January 2020. Geneva: World Health Organization; 2013. Report No.: 9241505346.

148. Bothamley G. The Tuberculosis Network European Trials Group (TBNET): new directions in the management of tuberculosis. Breathe (Sheff) 2017: 13(3): e65-e71.

149. Patrick E, Schramm SJ, Ormerod JT, Scolyer RA, Mann GJ, Mueller S, Yang JY. A multi-step classifier addressing cohort heterogeneity improves performance of prognostic biomarkers in three cancer types. Oncotarget 2017: 8(2): 2807-2815.

150. Archer KJ, Dumur CI, Ramakrishnan V. Graphical technique for identifying a monotonic variance stabilizing transformation for absolute gene intensity signals. BMC Bioinformatics 2004: 5: 60.

151. Bogdan M, Ghosh J, Tokdar S. A comparison of the Benjamini-Hochberg procedure with some Bayesian rules for multiple testing. Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen. Institute of Mathematical Statistics, 2008; pp. 211-230.

152. Morey LC, Blashfield RK, Skinner HA. A Comparison of Cluster Analysis Techniques Withing a Sequential Validation Framework. Multivariate Behav Res 1983: 18(3): 309-329.

153. Benesty J, Chen J, Huang Y, Cohen I. Pearson correlation coefficient. Noise reduction in speech processing. Springer, Berlin, 2009; pp. 1-4.

154. Held U. Tücken von Korrelationen: die Korrelationskoeffizienten von Pearson und Spearman. In: Swiss Med Forum; 2010: EMH Swiss Medical Publishers; 2010. p. 652-653.

155. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J. g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res 2016: 44(W1): W83-89.

156. Akaike H. A new look at the statistical model identification. IEEE Trans Control Syst Technol 1974: 19(6): 716-723.

157. Burnham K, Anderson D. A practical information-theoretic approach, 2002.

158. Liaw A, Wiener M. Classification and regression by randomForest. R news 2002: 2(3): 18-22.

159. Breiman L. Random forests. Mach learn 2001: 45(1): 5-32.

160. Archer KJ, Kimes RV. Empirical characterization of random forest variable importance measures. Computational statistics & data analysis 2008: 52(4): 2249-2260.

161. Tuszynski J. caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc. R package version 2008: 1.

162. Bradley A. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognit 1997: 30(7): 1145-1159.

163. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986: 42(1): 121-130.

164. Akter S, Xu D, Nagel SC, Bromfield JJ, Pelch K, Wilshire GB, Joshi T. Machine Learning Classifiers for Endometriosis Using Transcriptomics and Methylomics Data. Front Genet 2019: 10: 766.

165. Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, Flanagan JM, Vineis P. DNA methylation as a long-term biomarker of exposure to tobacco smoke. Epidemiology 2013: 24(5): 712-716.

166. De Meyer G, Shapiro F, Vanderstichele H, Vanmechelen E, Engelborghs S, De Deyn PP, Coart E, Hansson O, Minthon L, Zetterberg H, Blennow K, Shaw L, Trojanowski JQ. Diagnosis-independent Alzheimer disease biomarker signature in cognitively normal elderly people. Arch Neurol 2010: 67(8): 949-956.

167. Dervieux T, Greenstein N, Kremer J. Pharmacogenomic and metabolic biomarkers in the folate pathway and their association with methotrexate effects during dosage escalation in rheumatoid arthritis. Arthritis Rheum 2006: 54(10): 3095-3103.

168. Zhang Z. Reverse regression: a method for joint analysis of multiple endpoints in randomized clinical trials. Stat Sin 2014: 24(4): 1753-1769.

169. Wang H, Li G, Jiang G. Robust regression shrinkage and consistent variable selection through the LAD-Lasso. J Bus Econ Stat 2007: 25(3): 347-355.

170. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997: 16(4): 385-395.

171. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol 1996: 58(1): 267-288.

172. Urminsky O, Hansen C, Chernozhukov V. Using double-lasso regression for principled variable selection. Booth School of Business & Department of Economics and Center for Statistics MIT, Chicago, 2016.

173. Belloni A, Chernozhukov V, Hansen C. Inference on treatment effects after selection among high-dimensional controls. Rev Econ Stud 2014: 81(2): 608-650.

174. Belloni A, Chernozhukov V, Kato K. Valid post-selection inference in high-dimensional approximately sparse quantile regression models. J Am Stat Assoc 2019: 114(526): 749-758.

175. Heyckendorf J, Reimann M, Marwitz S, Lange C. Pathogen-free diagnosis of tuberculosis. Lancet Infect Dis 2021: 21(8): 1066.

176. Anderson ST, Kaforou M, Brent AJ, Wright VJ, Banwell CM, Chagaluka G, Crampin AC, Dockrell HM, French N, Hamilton MS, Hibberd ML, Kern F, Langford PR, Ling L, Mlotha R, Ottenhoff THM, Pienaar S, Pillay V, Scott JAG, Twahir H, Wilkinson RJ, Coin LJ, Heyderman RS, Levin M, Eley B. Diagnosis of childhood tuberculosis and host RNA expression in Africa. N Engl J Med 2014: 370(18): 1712-1723.

177. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, Wilkinson KA, Banchereau R, Skinner J, Wilkinson RJ, Quinn C, Blankenship D, Dhawan R, Cush JJ, Mejias A, Ramilo O, Kon OM, Pascual V, Banchereau J, Chaussabel D, O'Garra A. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature 2010: 466(7309): 973-977.

178. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, Banwell CM, Brent AJ, Crampin AC, Dockrell HM, Eley B, Heyderman RS, Hibberd ML, Kern F, Langford PR, Ling L, Mendelson M, Ottenhoff TH, Zgambo F, Wilkinson RJ, Coin LJ, Levin M. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. PLoS Med 2013: 10(10): e1001538.

179. Laux da Costa L, Delcroix M, Dalla Costa ER, Prestes IV, Milano M, Francis SS, Unis G, Silva DR, Riley LW, Rossetti ML. A real-time PCR signature to discriminate between tuberculosis and other pulmonary diseases. Tuberculosis (Edinb) 2015: 95(4): 421-425.

180. Maertzdorf J, McEwen G, Weiner J, 3rd, Tian S, Lader E, Schriek U, Mayanja-Kizza H, Ota M, Kenneth J, Kaufmann SH. Concise gene signature for point-of-care classification of tuberculosis. EMBO Mol Med 2016: 8(2): 86-95.

181. Sambarey A, Devaprasad A, Baloni P, Mishra M, Mohan A, Tyagi P, Singh A, Akshata JS, Sultana R, Buggi S, Chandra N. Meta-analysis of host response networks identifies a common core in tuberculosis. NPJ Syst Biol Appl 2017: 3: 4.

182. Singhania A, Verma R, Graham CM, Lee J, Tran T, Richardson M, Lecine P, Leissner P, Berry MPR, Wilkinson RJ, Kaiser K, Rodrigue M, Woltmann G, Haldar P, O'Garra A. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. Nat Commun 2018: 9(1): 2308.

183. Suliman S, Thompson EG, Sutherland J, Weiner J, 3rd, Ota MOC, Shankar S, Penn-Nicholson A, Thiel B, Erasmus M, Maertzdorf J, Duffy FJ, Hill PC, Hughes EJ, Stanley K, Downing K, Fisher ML, Valvo J, Parida SK, van der Spuy G, Tromp G, Adetifa IMO, Donkor S, Howe R, Mayanja-Kizza H, Boom WH, Dockrell HM, Ottenhoff THM, Hatherill M, Aderem A, Hanekom WA, Scriba TJ, Kaufmann SHE, Zak DE, Walzl G. Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis. Am J Respir Crit Care Med 2018: 197(9): 1198-1208.

184. Sutherland JS, Loxton AG, Haks MC, Kassa D, Ambrose L, Lee JS, Ran L, van Baarle D, Maertzdorf J, Howe R, Mayanja-Kizza H, Boom WH, Thiel BA, Crampin AC, Hanekom W, Ota MO, Dockrell H, Walzl G, Kaufmann SH, Ottenhoff TH. Differential gene expression of activating Fcγ receptor classifies active tuberculosis regardless of human immunodeficiency virus status or ethnicity. Clin Microbiol Infect 2014: 20(4): O230-238.

185. Ilievska-Poposka B, Metodieva M, Zakoska M, Vragoterova C, Trajkov D. Latent Tuberculosis Infection - Diagnosis and Treatment. Open Access Maced J Med Sci 2018: 6(4): 651-655.

186. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012: 28(6): 882-883.

187. Weiner 3rd J, Domaszewska T. tmod: an R package for general and multivariate enrichment analysis. PeerJ Preprints 2016: 4.

188. Du J, Yuan Z, Ma Z, Song J, Xie X, Chen Y. KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. Mol Biosyst 2014: 10(9): 2441-2447.

189. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, Stichweh D, Blankenship D, Li L, Munagala I, Bennett L, Allantaz F, Mejias A, Ardura M, Kaizer E, Monnet L, Allman W, Randall H, Johnson D, Lanier A, Punaro M, Wittkowski KM, White P, Fay J, Klintmalm G, Ramilo O, Palucka AK, Banchereau J, Pascual V. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity 2008: 29(1): 150-164.

190. Rodríguez Del Águila M, González-Ramírez A. Sample size calculation. Allergol Immunopathol (Madr) 2014: 42(5): 485-492.

191. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010: 5(9): 1315-1316.

192. Bloom CI, Graham CM, Berry MP, Wilkinson KA, Oni T, Rozakeas F, Xu Z, Rossello-Urgell J, Chaussabel D, Banchereau J, Pascual V, Lipman M, Wilkinson RJ, O'Garra A. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. PLoS One 2012: 7(10): e46191.

193. Blankley S, Graham CM, Turner J, Berry MP, Bloom CI, Xu Z, Pascual V, Banchereau J, Chaussabel D, Breen R, Santis G, Blankenship DM, Lipman M, O'Garra A. The Transcriptional Signature of Active Tuberculosis Reflects Symptom Status in Extra-Pulmonary and Pulmonary Tuberculosis. PLoS One 2016: 11(10): e0162220.

194. Bloom CI, Graham CM, Berry MP, Rozakeas F, Redford PS, Wang Y, Xu Z, Wilkinson KA, Wilkinson RJ, Kendrick Y, Devouassoux G, Ferry T, Miyara M, Bouvry D, Valeyre D, Gorochov G, Blankenship D, Saadatian M, Vanhems P, Beynon H, Vancheeswaran R, Wickremasinghe M, Chaussabel D, Banchereau J, Pascual V, Ho LP, Lipman M, O'Garra A. Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. PLoS One 2013: 8(8): e70630.

195. Sambarey A, Devaprasad A, Mohan A, Ahmed A, Nayak S, Swaminathan S, D'Souza G, Jesuraj A, Dhar C, Babu S, Vyakarnam A, Chandra N. Unbiased Identification of Blood-based Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks. EBioMedicine 2017: 15: 112-126.

196. Maertzdorf J, Ota M, Repsilber D, Mollenkopf HJ, Weiner J, Hill PC, Kaufmann SH. Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis. PLoS One 2011: 6(10): e26938.

197. Lange C, Mori T. Advances in the diagnosis of tuberculosis. Respirology 2010: 15(2): 220-240.

198. Nguyen MH, Levy NS, Ahuja SD, Trieu L, Proops DC, Achkar JM. Factors Associated With Sputum Culture-Negative vs Culture-Positive Diagnosis of Pulmonary Tuberculosis. JAMA Netw Open 2019: 2(2): e187617.

199. Hamada Y, Paulos L, Baruch NG, Cronin W. Proposed Approach for 4-Month Treatment of Culture-Negative Pulmonary Tuberculosis in Adults. Ann Am Thorac Soc 2016: 13(9): 1657-1658.

200. Mandal N, Anand PK, Gautam S, Das S, Hussain T. Diagnosis and treatment of paediatric tuberculosis: An insight review. Crit Rev Microbiol 2017: 43(4): 466-480.

201. Umair M, Siddiqui SA, Farooq MA. Diagnostic Accuracy of Sputum Microscopy in Comparison With GeneXpert in Pulmonary Tuberculosis. Cureus 2020: 12(11): e11383.

202. Acharya B, Acharya A, Gautam S, Ghimire SP, Mishra G, Parajuli N, Sapkota B. Advances in diagnosis of Tuberculosis: an update into molecular diagnosis of Mycobacterium tuberculosis. Mol Biol Rep 2020: 47(5): 4065-4075.

203. Zifodya JS, Kreniske JS, Schiller I, Kohli M, Dendukuri N, Schumacher SG, Ochodo EA, Haraka F, Zwerling AA, Pai M, Steingart KR, Horne DJ. Xpert Ultra versus Xpert MTB/RIF for pulmonary tuberculosis and rifampicin resistance in adults with presumptive pulmonary tuberculosis. Cochrane Database Syst Rev 2021: 2: Cd009593.

204. Vonasek B, Ness T, Takwoingi Y, Kay AW, van Wyk SS, Ouellette L, Marais BJ, Steingart KR, Mandalakas AM. Screening tests for active pulmonary tuberculosis in children. Cochrane Database Syst Rev 2021: 6(6): Cd013693.

205. Gillespie SH, Crook AM, McHugh TD, Mendel CM, Meredith SK, Murray SR, Pappas F, Phillips PP, Nunn AJ. Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis. N Engl J Med 2014: 371(17): 1577-1587.

206. Merle CS, Sismanidis C, Sow OB, Gninafon M, Horton J, Lapujade O, Lo MB, Mitchinson DA, Perronne C, Portaels F, Odhiambo J, Olliaro P, Rustomjee R, Lienhardt C, Fielding K. A pivotal registration phase III, multicenter, randomized tuberculosis controlled trial: design issues and lessons learnt from the Gatifloxacin for TB (OFLOTUB) project. Trials 2012: 13: 61.

207. Nunn A. An international multicentre controlled clinical trial to evaluate high-dose rifapentine and a quinolone in the treatment of pulmonary tuberculosis (RIFAQUIN). In: TM & IH; 2010; 2010. p. S22-S23.

208. Dorman SE, Nahid P, Kurbatova EV, Phillips PPJ, Bryant K, Dooley KE, Engle M, Goldberg SV, Phan HTT, Hakim J, Johnson JL, Lourens M, Martinson NA, Muzanyi G, Narunsky K, Nerette S, Nguyen NV, Pham TH, Pierre S, Purfield AE, Samaneka W, Savic RM, Sanne I, Scott NA, Shenje J, Sizemore E, Vernon A, Waja Z, Weiner M, Swindells S, Chaisson RE. Four-Month Rifapentine Regimens with or without Moxifloxacin for Tuberculosis. N Engl J Med 2021: 384(18): 1705-1718.

209. Van Deun A, Salim MA, Das AP, Bastian I, Portaels F. Results of a standardised regimen for multidrug-resistant tuberculosis in Bangladesh. Int J Tuberc Lung Dis 2004: 8(5): 560-567.

210. Trébucq A, Decroo T, Van Deun A, Piubello A, Chiang CY, Koura KG, Schwoebel V. Short-Course Regimen for Multidrug-Resistant Tuberculosis: A Decade of Evidence. J Clin Med 2019: 9(1).

211. Moodley R, Godec TR. Short-course treatment for multidrug-resistant tuberculosis: the STREAM trials. Eur Respir Rev 2016: 25(139): 29-35.

212. Udwadia ZF, Tornheim JA, Ganatra S, DeLuca A, Rodrigues CS, Gupta A. Few eligible for the newly recommended short course MDR-TB regimen at a large Mumbai private clinic. BMC Infect Dis 2019: 19(1): 94.

213. Caminero JA, Piubello A, Scardigli A, Migliori GB. Proposal for a standardised treatment regimen to manage pre- and extensively drug-resistant tuberculosis cases. Eur Respir J 2017: 50(1).

214. den Hertog AL, Montero-Martín M, Saunders RL, Blakiston M, Menting S, Sherchand JB, Lawson L, Oladimeji O, Abdurrahman ST, Cuevas LE, Anthony RM. Cytokine Kinetics in the First Week of Tuberculosis Therapy as a Tool to Confirm a Clinical Diagnosis and Guide Therapy. PLoS One 2015: 10(6): e0129552.

215. Gebremicael G, Kassa D, Quinten E, Alemayehu Y, Gebreegziaxier A, Belay Y, van Baarle D, Ottenhoff THM, Cliff JM, Haks MC. Host Gene Expression Kinetics During Treatment of Tuberculosis in HIV-Coinfected Individuals Is Independent of Highly Active Antiretroviral Therapy. J Infect Dis 2018: 218(11): 1833-1846.

216. Clifford V, He Y, Zufferey C, Connell T, Curtis N. Interferon gamma release assays for monitoring the response to treatment for tuberculosis: A systematic review. Tuberculosis (Edinb) 2015: 95(6): 639-650.

217. Denkinger CM, Pai M, Patel M, Menzies D. Gamma interferon release assay for monitoring of treatment response for active tuberculosis: an explosion in the spaghetti factory. J Clin Microbiol 2013: 51(2): 607-610.

218. Ahmed MIM, Ziegler C, Held K, Dubinski I, Ley-Zaporozhan J, Geldmacher C, von Both U. The TAM-TB Assay-A Promising TB Immune-Diagnostic Test With a Potential for Treatment Monitoring. Front Pediatr 2019: 7: 27.

219. Zhou Y, Du J, Hou HY, Lu YF, Yu J, Mao LY, Wang F, Sun ZY. Application of ImmunoScore Model for the Differentiation between Active Tuberculosis and Latent Tuberculosis Infection as Well as Monitoring Anti-tuberculosis Therapy. Front Cell Infect Microbiol 2017: 7: 457.

220. Essone PN, Kalsdorf B, Chegou NN, Loxton AG, Kriel M, Preyer R, Ernst M, Walzl G, Lange C. Bifunctional T-cell-derived cytokines for the diagnosis of tuberculosis and treatment monitoring. Respiration 2014: 88(3): 251-261.

221. Kontsevaya I, Lange C, Comella-Del-Barrio P, Coarfa C, DiNardo AR, Gillespie SH, Hauptmann M, Leschczyk C, Mandalakas AM, Martinecz A, Merker M, Niemann S, Reimann M, Rzhepishevska O, Schaible UE, Scheu KM, Schurr E, Abel Zur Wiesch P, Heyckendorf J. Perspectives for systems biology in the management of tuberculosis. Eur Respir Rev 2021: 30(160).

222. Tucci P, González-Sapienza G, Marin M. Pathogen-derived biomarkers for active tuberculosis diagnosis. Front Microbiol 2014: 5: 549.

223. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 2005: 21(9): 2067-2075.

224. Ranstam J, Cook J. LASSO regression. BJS 2018: 105(10): 1348-1348.

225. Backes C, Ludwig N, Leidinger P, Huwer H, Tenzer S, Fehlmann T, Franke A, Meese E, Lenhof HP, Keller A. Paired proteomics, transcriptomics and miRNomics in non-small cell lung cancers: known and novel signaling cascades. Oncotarget 2016: 7(44): 71514-71525.

226. Harcourt JL, Offermann MK. Multiple signaling cascades are differentially involved in gene induction by double stranded RNA in interferon-alpha-primed cells. Eur J Biochem 2001: 268(5): 1373-1381.

227. Reunanen J. Overfitting in making comparisons between variable selection methods. J Mach Learn Res 2003: 3(Mar): 1371-1382.

228. Schaffer C. Overfitting avoidance as bias. Mach learn 1993: 10(2): 153-178.

229. Liu H, Wu X, Zhang S. Feature selection using hierarchical feature clustering. In: Proceedings of the 20th ACM international conference on Information and knowledge management; 2011; New York: Association for Computing Machinery; 2011. p. 979-984.

230. Zhou N, Wang L. A modified T-test feature selection method and its application on the HapMap genotype data. Genomics Proteomics Bioinformatics 2007: 5(3-4): 242-249.

231. Ma S, Huang J. Penalized feature selection and classification in bioinformatics. Brief Bioinform 2008: 9(5): 392-403.

232. Chong I, Jun C. Performance of some variable selection methods when multicollinearity is present. Chemometr Intell Lab Syst 2005: 78(1-2): 103-112.

233. Wang H, Lengerich BJ, Aragam B, Xing EP. Precision Lasso: accounting for correlations and linear dependencies in high-dimensional genomic data. Bioinformatics 2019: 35(7): 1181-1187.

234. Fan L, Chen S, Li Q, Zhu Z. Variable selection and model prediction based on lasso, adaptive lasso and elastic net. In: 2015 4th International Conference on Computer Science and Network Technology (ICCSNT); 2015; Harbin, China: IEEE; 2015. p. 579-583.

235. Karamizadeh S, Abdullah S, Manaf A, Zamani M, Hooman A. An overview of principal component analysis. JSIP 2013: 4(3B): 173-175.

236. Saelens W, Cannoodt R, Saeys Y. A comprehensive evaluation of module detection methods for gene expression data. Nat Commun 2018: 9(1): 1090.

237. Ogutu JO, Schulz-Streeck T, Piepho HP. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. BMC Proc 2012: 6 Suppl 2(Suppl 2): S10.

238. Omranian N, Eloundou-Mbebi JM, Mueller-Roeber B, Nikoloski Z. Gene regulatory network inference using fused LASSO on multiple data sets. Sci Rep 2016: 6: 20533.

239. Loher P, Karathanasis N. Machine Learning Approaches Identify Genes Containing Spatial Information From Single-Cell Transcriptomics Data. Front Genet 2020: 11: 612840.

240. Gupta N, Jorwal P. Treatment Outcomes Associated with Multidrug-resistant Tuberculosis. J Glob Infect Dis 2018: 10(3): 125-128.

241. Leong S, Zhao Y, Joseph NM, Hochberg NS, Sarkar S, Pleskunas J, Hom D, Lakshminarayanan S, Horsburgh CR, Jr., Roy G, Ellner JJ, Johnson WE, Salgame P. Existing blood transcriptional classifiers

accurately discriminate active tuberculosis from latent infection in individuals from south India. Tuberculosis (Edinb) 2018: 109: 41-51.

242. Meng Z, Wang M, Guo S, Zhou Y, Lyu M, Hu X, Bai H, Wu Q, Tao C, Ying B. Novel Long Non-coding RNA and LASSO Prediction Model to Better Identify Pulmonary Tuberculosis: A Case-Control Study in China. Front Mol Biosci 2021: 8: 632185.

243. Waldmann P, Mészáros G, Gredler B, Fuerst C, Sölkner J. Evaluation of the lasso and the elastic net in genome-wide association studies. Front Genet 2013: 4: 270.

244. Richards SA, Whittingham MJ, Stephens PA. Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework. Behav Ecol Sociobiol 2011: 65(1): 77-89.

245. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 2007: 8: 25.

246. Grochtmann M, Grimm K. Classification trees for partition testing. STVR 1993: 3(2): 63-82.

247. Kirasich K, Smith T, Sadler B. Random forest vs logistic regression: binary classification for heterogeneous datasets. SMU Data Science Review 2018: 1(3): 9.

248. Hengl T, Nussbaum M, Wright MN, Heuvelink GBM, Gräler B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 2018: 6: e5518.

249. Lindner C. Automated image interpretation using statistical shape models. Statistical Shape and Deformation Analysis. Elsevier, 2017; pp. 3-32.

250. Muschelli J, Betz J, Varadhan R. Binomial regression in R. Handbook of Statistics. Elsevier, 2014; pp. 257-308.

251. Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics 2018: 19(1): 270.

252. Burel JG, Babor M, Pomaznoy M, Lindestam Arlehamn CS, Khan N, Sette A, Peters B. Host Transcriptomics as a Tool to Identify Diagnostic and Mechanistic Immune Signatures of Tuberculosis. Front Immunol 2019: 10: 221.

253. MacDougall D, Crummett WB. Guidelines for data acquisition and data quality evaluation in environmental chemistry. Anal Chem 1980: 52(14): 2242-2249.

254. Scriba TJ, Fiore-Gartland A, Penn-Nicholson A, Mulenga H, Kimbung Mbandi S, Borate B, Mendelsohn SC, Hadley K, Hikuam C, Kaskar M, Musvosvi M, Bilek N, Self S, Sumner T, White RG,

Erasmus M, Jaxa L, Raphela R, Innes C, Brumskine W, Hiemstra A, Malherbe ST, Hassan-Moosa R, Tameris M, Walzl G, Naidoo K, Churchyard G, Hatherill M. Biomarker-guided tuberculosis preventive therapy (CORTIS): a randomised controlled trial. Lancet Infect Dis 2021: 21(3): 354-365.

255. Scriba TJ, Penn-Nicholson A, Shankar S, Hraha T, Thompson EG, Sterling D, Nemes E, Darboe F, Suliman S, Amon LM, Mahomed H, Erasmus M, Whatney W, Johnson JL, Boom WH, Hatherill M, Valvo J, De Groote MA, Ochsner UA, Aderem A, Hanekom WA, Zak DE. Sequential inflammatory processes define human progression from M. tuberculosis infection to tuberculosis disease. PLoS Pathog 2017: 13(11): e1006687.

256. Alam A, Imam N, Ahmed MM, Tazyeen S, Tamkeen N, Farooqui A, Malik MZ, Ishrat R. Identification and Classification of Differentially Expressed Genes and Network Meta-Analysis Reveals Potential Molecular Signatures Associated With Tuberculosis. Front Genet 2019: 10: 932.

257. Barber DL, Mayer-Barber KD, Feng CG, Sharpe AH, Sher A. CD4 T cells promote rather than control tuberculosis in the absence of PD-1-mediated inhibition. J Immunol 2011: 186(3): 1598-1607.

258. Dupnik KM, Bean JM, Lee MH, Jean Juste MA, Skrabanek L, Rivera V, Vorkas CK, Pape JW, Fitzgerald DW, Glickman M. Blood transcriptomic markers of Mycobacterium tuberculosis load in sputum. Int J Tuberc Lung Dis 2018: 22(8): 950-958.

259. Chen Y, Cao S, Sun Y, Li C. Gene expression profiling of the TRIM protein family reveals potential biomarkers for indicating tuberculosis status. Microb Pathog 2018: 114: 385-392.

260. Hibbert L, Pflanz S, De Waal Malefyt R, Kastelein RA. IL-27 and IFN-alpha signal via Stat1 and Stat3 and induce T-Bet and IL-12Rbeta2 in naive T cells. J Interferon Cytokine Res 2003: 23(9): 513-522.

261. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One 2019: 14(11): e0224365.

262. Kristman V, Manno M, Côté P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol 2004: 19(8): 751-760.

263. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ, Jr. Selection Bias Due to Loss to Follow Up in Cohort Studies. Epidemiology 2016: 27(1): 91-97.

264. Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: an overview. Ann Intern Med 2001: 135(2): 112-123.

265. Chesov D, Alexandru S, Crudu V, Ciobanu N, Botnaru V, Heyckendorf J, Lange C. Failing treatment of multidrug-resistant tuberculosis: a matter of definition. Int J Tuberc Lung Dis 2019: 23(4): 522-524.

266. Naftalin CM, Verma R, Gurumurthy M, Hee KH, Lu Q, Yeo BCM, Tan KH, Lin W, Yu B, Seng KY, Lee LS, Paton NI. Adjunctive use of celecoxib with anti-tuberculosis drugs: evaluation in a whole-blood bactericidal activity model. Sci Rep 2018: 8(1): 13491.

267. Hamilton CD, Stout JE, Goodman PC, Mosher A, Menzies R, Schluger NW, Khan A, Johnson JL, Vernon AN. The value of end-of-treatment chest radiograph in predicting pulmonary tuberculosis relapse. Int J Tuberc Lung Dis 2008: 12(9): 1059-1064.

268. Xie YL, de Jager VR, Chen RY, Dodd LE, Paripati P, Via LE, Follmann D, Wang J, Lumbard K, Lahouar S, Malherbe ST, Andrews J, Yu X, Goldfeder LC, Cai Y, Arora K, Loxton AG, Vanker N, Duvenhage M, Winter J, Song T, Walzl G, Diacon AH, Barry CE, 3rd. Fourteen-day PET/CT imaging to monitor drug combination activity in treated individuals with tuberculosis. Sci Transl Med 2021: 13(579).

269. Roe JK, Thomas N, Gil E, Best K, Tsaliki E, Morris-Jones S, Stafford S, Simpson N, Witt KD, Chain B, Miller RF, Martineau A, Noursadeghi M. Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. JCI Insight 2016: 1(16): e87238.

270. Gjøen JE, Jenum S, Sivakumaran D, Mukherjee A, Macaden R, Kabra SK, Lodha R, Ottenhoff THM, Haks MC, Doherty TM, Ritz C, Grewal HMS. Novel transcriptional signatures for sputum-independent diagnostics of tuberculosis in children. Sci Rep 2017: 7(1): 5839.

271. Gliddon HD, Kaforou M, Alikian M, Habgood-Coote D, Zhou C, Oni T, Anderson ST, Brent AJ, Crampin AC, Eley B, Heyderman R, Kern F, Langford PR, Ottenhoff THM, Hibberd ML, French N, Wright VJ, Dockrell HM, Coin LJ, Wilkinson RJ, Levin M. Identification of Reduced Host Transcriptomic Signatures for Tuberculosis Disease and Digital PCR-Based Validation and Quantification. Front Immunol 2021: 12: 637164.

272. Huang HH, Liu XY, Liang Y, Chai H, Xia LY. Identification of 13 blood-based gene expression signatures to accurately distinguish tuberculosis from other pulmonary diseases and healthy controls. Biomed Mater Eng 2015: 26 Suppl 1: S1837-1843.

273. de Araujo LS, Vaas LA, Ribeiro-Alves M, Geffers R, Mello FC, de Almeida AS, Moreira AD, Kritski AL, Lapa ESJR, Moraes MO, Pessler F, Saad MH. Transcriptomic Biomarkers for Tuberculosis:

Evaluation of DOCK9. EPHA4, and NPC2 mRNA Expression in Peripheral Blood. Front Microbiol 2016: 7: 1586.

274. Qian Z, Lv J, Kelly GT, Wang H, Zhang X, Gu W, Yin X, Wang T, Zhou T. Expression of nuclear factor, erythroid 2-like 2-mediated genes differentiates tuberculosis. Tuberculosis (Edinb) 2016: 99: 56-62.

275. Rajan JV, Semitala FC, Mehta T, Seielstad M, Montalvo L, Andama A, Asege L, Nakaye M, Katende J, Mwebe S, Kamya MR, Yoon C, Cattamanchi A. A Novel, 5-Transcript, Whole-blood Gene-expression Signature for Tuberculosis Screening Among People Living With Human Immunodeficiency Virus. Clin Infect Dis 2019: 69(1): 77-83.

276. Roe J, Venturini C, Gupta RK, Gurry C, Chain BM, Sun Y, Southern J, Jackson C, Lipman MC, Miller RF, Martineau AR, Abubakar I, Noursadeghi M. Blood Transcriptomic Stratification of Short-term Risk in Contacts of Tuberculosis. Clin Infect Dis 2020: 70(5): 731-737.

277. Walter ND, Miller MA, Vasquez J, Weiner M, Chapman A, Engle M, Higgins M, Quinones AM, Rosselli V, Canono E, Yoon C, Cattamanchi A, Davis JL, Phang T, Stearman RS, Datta G, Garcia BJ, Daley CL, Strong M, Kechris K, Fingerlin TE, Reves R, Geraci MW. Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: a Case-Control Study with Systematic Cross-Classifier Evaluation. J Clin Microbiol 2016: 54(2): 274-282.

278. Verhagen LM, Zomer A, Maes M, Villalba JA, Del Nogal B, Eleveld M, van Hijum SA, de Waard JH, Hermans PW. A predictive signature gene set for discriminating active from latent tuberculosis in Warao Amerindian children. BMC Genomics 2013: 14: 74.

279. Darboe F, Mbandi SK, Thompson EG, Fisher M, Rodo M, van Rooyen M, Filander E, Bilek N, Mabwe S, Hatherill M, Zak DE, Penn-Nicholson A, Scriba TJ. Diagnostic performance of an optimized transcriptomic signature of risk of tuberculosis in cryopreserved peripheral blood mononuclear cells. Tuberculosis (Edinb) 2018: 108: 124-126.

280. Kulkarni V, Queiroz ATL, Sangle S, Kagal A, Salvi S, Gupta A, Ellner J, Kadam D, Rolla VC, Andrade BB, Salgame P, Mave V. A Two-Gene Signature for Tuberculosis Diagnosis in Persons With Advanced HIV. Front Immunol 2021: 12: 631165.

281. Jenum S, Bakken R, Dhanasekaran S, Mukherjee A, Lodha R, Singh S, Singh V, Haks MC, Ottenhoff TH, Kabra SK, Doherty TM, Ritz C, Grewal HM. BLR1 and FCGR1A transcripts in peripheral blood associate with the extent of intrathoracic tuberculosis in children and predict treatment outcome. Sci Rep 2016: 6: 38841.

282. Day CL, Abrahams DA, Bunjun R, Stone L, de Kock M, Walzl G, Wilkinson RJ, Burgers WA, Hanekom WA. PD-1 Expression on Mycobacterium tuberculosis-Specific CD4 T Cells Is Associated With Bacterial Load in Human Tuberculosis. Front Immunol 2018: 9: 1995.

283. Hu JF, Zhang W, Zuo W, Tan HQ, Bai W. Inhibition of the PD-1/PD-L1 signaling pathway enhances innate immune response of alveolar macrophages to mycobacterium tuberculosis in mice. Pulm Pharmacol Ther 2020: 60: 101842.

284. Tezera LB, Bielecka MK, Ogongo P, Walker NF, Ellis M, Garay-Baquero DJ, Thomas K, Reichmann MT, Johnston DA, Wilkinson KA, Ahmed M, Jogai S, Jayasinghe SN, Wilkinson RJ, Mansour S, Thomas GJ, Ottensmeier CH, Leslie A, Elkington PT. Anti-PD-1 immunotherapy leads to tuberculosis reactivation via dysregulation of TNF-α. Elife 2020: 9.

285. Pan L, Wei N, Jia H, Gao M, Chen X, Wei R, Sun Q, Gu S, Du B, Xing A, Zhang Z. Genome-wide transcriptional profiling identifies potential signatures in discriminating active tuberculosis from latent infection. Oncotarget 2017: 8(68): 112907-112916.

286. Tarui M, Shindou H, Kumagai K, Morimoto R, Harayama T, Hashidate T, Kojima H, Okabe T, Nagano T, Nagase T, Shimizu T. Selective inhibitors of a PAF biosynthetic enzyme lysophosphatidylcholine acyltransferase 2. J Lipid Res 2014: 55(7): 1386-1396.

287. Morimoto R, Shindou H, Tarui M, Shimizu T. Rapid production of platelet-activating factor is induced by protein kinase Cα-mediated phosphorylation of lysophosphatidylcholine acyltransferase 2 protein. J Biol Chem 2014: 289(22): 15566-15576.

288. Singh A, Yamamoto M, Kam SH, Ruan J, Gauvreau GM, O'Byrne PM, FitzGerald JM, Schellenberg R, Boulet LP, Wojewodka G, Kanagaratham C, De Sanctis JB, Radzioch D, Tebbutt SJ. Gene-metabolite expression in blood can discriminate allergen-induced isolated early from dual asthmatic responses. PLoS One 2013: 8(7): e67907.

289. Wang J, Teng JL, Zhao D, Ge P, Li B, Woo PC, Liu CH. The ubiquitin ligase TRIM27 functions as a host restriction factor antagonized by Mycobacterium tuberculosis PtpA during mycobacterial infection. Sci Rep 2016: 6: 34827.

290. Smith CL, Dickinson P, Forster T, Craigon M, Ross A, Khondoker MR, France R, Ivens A, Lynn DJ, Orme J, Jackson A, Lacaze P, Flanagan KL, Stenson BJ, Ghazal P. Identification of a human neonatal immune-metabolic network associated with bacterial infection. Nat Commun 2014: 5: 4649.

291. Izquierdo-Bouldstridge A, Bustillos A, Bonet-Costa C, Aribau-Miralbés P, García-Gomis D, Dabad M, Esteve-Codina A, Pascual-Reguant L, Peiró S, Esteller M, Murtha M, Millán-Ariño L, Jordan

A. Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats. Nucleic Acids Res 2017: 45(20): 11622-11642.

292. Martino F, Pal M, Muñoz-Hernández H, Rodríguez CF, Núñez-Ramírez R, Gil-Carton D, Degliesposti G, Skehel JM, Roe SM, Prodromou C, Pearl LH, Llorca O. RPAP3 provides a flexible scaffold for coupling HSP90 to the human R2TP co-chaperone complex. Nat Commun 2018: 9(1): 1501.

293. Kurachi M, Barnitz RA, Yosef N, Odorizzi PM, DiIorio MA, Lemieux ME, Yates K, Godec J, Klatt MG, Regev A, Wherry EJ, Haining WN. The transcription factor BATF operates as an essential differentiation checkpoint in early effector CD8+ T cells. Nat Immunol 2014: 15(4): 373-383.

294. Guler R, Mpotje T, Ozturk M, Nono JK, Parihar SP, Chia JE, Abdel Aziz N, Hlaka L, Kumar S, Roy S, Penn-Nicholson A, Hanekom WA, Zak DE, Scriba TJ, Suzuki H, Brombacher F. Batf2 differentially regulates tissue immunopathology in Type 1 and Type 2 diseases. Mucosal Immunol 2019: 12(2): 390-402.

295. Maertzdorf J, Kaufmann SH, Weiner J, 3rd. Toward a unified biosignature for tuberculosis. Cold Spring Harb Perspect Med 2014: 5(1): a018531.

296. Thakkinstian A, McEvoy M, Chakravarthy U, Chakrabarti S, McKay GJ, Ryu E, Silvestri G, Kaur I, Francis P, Iwata T, Akahori M, Arning A, Edwards AO, Seddon JM, Attia J. The association between complement component 2/complement factor B polymorphisms and age-related macular degeneration: a HuGE review and meta-analysis. Am J Epidemiol 2012: 176(5): 361-372.

297. Sargent CA, Young C, Marsh S, Ferguson-Smith MA, Affara NA. The glycerol kinase gene family: structure of the Xp gene, and related intronless retroposons. Hum Mol Genet 1994: 3(8): 1317-1324.

298. Larrouy-Maumus G, Biswas T, Hunt DM, Kelly G, Tsodikov OV, de Carvalho LP. Discovery of a glycerol 3-phosphate phosphatase reveals glycerophospholipid polar head recycling in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 2013: 110(28): 11320-11325.

299. Diamond MS, Farzan M. The broad-spectrum antiviral functions of IFIT and IFITM proteins. Nat Rev Immunol 2013: 13(1): 46-57.

300. Mandalakas AM, Detjen AK, Hesseling AC, Benedetti A, Menzies D. Interferon-gamma release assays and childhood tuberculosis: systematic review and meta-analysis. Int J Tuberc Lung Dis 2011: 15(8): 1018-1032.

301. Reich NC. A death-promoting role for ISG54/IFIT2. J Interferon Cytokine Res 2013: 33(4): 199-205.

302. Ranjbar S, Haridas V, Jasenosky LD, Falvo JV, Goldfeld AE. A Role for IFITM Proteins in Restriction of Mycobacterium tuberculosis Infection. Cell Rep 2015: 13(5): 874-883.

303. Yang G, Xu Y, Chen X, Hu G. IFITM1 plays an essential role in the antiproliferative action of interferon-gamma. Oncogene 2007: 26(4): 594-603.

304. Causeret F, Sumia I, Pierani A. Kremen1 and Dickkopf1 control cell survival in a Wnt-independent manner. Cell Death Differ 2016: 23(2): 323-332.

305. Tornheim JA, Madugundu AK, Paradkar M, Fukutani KF, Queiroz ATL, Gupte N, Gupte AN, Kinikar A, Kulkarni V, Balasubramanian U, Sreenivasamurthy S, Raja R, Pradhan N, Shivakumar S, Valvi C, Hanna LE, Andrade BB, Mave V, Pandey A, Gupta A. Transcriptomic Profiles of Confirmed Pediatric Tuberculosis Patients and Household Contacts Identifies Active Tuberculosis, Infection, and Treatment Response Among Indian Children. J Infect Dis 2020: 221(10): 1647-1658.

306. Pai SG, Carneiro BA, Mota JM, Costa R, Leite CA, Barroso-Sousa R, Kaplan JB, Chae YK, Giles FJ. Wnt/beta-catenin pathway: modulating anticancer immune response. J Hematol Oncol 2017: 10(1): 101.

307. Lin SL, Li B, Rao S, Yeo EJ, Hudson TE, Nowlin BT, Pei H, Chen L, Zheng JJ, Carroll TJ, Pollard JW, McMahon AP, Lang RA, Duffield JS. Macrophage Wnt7b is critical for kidney repair and regeneration. Proc Natl Acad Sci U S A 2010: 107(9): 4194-4199.

308. Burgin AB, Magnusson OT, Singh J, Witte P, Staker BL, Bjornsson JM, Thorsteinsdottir M, Hrafnsdottir S, Hagen T, Kiselyov AS, Stewart LJ, Gurney ME. Design of phosphodiesterase 4D (PDE4D) allosteric modulators for enhancing cognition with improved safety. Nat Biotechnol 2010: 28(1): 63-70.

309. Subbian S, Tsenova L, O'Brien P, Yang G, Koo MS, Peixoto B, Fallows D, Dartois V, Muller G, Kaplan G. Phosphodiesterase-4 inhibition alters gene expression and improves isoniazid-mediated clearance of Mycobacterium tuberculosis in rabbit lungs. PLoS Pathog 2011: 7(9): e1002262.

310. Liu H, Palmer D, Jimmo SL, Tilley DG, Dunkerley HA, Pang SC, Maurice DH. Expression of phosphodiesterase 4D (PDE4D) is regulated by both the cyclic AMP-dependent protein kinase and mitogen-activated protein kinase signaling pathways. A potential mechanism allowing for the coordinated regulation of PDE4D activity and expression in cells. J Biol Chem 2000: 275(34): 26615-26624.

311. Yang C, Song C, Jung S, Lee K, Kim S, Lee J, Shin A, Oh J, Kwon Y, Kim H. The Phospholipase-Protein Kinase C-MEK-ERK Pathway is Essential in Mycobacteria-induced CCL3 and CCL4 Expression in Human Monocytes. Immune Netw 2005: 5(4): 237-246.

312. Shenoy AR, Wellington DA, Kumar P, Kassa H, Booth CJ, Cresswell P, MacMicking JD. GBP5 promotes NLRP3 inflammasome assembly and immunity in mammals. Science 2012: 336(6080): 481-485.

313. Fakruddin JM, Lempicki RA, Gorelick RJ, Yang J, Adelsberger JW, Garcia-Pineres AJ, Pinto LA, Lane HC, Imamichi T. Noninfectious papilloma virus-like particles inhibit HIV-1 replication: implications for immune control of HIV-1 infection by IL-27. Blood 2007: 109(5): 1841-1849.

314. Pflanz S, Timans JC, Cheung J, Rosales R, Kanzler H, Gilbert J, Hibbert L, Churakova T, Travis M, Vaisberg E, Blumenschein WM, Mattson JD, Wagner JL, To W, Zurawski S, McClanahan TK, Gorman DM, Bazan JF, de Waal Malefyt R, Rennick D, Kastelein RA. IL-27, a heterodimeric cytokine composed of EBI3 and p28 protein, induces proliferation of naive CD4+ T cells. Immunity 2002: 16(6): 779-790.

315. Abdalla AE, Li Q, Xie L, Xie J. Biology of IL-27 and its role in the host immunity against Mycobacterium tuberculosis. Int J Biol Sci 2015: 11(2): 168-175.

316. Charlot-Rabiega P, Bardel E, Dietrich C, Kastelein R, Devergne O. Signaling events involved in interleukin 27 (IL-27)-induced proliferation of human naive CD4+ T cells and B cells. J Biol Chem 2011: 286(31): 27350-27362.

317. Cheng C, Wang Q, Zhu M, Liu K, Zhang Z. Integrated analysis reveals potential long non-coding RNA biomarkers and their potential biological functions for disease free survival in gastric cancer patients. Cancer Cell Int 2019: 19: 123.

318. Schoenfeld AK, Lahrsen E, Alban S. Regulation of Complement and Contact System Activation via C1 Inhibitor Potentiation and Factor XIIa Activity Modulation by Sulfated Glycans - Structure-Activity Relationships. PLoS One 2016: 11(10): e0165493.

319. Weiner J, Domaszewska T, Donkor S, Kaufmann SHE, Hill PC, Sutherland JS. Changes in Transcript, Metabolite, and Antibody Reactivity During the Early Protective Immune Response in Humans to Mycobacterium tuberculosis Infection. Clin Infect Dis 2020: 71(1): 30-40.

320. Son M, Diamond B, Santiago-Schwarz F. Fundamental role of C1q in autoimmunity and inflammation. Immunol Res 2015: 63(1-3): 101-106.

321. Liu B, Liao J, Rao X, Kushner SA, Chung CD, Chang DD, Shuai K. Inhibition of Stat1-mediated gene activation by PIAS1. Proc Natl Acad Sci U S A 1998: 95(18): 10626-10631.

322. Yao K, Chen Q, Wu Y, Liu F, Chen X, Zhang Y. Unphosphorylated STAT1 represses apoptosis in macrophages during Mycobacteriumtuberculosis infection. J Cell Sci 2017: 130(10): 1740-1751.

323. Manca C, Tsenova L, Freeman S, Barczak AK, Tovey M, Murray PJ, Barry C, Kaplan G. Hypervirulent M. tuberculosis W/Beijing strains upregulate type I IFNs and increase expression of negative regulators of the Jak-Stat pathway. J Interferon Cytokine Res 2005: 25(11): 694-701.

324. Li X, Wang F, Xu X, Zhang J, Xu G. The Dual Role of STAT1 in Ovarian Cancer: Insight Into Molecular Mechanisms and Application Potentials. Front Cell Dev Biol 2021: 9: 636595.

325. Esquivel-Solís H, Quiñones-Falconi F, Zarain-Herzberg A, Amieva-Fernández RI, López-Vidal Y. Impaired activation of Stat1 and c-Jun as a possible defect in macrophages of patients with active tuberculosis. Clin Exp Immunol 2009: 158(1): 45-54.

326. Sugawara I, Yamada H, Mizuno S. STAT1 knockout mice are highly susceptible to pulmonary mycobacterial infection. Tohoku J Exp Med 2004: 202(1): 41-50.

327. Pan G, Bauer JH, Haridas V, Wang S, Liu D, Yu G, Vincenz C, Aggarwal BB, Ni J, Dixit VM. Identification and functional characterization of DR6, a novel death domain-containing TNF receptor. FEBS Lett 1998: 431(3): 351-356.

328. Bidère N, Su HC, Lenardo MJ. Genetic disorders of programmed cell death in the immune system. Annu Rev Immunol 2006: 24: 321-352.

329. Zeng Q, Subramaniam VN, Wong SH, Tang BL, Parton RG, Rea S, James DE, Hong W. A novel synaptobrevin/VAMP homologous protein (VAMP5) is increased during in vitro myogenesis and present in the plasma membrane. Mol Biol Cell 1998: 9(9): 2423-2437.

330. Ikezawa M, Tajika Y, Ueno H, Murakami T, Inoue N, Yorifuji H. Loss of VAMP5 in mice results in duplication of the ureter and insufficient expansion of the lung. Dev Dyn 2018: 247(5): 754-762.

331. Stow JL, Manderson AP, Murray RZ. SNAREing immunity: the role of SNAREs in the immune system. Nat Rev Immunol 2006: 6(12): 919-929.

6. Appendix: Supplement tables

Supplement table 1 Overview about biomarkers considered for treatment monitoring and outcome prediction in TB

Group	Systematic	Marker	Description	Usage
Pathogen				
	Sputum			
		TCC	Time to culture conversion	Survival prediction, treatment response, treatment outcome [19]
		TTP+	Time to positivity	Treatment response, MDR-TB suspicion when delayed [19]
		TB-RNA	Quantification of mycobacterial RNA	Treatment response [71, 107, 108]
	Sputum & others			
		TB-DNA	Mycobacterial DNA in sputum and urine	Treatment response [76]
		TB-RNA	Mycobacterial RNA in sputum	Treatment response [19]
		TB-metabolomes	Changes in mycobacterial metabolomics activity	Treatment response [19]
		TB-antigens	Mycobacterial antigens in sputum, urine and serum	Treatment response [70]
	Whole Blood			
		Whole blood bactericidal activity	Blood culture testing for bactericidal activity against M. tuberculosis	Treatment response [265]

Group	Systematic	Marker	Description	Usage
Host				
	Clinical Parameter			
		Clinical score	Evaluation of changes in patient characteristics during treatment	Treatment response [81, 82].
	Imaging			
		Chest X-rays	Changes of X-ray results	Treatment response [266]
		Chest CT scan	Changes of CT scan results	Treatment response [266]
		PET/CT scan	Changes of PET/CT scan results	Treatment response [72, 267]
	Chemokines, cytokines, proteins, and peptides			
		Acute phase proteins	Specific proteins and peptides in serum and plasma	Therapy response, treatment outcome [19, 84]
		Cytokines and chemokines	Specific cytokines and chemokines in sputum, plasma and serum	Therapy response, therapy outcome prediction [19]
		Nutritional markers	Specific markers in plasma	Determination of disease severity, treatment response [19]
	Assays and cells			
		Immunophenotype of certain cells	Differences in cytokine response of antigen-specific T-cells in whole blood and PBMC	Therapy response, outcome prediction [19]
		Immune cells without stimulation	Surface marker expression in whole blood and PBMC	Treatment response, disease severity [19]
		Antibodies	High throughput proof off circulation TB antigens in serum	Treatment response [19]

Group	Systematic	Marker	Description	Usage
Host				
	OMICS			
		Gene expression levels	RNA patterns for different signatures or candidate gene expression	Treatment response & outcome [85, 111, 117-119].
		Protein levels	Protein (patterns) from serum proteins	Treatment response [90, 91]
		Metabolomics	Pattern identification and metabolomic changes with mass spectrometry	Treatment response [94-96].
	Other			
		Volatile organic compounds	Identification of mycobacterial parts in breath	Treatment response [19]

TB= Tuberculosis; TTP = Time to culture positivity; TCC= Time to culture conversion; CT= Computer tomography; PET=Positron emission tomography, PBMC= Peripheral blood mononuclear cell

Supplement table 2 Existing tuberculosis host RNA signatures and models for diagnosis, treatment monitoring and therapy outcome predictions

Aim	Pubmed ID	Number of genes included	Population	Intended application	performance
Diagnostic					
	24785206 [175]	42	Children HIV + and -	TB vs LTBI	Sensitivity=0.83 Specificity= 0.84
	20725040 [176]	86 393	Adults HIV - HIV -	TB vs ODs TB vs LTBI and controls	p<0.001 – p<0.01
	26025597 [178]	3	Adults HIV -	TB vs. ODs and LTBI	AUC = 0.89
	29559120 [240]	24	Adults HIV -	TB vs. LTBI	AUC=0.98
	27734027 [268]	1 4	Adults HIV - HIV -	TB vs controls TB vs. ODs	AUC= 0.52 – 0.83
	28724962 [269]	7	Children HIV -	TB vs controls and ODs	AUC=0.81

Aim	Pubmed ID	Number of genes included	Population	Intended application	performance
Diagnostic					
	33763081 [270]	3	Adults HIV + and -	TB vs LTBI	AUC=0.94 – 0.97
	26405955 [271]	13	Adults HIV -	TB vs. ODs and LTBI	Sensitivity=0.92 Specificity= 0.95
	24167453 [177]	27	Adults HIV + and -	TB vs LTBI	Sensitivity=0.95 Specificity=0.90
		44		TB vs ODs	Sensitivity=0.93 Specificity=0.88
	26682570 [179]	4	Adults HIV -	TB vs controls	Sensitivity=0.88 Specificity=0.75
	27826286 [272]	1	Adults	TB vs controls	AUC=0.88
		1		TB vs controls	AUC=0.73
		1		TB vs controls	AUC=0.66

Aim	Pubmed ID	Number of genes included	Population	Intended application	performance
Diagnostic					
	27450006 [273]	17	Adults HIV -	TB vs. ODs and LTBI	AUC=0.71 – 0.88
	23940611 [193]	144	Adults HIV -	TB vs. ODs	p>0.001 – p<0.05
	30462176 [274]	5	Adults HIV +	TB vs controls (active case finding among people living with HIV)	AUC=0.87
	30919880 [275]	3	Adults HIV -	Incipient TB vs controls	PPV=5.6-50% NPV=99.3%
	29921861 [181]	20	Adults HIV -	TB vs. ODs and LTBI	AUC=0.74 – 1
	29624071 [182]	2	Adults HIV -	Incipient TB vs controls	AUC= 0.77 – 0.86
		4		Incipient TB vs controls	
	26907218, 30646264 [113, 114]	3	Adults HIV + and -	TB vs. ODs and LTBI	AUC=0.84-0.9

Aim	Pubmed ID	Number of genes included	Population	Intended application	performance
Diagnostic					
	26582831 [276]	51	Adults HIV -	TB vs LTBI	AUC=0.91-0.97
	28065665[180]	10	Adults HIV + and -	TB vs. LTBI and controls	Accuracy=0.74 – 0.93
	23375113 [277]	9	Adults HIV -	TB vs. LTBI and controls	Prediction error =11%
	27706152 [192]	380	Adults HIV -	TB vs. ODs and LTBI	AUC=0.87
	29523312 [278]	11	Adolescents HIV -	TB vs. LTBI and incipient TB and controls	AUC=0.97
	32451443 [253]	6	Adults HIV + and -	TB vs controls	AUC=0.85
	33508221 [116]	13	Adults HIV + and -	TB vs LTBI, controls and ODs	
	33692804 [279]	2	Adults HIC +	TB vs controls	AUC= 0.95 – 1

Aim	Pubmed ID	Number of genes included	Population	Intended application	performance
Diagnostic					
	32330522 [115]	7	Adults HIV -	TB vs controls TB vs- LTBI	0.86 0.88
Treatment response and monitoring					
	26907218, 30646264 [113, 114]	3	Adults HIV + and -	Baseline Treatment 1 week Treatment 2 weeks Treatment 4 weeks Treatment 6 months	Not given
	27826286 [272]	1	Adults	Controls Baseline Treatment 2 weeks Treatment 12 weeks	p=0.004
	31297103 [111]	11	Adults HIV +	Baseline Treatment 8 weeks	AUC=0.63

Aim	Pubmed ID	Number of genes included	Population	Intended application	performance
Treatment response and monitoring					
	32451443 [117]	6	Adults HIV + and -	Baseline Treatment 1 week Treatment 4 weeks	AUC=0.77 – 0.88
	27941850 [280]	11	Children HIV -	Baseline Treatment 8 weeks Treatment 24 weeks	Not given
	23940611 [193]	144	Adults HIV -	Pre-treatment vs. post-treatment Good therapy response vs inadequate therapy response	p<0.001 p<0.001
	23056259 [191]	320	Adults HIV-	Baseline treatment 2 weeks Treatment 24 weeks Treatment 48 weeks	p<0.001 – p<0.5
	29050771 [118]	13 32	Adolescents HIV -	Cure vs. failure Cure vs. failure	AUC= 0.70 – 0.99

Aim	Pubmed ID	Number of genes included	Population	Intended application	performance
Outcome prediction					
	31297103 [111]	11	Adults HIV + and -	Relapse vs- no relapse	AUC=0.72
				Culture conversion after 2 months	AUC= 0.46 – 0.73
	32451443 [117]	6	Adults HIV + and -	Cure vs failure	AUC=0.77 – 0.95
	27941850 [280]	11	Children HIV -	Smear and culture positivity after 2 and 6 months X-Ray changes BMI changes	AUC = 0.52 – 0.62
	32647325 [119]	12 + 6 clinical parameters	Adults HIV -	Cure vs failure	AUC=0.88
	26907218, 30646264 [113, 114]	3	Adults HIV + and -	Cure vs failure	AUC=0.93

HIV= Human immunodeficiency virus; +=positive, -=negative; TB=tuberculosis; LTBI= latent tuberculosis infection; AUC= area under the curve, OD=other disease, PPV=positive predictive value; NPV= negative predictive value

Supplement table 3 Overview of the signatures described in the literature, which are compared with the performance of the therapy end model described in this work.

First main authors	Signature abbreviation for this work including numbers of genes	Described comparison in original article
Anderson [175]	Anderson44	ATB vs LTBI
Berry [176]	Berry83	ATB vs LTBI and controls
Blankley [192]	Blankley4	Pulmonary TB vs. extra-pulmonary TB & Sarcoidosis
Kaforou [177]	Kaforou 27	ATB vs LTBI
	Kaforou44	ATB vs other diseases
	Kaforou53	ATB vs LTBI and other diseases
Laux da Costa [178]	Laux da Costa5	ATB vs other diseases
Maerztdorf [179]	Maertzdorf3	ATB vs LTBI and controls
Penn-Nicholson [117]	RISK6	ATB vs LTBI and controls
		Cure vs failure of DS-TB patients after therapy start
Sambaray [194]	Sambaray10	ATB vs LTBI and controls
Singhania [181]	Singhania20	ATB vs LTBI and controls
Suliman [182]	Suliman4	ATB vs LTBI
Sutherland [183]	Sutherland4	ATB vs LTBI
Sweeney [114]	Sweeney3	ATB vs LTBI and controls

First main authors	Signature abbreviation for this work including numbers of genes	Described comparison in original article
Thompson[118]	Thompson9	ATB vs LTBI
		Cure vs failure of DS-TB patients after therapy start
	Thompson16	ATB vs LTBI
		Cure vs failure of DS-TB patients after therapy start
	Thompson32	ATB vs LTBI
		Cure vs failure of DS-TB patients after therapy start
Zak [110]	Zak16	ATB vs LTBI

TB= Tuberculosis, LTBI= Latent tuberculosis infection, DS-TB= Drug-susceptible tuberculosis, ATB= Active tuberculosis

Supplement table 4 Model parameters of three steps leading to the end-of-therapy model for patients with tuberculosis.

	Therapy outcome score (TOS)	**Therapy progression score (TPS)**	**End-of-therapy (EOT) list**	**Therapy end model (TB22)**
Type	Random forest for initial gene selection, GLM for score creation, logistic regression	GLM	Lasso	RF GLM
Outcome information	Therapy outcome prediction	Progression of therapy	Clinical therapy end timepoint reached	1. Binary classification for therapy end (yes/no. Probability (P) threshold for end-of therapy ≥0.5) 2. Numeric probability value for monitoring therapy response
Variables	Dependent: score value Independent: CD274, FAM20A, LPCAT2, TRIM27, GYG1, HIST1H1B	Dependent: Remaining days of therapy at sampling timepoint Independent: RPAP3, A_33_P3281041, BATF2, C2, GK, IFIT2, IFITM1, KREMEN1, PDE4D, multidrug-resistant (yes/no) strain. Days since therapy start subtracted from calculated value.	Clinical therapy end timepoint reached Independent: 44.000 genes in transcriptomic data	Dependent: Clinical therapy end timepoint reached /probability of therapy end Independent: sub-model 1 and 2; BATF2, GBP5, IFITM1, IL27, KCNJ2-AS1, SERPING, STAT1, TNFRSF21, VAMP5
Model building data set	All GICs patients with available outcome data (cross validation, split ratio: 0.7:0.3) including DS-TB and MDR-TB patients	All GICs patients (cross validation, split ratio: 0.7:0.3) including DS-TB and MDR-TB patients	All patients from DS-GIC	All patients from DS-GIC (cross validation, split ratio: 0.7:0.3), first validation in MDR-GIC

	Therapy outcome score (TOS)	Therapy progression score (TPS)	End-of-therapy (EOT) list	Therapy end model (TB22)
Model performance parameter	Overall GLM model's p-value <0.001 -AUC for TOS in therapy naïve patients for the ability to differentiate between cure and failure: 0.85 (95% CI: 0.78 – 0.92) -AUC for TOS in therapy naïve patients for the ability to differentiate between survivors and deceased: 0.96 (0.88 –1) -AUC for TOS in therapy naïve patients to predict smear conversion before and after 2 months: 0.59 (0.44 – 0.73) -AUC for TOS in therapy naïve patients to predict culture conversion before and after 2 months: 0.70 (0.55 – 0.85) -AUC for TOS in therapy naïve patients to predict culture conversion before and after 6 months: 0.56 (0.36 – 0.77) -AUC for TOS to correlate with the presence of positive culture at sampling timepoint: 0.66 (0.53 – 0.80)	Overall GLM model's p-Value: <0.001 Coefficient of determination R^2: 0.53	64 out of 44.000 targets with $\beta>0$	Out of bag estimate of error rate: 5.78% Classification error specificity: 0.20 Classification error sensitivity: 0.02 Mean decrease in Gini: TPS: 23.4 TNFRSF21: 4.74 BATF2: 4.73 IFITM1: 4.62 IL27: 3.32 GBP5: 3.17 SERPING1: 2.81 STAT1: 2.32 KCNJ2-AS1: 2.20 VAMP5: 2.1 TOS: 1.89 Overall GLM model's p-value: <0.001 Coefficient of determination R^2: 0.64 Difference between RF-Probabilities (μP=0.244) and GLM Model

					(μP=0.234) not significant with p=0.686

RF= Random forest, GLM= generalized linear model, GIC= German identification cohort, DS-TB= Drug-susceptible tuberculosis, MDR-TB= Multidrug-resistant tuberculosis, TOS= Therapy outcome score, TPS= Therapy progression score, EOT= End of therapy, AUC= Area under the curve [143]

Supplement table 5 Involved KEGG pathways for different comparisons

KEGG pathway identifier	Pathway	Category	Number of genes	Upregulated genes	Downregulated genes	Contrast
path:hsa00010	Glycolysis /gluconeogenesis	Metabolism	55	6	18	Before culture conversion vs. after culture conversion
				4	15	Therapy naive patients vs. patients under therapy
				2	5	Healthy controls vs. calculated therapy end
path:hsa00020	Citrate cycle (TCA cycle)	Metabolism	29	0	1	Disconcordance between observed and calculated therapy end
path:hsa00030	Pentose phosphate pathway	Metabolism	22	0	3	Healthy controls vs. calculated therapy end
path:hsa00051	Fructose and mannose metabolism	Metabolism	28	7	2	DS-TB patients vs. MDR-TB patients at baseline
path:hsa00061	Fatty acid biosynthesis	Metabolism	10	1	0	Favourable outcomes vs unfavourable outcomes at baseline
				1	1	Patients with cavitary disease vs patients without cavitary disease at baseline
path:hsa00062	Fatty acid elongation	Metabolism	12	3	0	Healthy controls vs. TB patients at baseline
				3	0	Healthy controls vs. calculated therapy end
path:hsa00100	Steroid biosynthesis	Metabolism	17	0	1	Favourable outcomes vs unfavourable

						outcomes at baseline
path:hsa00120	Primary bile acid biosynthesis	Metabolism	15	3	1	Healthy controls vs. calculated therapy end
path:hsa00130	Ubiquinone and other terpenoid-quinone biosynthesis	Metabolism	6	1	0	Patients under therapy vs. patients after therapy end
path:hsa00230	Purine metabolism	Metabolism	98	2	8	Healthy controls vs. calculated therapy end
path:hsa00240	Pyrimidine metabolism	Metabolism	37	2	1	Patients under therapy vs. patients after therapy end
path:hsa00260	Glycine, serine and threonine metabolism	Metabolism	32	2	0	Patients under therapy vs. patients after therapy end
path:hsa00270	Cysteine and methionine metabolism	Metabolism	38	2	0	Patients under therapy vs. patients after therapy end
path:hsa00280	Valine, leucine and isoleucine degradation	Musculoskeletal	39	4	9	Therapy naive patients vs. patients under therapy
path:hsa00290	Valine, leucine and isoleucine biosynthesis	Musculoskeletal	3	1	0	Patients under therapy vs. patients after therapy end
path:hsa00310	Lysine degradation	Musculoskeletal	36	1	0	Disconcordance between observed and calculated therapy end
path:hsa00400	Phenylalanine, tyrosine and tryptophan biosynthesis	Musculoskeletal	4	1	0	Patients under therapy vs. patients after therapy end
path:hsa00440	Phosphonate and phosphinate	Musculoskeletal	4	0	3	Before culture conversion vs. after culture

	metabolism					conversion
path:hsa00520	Amino sugar and nucleotide sugar metabolism	Metabolism	31	0	1	Favourable outcomes vs unfavourable outcomes at baseline
				10	5	Before culture conversion vs. after culture conversion
path:hsa00524	Neomycin, kanamycin and gentamicin biosynthesis	Metabolism	4	1	0	Patients with cavitary disease vs patients without cavitary disease at baseline
path:hsa00561	Glycerolipid metabolism	Metabolism	36	1	0	Favourable outcomes vs unfavourable outcomes at baseline
path:hsa00562	Inositol phosphate metabolism	Cell communication	55	0	1	Favourable outcomes vs unfavourable outcomes at baseline
path:hsa00620	Pyruvate metabolism	Metabolism	41	0	1	Disconcordance between observed and calculated therapy end
				1	11	Therapy naive patients vs. patients under therapy
path:hsa00630	Glyoxylate and dicarboxylate metabolism	Metabolism	20	0	1	Disconcordance between observed and calculated therapy end
path:hsa00740	Riboflavin metabolism	Metabolism	6	0	4	Before culture conversion vs. after culture conversion
path:hsa00740	Riboflavin metabolism	Metabolism	6	1	0	Patients under therapy vs. patients after therapy end

path:hsa00860	Porphyrin and chlorophyll metabolism	Metabolism	28	2	0	Patients under therapy vs. patients after therapy end
path:hsa00910	Nitrogen metabolism	Metabolism	16	3	0	Healthy controls vs. TB patients at baseline
path:hsa01040	Biosynthesis of unsaturated fatty acids	Metabolism	14	3	1	Healthy controls vs. TB patients at baseline
				3	0	Healthy controls vs. calculated therapy end
path:hsa01200	Carbon metabolism	Metabolism	97	3	0	Patients with cavitary disease vs patients without cavitary disease at baseline
				0	2	Disconcordance between observed and calculated therapy end
path:hsa01210	2-Oxocarboxylic acid metabolism	Metabolism	14	0	1	Disconcordance between observed and calculated therapy end
path:hsa01230	Biosynthesis of amino acids	Metabolism	63	2	0	Patients with cavitary disease vs patients without cavitary disease at baseline
path:hsa01521	EGFR tyrosine kinase inhibitor resistance	Cell communication	74	7	2	Healthy controls vs. TB patients at baseline
path:hsa03008	Ribosome biogenesis in eukaryotes	DNA/RNA regulation	43	4	15	Before culture conversion vs. after culture conversion
path:hsa03010	Ribosome	DNA/RNA regulation	95	8	4	Healthy controls vs. TB patients at baseline
path:hsa03022	Basal transcription factors	DNA/RNA	36	4	1	Healthy controls vs. calculated therapy end

		regulation				
path:hsa03030	DNA replication	DNA/RNA regulation	30	4	1	Healthy controls vs. calculated therapy end
path:hsa03060	Protein export	Cell life cycle	17	1	0	Favourable outcomes vs unfavourable outcomes at baseline
path:hsa03320	PPAR signaling pathway	DNA/RNA regulation	65	2	14	Therapy naive patients vs. patients under therapy
path:hsa03420	Nucleotide excision repair	DNA/RNA regulation	40	1	0	Disconcordance between observed and calculated therapy end
				5	1	Healthy controls vs. calculated therapy end
path:hsa03430	Mismatch repair	DNA/RNA regulation	22	8	2	Before culture conversion vs. after culture conversion
				3	0	Therapy naive patients vs. patients after two weeks of therapy
				7	0	Therapy naive patients vs. patients under therapy
path:hsa03440	Homologous recombination	DNA/RNA regulation	32	9	1	Therapy naive patients vs. patients under therapy
path:hsa03460	Fanconi anemia pathway	Cell life cycle	31	2	0	Patients with cavitary disease vs patients without cavitary disease at baseline

path:hsa04010	MAPK signaling pathway	Neuro transmission	272	26	42	Therapy naive patients vs. patients under therapy
path:hsa04015	Rap1 signaling pathway	Neuro transmission	186	4	0	Patients with cavitary disease vs patients without cavitary disease at baseline
				1	7	Patients under therapy vs. patients after therapy end
				23	31	Therapy naive patients vs. patients under therapy
				14	4	Healthy controls vs. TB patients at baseline
path:hsa04022	cGMP-PKG signaling pathway	Neuro transmission	152	22	7	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04024	cAMP signaling pathway	Neuro transmission	195	9	0	Therapy naive patients vs. patients after two weeks of therapy
path:hsa04068	FoxO signaling pathway	Cell life cycle	119	14	21	Therapy naive patients vs. patients under therapy
				10	3	Healthy controls vs. TB patients at baseline
				17	5	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04070	Phosphatidylinositol signaling system	Cell communication	73	0	1	Favourable outcomes vs unfavourable outcomes at baseline
path:hsa04071	Sphingolipid signaling pathway	Cell	102	10	1	Healthy controls vs. TB patients at baseline

		communication				
path:hsa04080	Neuroactive ligand-receptor interaction	Neuro transmission	301	51	67	Before culture conversion vs. after culture conversion
path:hsa04110	Cell cycle	Cell life cycle	116	0	5	Patients under therapy vs. patients after therapy end
path:hsa04130	SNARE interactions in vesicular transport	Neuro transmission	26	1	2	Patients under therapy vs. patients after therapy end
path:hsa04136	Autophagy - other	Cell life cycle	18	3	0	Healthy controls vs. calculated therapy end
path:hsa04140	Autophagy - animal	Cell life cycle	101	9	2	Healthy controls vs. calculated therapy end
path:hsa04144	Endocytosis	Cell life cycle	178	22	7	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04145	Phagosome	Immune system	120	1	2	Therapy naive patients vs. patients after two weeks of therapy
path:hsa04150	mTOR signaling pathway	Immune system	113	15	21	Therapy naive patients vs. patients under therapy
path:hsa04210	Apoptosis	Cell life cycle	124	0	2	Favourable outcomes vs unfavourable outcomes at baseline
path:hsa04218	Cellular senescence	Cell life cycle	140	0	7	Patients under therapy vs. patients after therapy end
path:hsa04270	Vascular smooth muscle contraction	Musculoskeletal	112	2	0	Favourable outcomes vs unfavourable outcomes at baseline

				17	8	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04310	Wnt signaling pathway	Cell communication	129	17	14	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04330	Notch signaling pathway	Neuro transmission	45	0	1	Favourable outcomes vs unfavourable outcomes at baseline
path:hsa04340	Hedgehog signaling pathway	Neuro transmission	37	3	4	Healthy controls vs. calculated therapy end
				4	12	Before culture conversion vs. after culture conversion
				5	1	Healthy controls vs. TB patients at baseline
path:hsa04350	TGF-beta signaling pathway	Neuro transmission	79	12	9	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04360	Axon guidance	Neuro transmission	143	3	6	Patients under therapy vs. patients after therapy end
				15	27	Therapy naive patients vs. patients under therapy
path:hsa04371	Apelin signaling pathway	Tissue-related function	118	10	5	Healthy controls vs. calculated therapy end
			118	0	2	Disconcordance between observed and calculated therapy end
path:hsa04512	ECM-receptor interaction	Neuro transmission	78	6	2	Healthy controls vs. calculated therapy end
				0	2	Therapy naive patients vs. patients after two

						weeks of therapy
path:hsa04514	Cell adhesion molecules	Neuro transmission	119	2	3	Therapy naive patients vs. patients after two weeks of therapy
path:hsa04520	Adherens junction	Cell life cycle	64	4	14	Therapy naive patients vs. patients under therapy
path:hsa04550	Signaling pathways regulating pluripotency of stem cells	Cell life cycle	126	17	24	Therapy naive patients vs. patients under therapy
path:hsa04613	Neutrophil extracellular trap formation	Immune system	153	2	6	Patients under therapy vs. patients after therapy end
path:hsa04620	Toll-like receptor signaling pathway	Immune system	94	9	0	Healthy controls vs. TB patients at baseline
path:hsa04622	RIG-I-like receptor signaling pathway	Immune system	59	6	0	Healthy controls vs. TB patients at baseline
path:hsa04650	Natural killer cell mediated cytotoxicity	Immune system	114	3	0	Patients with cavitary disease vs patients without cavitary disease at baseline
path:hsa04659	Th17 cell differentiation	Immune system	99	2	1	Favourable outcomes vs unfavourable outcomes at baseline
				19	7	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04660	T cell receptor signaling pathway	Immune system	100	2	3	Therapy naive patients vs. patients after two weeks of therapy

path:hsa04664	Fc epsilon RI signaling pathway	Immune system	63	2	0	Patients with cavitary disease vs patients without cavitary disease at baseline
path:hsa04670	Leukocyte transendothelial migration	Immune system	95	4	0	Patients with cavitary disease vs patients without cavitary disease at baseline
path:hsa04720	Long-term potentiation	Neuro transmission	64	6	18	Before culture conversion vs. after culture conversion
path:hsa04723	Retrograde endocannabinoid signaling	Neuro transmission	126	2	0	Favourable outcomes vs unfavourable outcomes at baseline
				19	33	Before culture conversion vs. after culture conversion
				0	2	Disconcordance between observed and calculated therapy end
				0	7	Patients under therapy vs. patients after therapy end
path:hsa04724	Glutamatergic synapse	Neuro transmission	95	4	7	Healthy controls vs. calculated therapy end
				22	12	Therapy naive patients vs. patients under therapy
				14	12	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04725	Cholinergic synapse	Neuro transmission	99	25	15	Before culture conversion vs. after culture conversion

				0	2	Disconcordance between observed and calculated therapy end
path:hsa04727	GABAergic synapse	Neuro transmission	75	0	2	Disconcordance between observed and calculated therapy end
				1	4	Patients under therapy vs. patients after therapy end
				5	0	Therapy naive patients vs. patients after two weeks of therapy
path:hsa04740	Olfactory transduction	Neuro transmission	92	7	5	Healthy controls vs. calculated therapy end
path:hsa04742	Taste transduction	Neuro transmission	51	5	15	Before culture conversion vs. after culture conversion
path:hsa04744	Photo transduction	Neuro transmission	24	1	0	Disconcordance between observed and calculated therapy end
path:hsa04910	Insulin signaling pathway	Hormones	121	2	0	Favourable outcomes vs unfavourable outcomes at baseline
				3	1	Patients with cavitary disease vs patients without cavitary disease at baseline
				17	22	Therapy naive patients vs. patients under therapy

path:hsa04911	Insulin secretion	Hormones	78	5	0	Therapy naive patients vs. patients after two weeks of therapy
path:hsa04911	Insulin secretion	Hormones	78	13	8	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04913	Ovarian steroidogenesis	Hormones	43	6	13	Before culture conversion vs. after culture conversion
path:hsa04916	Melanogenesis	Hormones	86	4	7	Healthy controls vs. calculated therapy end
path:hsa04917	Prolactin signaling pathway	Hormones	67	8	19	Before culture conversion vs. after culture conversion
				9	3	Healthy controls vs. TB patients at baseline
path:hsa04918	Thyroid hormone synthesis	Hormones	64	4	0	Therapy naive patients vs. patients after two weeks of therapy
				1	6	Healthy controls vs. TB patients at baseline
path:hsa04925	Aldosterone synthesis and secretion	Hormones	85	5	1	Therapy naive patients vs. patients after two weeks of therapy
path:hsa04926	Relaxin signaling pathway	Hormones	112	8	3	Healthy controls vs. calculated therapy end
path:hsa04929	GnRH secretion	Hormones	61	2	0	Patients with cavitary disease vs patients without cavitary disease at baseline
path:hsa04930	Type II diabetes mellitus	Hormones	43	2	0	Patients with cavitary disease vs patients without cavitary disease at baseline
				0	1	Disconcordance between observed and

						calculated therapy end
path:hsa04933	AGE-RAGE signaling pathway in diabetic complications	Hormones	96	14	7	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04934	Cushing syndrome	Hormones	136	2	2	Patients with cavitary disease vs patients without cavitary disease at baseline
path:hsa04935	Growth hormone synthesis, secretion and action	Hormones	110	16	7	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04960	Aldosterone-regulated sodium reabsorption	Hormones	36	6	12	Before culture conversion vs. after culture conversion
				3	0	Therapy naive patients vs. patients after two weeks of therapy
				4	9	Therapy naive patients vs. patients under therapy
				7	3	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04962	Vasopressin-regulated water reabsorption	Hormones	34	1	0	Disconcordance between observed and calculated therapy end
path:hsa04966	Collecting duct acid secretion	Metabolism	20	1	0	Favourable outcomes vs unfavourable outcomes at baseline
path:hsa04971	Gastric acid secretion	Metabolism	66	0	1	Favourable outcomes vs unfavourable outcomes at baseline

path:hsa04973	Carbohydrate digestion and absorption	Metabolism	40	2	0	Patients with cavitary disease vs patients without cavitary disease at baseline
				4	0	Therapy naive patients vs. patients after two weeks of therapy
path:hsa04974	Protein digestion and absorption	Metabolism	77	9	2	Healthy controls vs. calculated therapy end
path:hsa04975	Fat digestion and absorption	Metabolism	32	1	0	Favourable outcomes vs unfavourable outcomes at baseline
path:hsa04976	Bile secretion	Metabolism	70	6	20	Before culture conversion vs. after culture conversion
				6	14	Therapy naive patients vs. patients under therapy
				11	7	DS-TB patients vs. MDR-TB patients at baseline
path:hsa04978	Mineral absorption	Metabolism	47	3	5	Healthy controls vs. TB patients at baseline
path:hsa05014	Amyotrophic lateral sclerosis	Neurological disease	283	19	7	Healthy controls vs. TB patients at baseline
path:hsa05022	Pathways of neurodegeneration - multiple diseases	Neurological disease	398	25	10	Healthy controls vs. TB patients at baseline
path:hsa05031	Amphetamine addiction	Neurological disease	59	11	6	DS-TB patients vs. MDR-TB patients at baseline
path:hsa05032	Morphine addiction	Neurological	81	0	2	Disconcordance between observed and

		disease				calculated therapy end
				0	4	Patients under therapy vs. patients after therapy end
path:hsa05033	Nicotine addiction	Neurological disease	32	3	14	Before culture conversion vs. after culture conversion
				0	2	Disconcordance between observed and calculated therapy end
				0	6	Patients under therapy vs. patients after therapy end
path:hsa05100	Bacterial invasion of epithelial cells	Infectious disease	66	2	0	Patients with cavitary disease vs patients without cavitary disease at baseline
path:hsa05110	Vibrio cholerae infection	Infectious disease	41	0	1	Favourable outcomes vs unfavourable outcomes at baseline
				3	13	Before culture conversion vs. after culture conversion
path:hsa05120	Epithelial cell signaling in Helicobacter pylori infection	Infectious disease	58	1	2	Therapy naive patients vs. patients after two weeks of therapy
path:hsa05131	Shigellosis	Infectious disease	202	4	0	Patients with cavitary disease vs patients without cavitary disease at baseline
				2	0	Disconcordance between observed and calculated therapy end

path:hsa05142	Chagas disease	Infectious disease	97	15	7	DS-TB patients vs. MDR-TB patients at baseline
path:hsa05161	Hepatitis B	Infectious disease	150	20	9	DS-TB patients vs. MDR-TB patients at baseline
path:hsa05164	Influenza A	Infectious disease	150	13	6	Healthy controls vs. TB patients at baseline
path:hsa05166	Human T-cell leukemia virus 1 infection	Infectious disease	202	25	10	DS-TB patients vs. MDR-TB patients at baseline
path:hsa05168	Herpes simplex virus 1 infection	Infectious disease	250	34	56	Before culture conversion vs. after culture conversion
				5	3	Therapy naive patients vs. patients after two weeks of therapy
path:hsa05171	Coronavirus disease - COVID-19	Infectious disease	213	19	6	Healthy controls vs. TB patients at baseline
path:hsa05200	Pathways in cancer	Cancer	493	0	4	Disconcordance between observed and calculated therapy end
path:hsa05202	Transcriptional misregulation in cancer	Cancer	170	0	2	Disconcordance between observed and calculated therapy end
path:hsa05205	Proteoglycans in cancer	Cancer	190	4	0	Patients with cavitary disease vs patients without cavitary disease at baseline

				19	35	Therapy naive patients vs. patients under therapy
path:hsa05206	MicroRNAs in cancer	Cancer	151	0	2	Favourable outcomes vs unfavourable outcomes at baseline
path:hsa05210	Colorectal cancer	Cancer	82	8	3	Healthy controls vs. TB patients at baseline
				12	4	DS-TB patients vs. MDR-TB patients at baseline
path:hsa05213	Endometrial cancer	Cancer	55	2	0	Patients with cavitary disease vs patients without cavitary disease at baseline
path:hsa05214	Glioma	Cancer	67	0	4	Patients under therapy vs. patients after therapy end
				4	0	Therapy naive patients vs. patients after two weeks of therapy
path:hsa05216	Thyroid cancer	Cancer	34	2	0	Patients with cavitary disease vs patients without cavitary disease at baseline
				0	1	Disconcordance between observed and calculated therapy end
path:hsa05217	Basal cell carcinoma	Cancer	52	0	1	Favourable outcomes vs unfavourable outcomes at baseline
path:hsa05220	Chronic myeloid leukaemia	Cancer	73	6	15	Therapy naive patients vs. patients under therapy

path:hsa05230	Central carbon metabolism in cancer	Cancer	64	1	6	Healthy controls vs. calculated therapy end
				2	0	Patients with cavitary disease vs patients without cavitary disease at baseline
				4	0	Therapy naive patients vs. patients after two weeks of therapy
path:hsa05231	Choline metabolism in cancer	Cancer	83	3	1	Patients under therapy vs. patients after therapy end
				5	0	Therapy naive patients vs. patients after two weeks of therapy
path:hsa05320	Autoimmune thyroid disease	Autoimmune disease	52	13	1	Therapy naive patients vs. patients under therapy
path:hsa05321	Inflammatory bowel disease	Autoimmune disease	57	11	5	DS-TB patients vs. MDR-TB patients at baseline
path:hsa05323	Rheumatoid arthritis	Autoimmune disease	82	0	2	Therapy naive patients vs. patients after two weeks of therapy
path:hsa05340	Primary immunodeficiency	Autoimmune disease	33	0	1	Favourable outcomes vs unfavourable outcomes at baseline
				1	2	Therapy naive patients vs. patients after two weeks of therapy
path:hsa05415	Diabetic cardiomyopathy	Cardiovascular	187	21	44	Before culture conversion vs. after culture

		disease				conversion
path:hsa05416	Viral myocarditis	Cardiovascular disease	57	3	6	Healthy controls vs. TB patients at baseline
path:hsa05418	Fluid shear stress and atherosclerosis	Cardiovascular disease	128	2	2	Therapy naive patients vs. patients after two weeks of therapy

Supplement table 6 Presentation of the genes in the signature with regard to their biological function, pathways and previous mention in the scientific literature with regard to tuberculosis

Gene	Name	Short description	Previous TB-Biomarker or immunological related description in literature	Related Pathways
CD274	Programmed cell death 1 ligand	Protein coding gene. The protein PD-L1 expressed by CD274 binds to the PD-1 receptor, which is found on activated T and B cells. In activated CD4 cells PD-1 is upregulated and the binding of PD-L1 can thereby induce the production of IL-10 [62, 63].	• Upregulated in patients with active TB compared to latent controls [254] • Upregulated in latent infected compared to healthy controls [255] • Potential biomarker for Progress from latent to active TB [176, 181] • Correlation between CD274 and detectable acid-fast bacilli on sputum smear microscopy [281] • Increased sensitivity to infection in PD-1 knock out mice with equal CD4 cell count [256] • Actor in programmed cell-death [282] CD274-inhibitor immune therapy was shown to be associated with increased growth of *M. tuberculosis* [283]	•Translocation of ZAP-70 to immunological synapse •T-cell co-signaling pathway: ligand-receptor interactions •CD28 co-stimulation •Cell adhesion molecules (CAMs) •T-cell receptor signaling pathway •NF-kappaB Signaling •Class I MHC mediated antigen processing and presentation •Innate immune system
FAM20A	Golgi associated secretory pathway	Protein coding gene. FAM20A belongs to the secreted proteins and is found in many tissues. It is	• Upregulated in patients with active TB compared to latent controls [254]	•Regulation of insulin-like growth factor (IGF) transport and uptake by insulin-like growth factor

	Pseudokinase	involved in haematopoiesis and also plays an important role in biomineralization [64-66].	• Potential biomarker for progress from latent to active TB [176, 181, 284] • Possible diagnostic biomarker regardless of HIV status [177]	binding proteins (IGFBPs) • Regulation of insulin-like growth factor (IGF) transport and uptake by insulin-like growth factor binding proteins (IGFBPs) •Post-translational protein phosphorylation
LPCAT2	Lyso-phosphatidylcholine Acyltransferase 2	Protein coding gene. Acts as a biocatalyst for the biosynthesis of the platelet-activating factor and the glycerophospholipid precursor [285, 286].	• Upregulated in patients with active TB compared to latent controls [254] • Possible diagnostic biomarker ability [176] • Upregulated in latent infected compared to healthy controls [255] • Link between LPCAT2 and Asthma has been shown, so might be an indicator in pathogenic lung functions [287]	•Acyl chain remodelling of PE •Glycerophospholipid biosynthesis •Metabolism
TRIM27	Tripartite Motif Containing 27	Protein coding gene. This gene encodes a protein with zinc-binding domains, a V-Box type 1 and 2, a RING and a coiled-coil region. TRIM27 has a repressive effect on cellular transcriptions [69, 70]	• Influence of TRIM 27 at macrophages and survival ability of M. tuberculosis [288] • Potential Biomarker for progress from latent to active TB [254, 258] • Host restriction factor antagonized by *M. tuberculosis* [288]	•PI3/AKT signaling •PTEN regulation •Infection with *M. tuberculosis*
GYG1	Glycogenin-1 / glycogenin gkycosyltransferase	Protein coding gene. Forms oligosaccharide primers for the synthesis of glycogen by autoglycolisation and provides a	• GYG1 was significantly different expressed in patients with low bacterial load in comparison to patients with high bacterial load [257]	•Glycogen storage disease type 0 (muscle GYS1) •Glycogen storage diseases •Glycogen metabolism

		core protein of glycogen as a result of the whole process [61].	• Upregulated in patients with active TB compared to latent controls [254] • Correlates with infection risk of newborns; suspected to be linked to innate immune system [257, 289]	•Galactose metabolism •Diseases of metabolism •Glucose metabolism •Angiopoietin like protein 8 regulatory pathway •Glycosaminoglycan metabolism •HIV life cycle •Innate immune system •Metabolism
HIST1H1B	H1.5 Linker Histone	Protein coding gene. As cluster member HIST1H1B is involved, among other things, in the regulation of transcriptome and the compaction of chromatin in cells [290].	• Discriminative ability between memory CD4 cells in latent infected and healthy controls [251]	•Apoptosis induced DNA fragmentation •CDK-mediated phosphorylation and removal of Cdc6 •DNA damage/telomere stress induced senescence •Granzyme-A pathway •Cellular senescence
A_33_P332 71041	-	-	-	
RPAP3	RNA polymerase II-associated protein 3	Protein coding gene. Provide a part of the RNA and transcription processing apparatus [291].	• Upregulated in patients with active TB compared to latent controls [254]	•NF-kappaB signaling
BATF2	Basic leucine zipper	Protein coding gene. BATF2 is	• Upregulated in patients with active TB	•T-cell receptor signaling pathway

	ATF-like transcription factor 2	expressed in immune cells and is involved in the control of linear specific differentiation of these cells [292].	compared to latent controls [254] • Diagnostic biomarker ability [110, 176, 181, 275, 293, 294] • Possible biomarker for treatment response/outcome prediction [118] • Upregulated in latent infected compared to healthy controls [255] • plays a role in the differentiation of interleukin-17 producing helper T-cells and CD8 cells [292]	•PD-L1 expression and PD-1 checkpoint pathway in cancer
C2	Complement component 2	Protein encoding gene for serum glycoprotein. As part of the classical component pathways, it plays a role in both the innate and the acquired immune system [295].	• Upregulated in patients with active TB compared to latent controls [254]	•Complement pathway •Immune response lectin induced complement pathway •Creation of C4 and C4 activators •Innate immune system
GK3P	Glycerol kinase 3 pseudogene	Pseudogen, which as a key enzyme plays an important role in metabolism and in particular in glycerol absorption [296].	• Upregulated in patients with active TB compared to latent controls [254] • Influence on glycerol processing. M. *tuberculosis* is able to recycle the glycerol 3-phosphates resulting from glycerol to glycerophospholipid [297] • Increased need for glycerol in patients metabolism with active TB [297]	•Polyol metabolism •Glycerol degradation via glycerol kinase pathway •Sn-glycerol 3-phosphate from glycerol

IFIT2	Interferon induced protein with tetratricopeptide repeats 2	Protein coding gene. Part of interferon gamma signaling pathway. Therefore it plays an important role in cytokine regulation in adapted and innate immune system [298].	• Upregulated in patients with active TB compared to latent controls [254] • Possible diagnostic biomarker ability [176] • Upregulated in latent infected compared to healthy controls [255] • Key role in in interferon-γ signaling with *M. tuberculosis* specific interferon-γ response (IGRA) [299, 300]	•Type II interferon signaling (IFNG) •Immune response IFN alpha/beta signaling pathway •Interferon gamma signaling •Cytokine signaling in immune system •Innate immune system
IFITM1	Interferon Induced Transmembrane Protein 1	Protein coding gene. Protein inhibits cell infection [298].	• Upregulated in patients with active TB compared to latent controls [254] • Expression level correlates with bacterial load [301] • Possible diagnostic biomarker ability [176] • Possible marker for therapy response [191] • Upregulated in latent infected compared to healthy controls [255] • Correlation with *M. tuberculosis* [301] • T-cell death is supported and the survival of mycobacteria is inhibited [288] • Key role in in interferon-γ signaling with *M. tuberculosis* specific interferon-γ response (IGRA) [299, 302]	•Immune response IFN alpha/beta signaling pathway •B cell receptor signaling pathway (KEGG) •Immunoregulatory interactions between a lymphoid and a non-lymphoid cell •Interferon gamma signaling •Cytokine signaling in immune system •Class I MHC mediated antigen processing and presentation •Innate immune system

KREMEN1	Kringle Containing Transmembrane Protein 1	Protein coding gene, binds to Dickkopf1 genes, which makes KREMEN 1 to an important Wnt/β catenin signaling inhibitor [303].	• Upregulated in patients with active TB compared to latent controls [254] • Possible biomarker in paediatric TB [304] • Possible diagnostic biomarker ability [176] • KREMEN1 is an important protein in Wnt/β-catenin signaling [303], which controls the key functions of all cells - proliferation, differentiation, migration, genetic stability, apoptosis, and stem cell renewal [305]. • Association between Wnt/β-catenin signaling and macrophages is also known [306]	•Negative regulation of TCF-dependent signaling by WNT ligand antagonists •Wnt signaling network •Misspliced GSK3beta mutants stabilize beta-catenin •Wnt signaling pathway and pluripotency •WNT signaling •HIV life cycle •Signaling by GPCR
PDE4D	Phosphodiesterase 4D	Protein coding gene whose enzyme hydrolyzes cAMP and cGMP in cells [307]	• Modifies gene expression and improves isoniazide-mediated clearance of *M. tuberculosis* in rabbit lungs [308] • PDE4D is regulated by the protein kinase C (PKC)-Raf-MEK-ERK [309]. This protein kinase C zeta signaling cascade also plays an important role in the extracellular activation of monocytes and macrophages via the TOLL-2 receptor during infection with *M. tuberculosis* [310].	•Neurophysiological process glutamate regulation of dopamine D1A receptor signaling •Human thyroid stimulating hormone (TSH) signaling pathway •Signal transduction •PKA signaling •Parathyroid hormone synthesis, secretion and action •Regulation of CFTR activity (norm and CF)

				•ADP signaling through P2Y purinoceptor 12 •G alpha (s) signaling events •Metabolism of nucleotides •GABAergic synapse •Myometrial relaxation and contraction pathways •cAMP signaling pathway •DAG and IP3 signaling •Sweet taste Signaling •Signaling by GPCR •Metabolism
GPB5	Guanylate binding protein 5	Protein-coding gene that acts as an activator of the NLRP3 inflammosome in the innate immune system and is thus involved in the immune response to inflammatory processes [311].	• Upregulated in patients with active TB compared to latent controls [254] • Possible biomarker for progress from latent to active TB [110, 176, 181, 294] • Discrimitative ability between healthy persons and TB patients [257] • Diagnostic discrimination between TB and other pulmonary diseases [178] • Possible biomarker for treatment response/outcome prediction [118] • Upregulated in latent infected compared to healthy controls [255]	•NF-kB (NFkB) Pathway •NOD-like receptor signaling pathway •Interferon gamma signaling •Cytokine signaling in immune system •Innate immune system

			• Possible diagnostic biomarker regardless of HIV status [177]	
IL27	Interleukin 27	Protein-coding gene with pro- and anti-inflammatory properties. Influences formation, proliferation suppression and stimulation of T-cells and B-cells of the innate immune system [259, 312, 313].	• Upregulated in patients with active TB compared to latent controls [254] • Inhibits autophagy formation, production of pro- inflammatory cytokines and phagosomal acidification in macrophages. Supresses Th1 cell response and Th17 cell development; induces IL-10 producing Tr1 cells in T-cells [314] • Regulation of naïve B- and T-cells [315]	•Th1 differentiation pathway •Interleukin-6 family signaling •PEDF induced signaling •Cytokine signaling in immune system •Innate immune system
KCNJ2-AS1	Potassium inwardly rectifying channel subfamily J member 2 - antisense RNA 1	Long non-coding RNA [316]	• -	•GABA receptor activation
SERPING1	Serpin family G member 1	A protein encoding gene which is an inhibitor of C1 complex, fXIa, fXIIa, chymotrypsin and kallikrein [317].	• Upregulated in patients with active TB compared to latent controls [254] • Different expressed levels between QDT+ and QFT- tested persons [318] • Possible biomarker for progress from latent to active TB [110, 176, 181] • Possible biomarker for treatment	•Complement pathway •Immune response lectin induced complement pathway •Formation of fibrin clot (Clotting Cascade) •Cell adhesion •Plasmin signaling

			response/outcome prediction [118] • Upregulated in latent infected compared to healthy controls [255] • Possible diagnostic biomarker regardless of HIV status[177] • Plays as a C1 inhibitor a role in apoptotic cell clearance [319]	•Complement and coagulation cascades •Pertussis •Response to elevated platelet cytosolic Ca2+
STAT1	Signal transducer and activator of transcription 1	Protein coding gene that mediates intracellular response to interferons, cytokines and growth factors [320]	• Upregulated in patients with active TB compared to latent controls [254] • Possible biomarker for progress from latent to active TB [110, 176, 181] • Possible marker for therapy response [118] • Upregulated in latent infected compared to healthy controls [255] • Involved in macrophage apoptosis and phagocytosis after Infection with M. Tuberculosis [321] • JAK/STAT signaling cascade has an important influence on *M. tuberculosis* phagocytosis [322] • Influences the activation of those due to interferon-γ stimulation[323, 324] • STAT1 knock-out mice showed a more	•Immune response •Oncostatin M signaling via JAK-Stat in human cells •Peginterferon alpha-2a/peginterferon alpha-2b pathway (hepatocyte), pharmacodynamics •Development thrombopoetin signaling via JAK-STAT pathway •EGF receptor (ErbB1) signaling pathway •Interleukin-6 family signaling •FGFR1 mutant receptor activation •IL-7 signaling pathway •Type II interferon signaling (IFNG) •Type III interferon signaling pathways •Growth hormone receptor signaling •p38 MAPK signaling pathway

		severe and more frequent fatal course of disease than mice with existing STAT genes [325]	•Signaling events mediated by TCPTP •Jak/STAT signaling pathway intracellular regulation •IL-6 signaling pathway •Integrated cancer pathway •Interferon type I signaling pathways •FGF signaling pathway •PDGFR-beta signaling pathway •Interleukin-11 signaling pathway •TGF-beta receptor signaling •P38 MAPK signaling pathway •Signaling by FGFR2 in disease •Thymic stromal lympho-poietin (TSLP) signaling pathway •BRCA1 pathway •Development EPO-induced Jak-STAT pathway •AGE/RAGE pathway •Th1 differentiation pathway •Endochondral ossification •Immune response IFN gamma signaling pathway •Glucocorticoid receptor regulatory network

	•CXCR4-mediated signaling events •Human thyroid stimulating hormone (TSH) signaling pathway •IL-2 signaling pathway •TNF signaling •Bacterial infections in CF airways •IL-1 family signaling pathways •RANK signaling in osteoclasts •AGE-RAGE signaling pathway in diabetic complications •Kit receptor signaling pathway •Jak-Stat signaling pathway •ErbB1 downstream signaling •Immune response IFN alpha/beta signaling pathway •Development EGFR signaling via small GTPases •Toll comparative pathway •UVA-induced MAPK signaling •G-protein signaling Ras family GTPases in kinase cascades (scheme) •IL12-mediated signaling events

	•MAPK pathway •CNTF signaling •Development IGF-1 receptor signaling •Interleukin-4 and 13 signaling •Toxoplasmosis •Integrated breast cancer pathway •C-type lectin receptor signaling pathway •Osteoclast differentiation •Brain-derived neurotrophic factor (BDNF) signaling pathway •Type I interferon signaling pathways •Immune response IL-23 signaling pathway •Development ERBB-family signaling •Common cytokine receptor gamma-chain family signaling pathways •Prolactin signaling pathway •Thyroid hormone signaling pathway •Adipogenesis •Endometrial cancer •Necroptosis •Colorectal cancer metastasis

		•TCR signaling •Mesodermal commitment pathway •Tuberculosis •ErbB signaling pathway •EGF/EGFR signaling pathway •T-cell receptor signaling pathway •Th17 cell differentiation •Transport of the SLBP independent mature mRNA •NOD-like receptor signaling pathway •Interferon gamma signaling •JAK-STAT signaling pathway (KEGG) •4-1BB pathway •MAPK-Erk pathway •Allograft rejection •Regulation of lipid metabolism Insulin signaling-generic cascades •CCR5 pathway in macrophages •NF-kappaB signaling •Measles •Toll-like receptor signaling pathway •IL-2 pathway

				•Human cytomegalovirus infection •p70S6K signaling •Pathways in cancer •Herpes simplex virus 1 infection •Integrin pathway •PI3K-Akt signaling pathway •Akt signaling •GPCR pathway •TGF-beta pathway •HIV life cycle •Cytokine signaling in immune system •RET signaling •ERK signaling •Signaling by GPCR •Innate immune system
TNFRSF21	TNF Receptor Superfamily Member 21	Apoptosis promoting Protein coding gene [326]	• TNF superfamily members induce apoptosis in T-cells as immune response to infection [327]	• Apoptosis-related network due to altered Notch3 in ovarian cancer • Apoptosis and autophagy • Apoptosis modulation and signaling • TNFR1 pathway • TRAF pathway • Regulation of lipid metabolism by peroxisome

				proliferator-activated receptor alpha (PPARalpha) • Akt signaling • PEDF induced signaling • PAK pathway • ERK signaling • Metabolism
VAMP5	Vesicle-associated membrane protein 5	Protein coding gene that is involved in docking and fusion of cell membranes and vesicles [328]	• Upregulated in patients with active TB compared to latent controls [254]. • Possible diagnostic biomarker ability [176, 181] • Upregulated in latent infected compared to healthy controls [255] • Possible diagnostic biomarker regardless of HIV status[177] • VAMP5 knock out mice had insufficient lung development at birth [329] • Was found the bronchial epithelial cells predominantly [329] • Belongs to the soluble-N-ethylmaleimide-sensitive-factor accessory protein receptor proteins (SNARE) [329], which are essentially important for the immune system [330]	• Nicotine pathway (dopaminergic neuron) • Pharmacodynamics

	• SNARE proteins have an essential function for granulocytes, since SNARE are involved with different secretory granule [330]

Danksagung

An dieser Stelle möchte ich allen beteiligten Personen meinen großen Dank aussprechen, die mich bei der Anfertigung meiner Doktorarbeit unterstützt haben.

Das Forschungszentrum Borstel und vor allem die Forschungsgruppe der klinischen Infektiologie haben mir die Chance zum Verfassen einer Doktorarbeit gegeben, welche ohne die Unterstützung durch das Deutsche Zentrum für Infektionsforschung (DZIF) nicht möglich gewesen wäre. Mein besonderer Dank gilt Professor Christoph Lange und Professor Jan Heyckendorf für die hervorragende Betreuung und die enorme Unterstützung bei der Durchführung der gesamten Arbeit.

Darüber hinaus möchte ich Franziska Daduna und Jessica Hofmeister sowie allen technischen Assistentinnen und Assistenten danken, ohne deren Einsatz im Labor die hiesige Arbeit nicht möglich gewesen wäre. Weiterhin wäre die Arbeit nicht möglich ohne Dagmar Schaubs Datenmanagement. Neben der fachlichen und technischen Unterstützung möchte ich der Forschungsgruppe aber auch für die zwischenmenschliche Unterstützung danken, ohne die es deutlich schwerer gewesen wäre. Hierbei sind auch PD Dr. Barbara Kalsdorf sowie Dr. Patricia Sanchez-Caballo sowie die Doktorandinnen und Doktoranden Caroline Oertel, Johanna Eggeling, Niklas Köhler und Hans-Peter Grobbel zu nennen. Außerhalb der Forschungsgruppe gilt mein Dank Dr. Anna Störmer, die immer als Ansprechpartnerin für Doktorandinnen und Doktoranden zur Verfügung steht, sowie Professor Torsten Goldmann, welcher sich bereit erklärt hat, im Rahmen des Graduiertenprogramms als Zweitbetreuer zur Verfügung zu stehen. Auch der DZIF-Study Cohort Gruppe möchte ich meinen Dank aussprechen, sowie Dr. Andrew DiNardo, Professor Anna Mandalakas und Dr. Cristian Coarfa für ihre Unterstützung und den fachlichen Input.

Aber auch auf privater Ebene gab es so viele Leute, die mich all die Zeit unterstützt haben und ohne die ich vermutlich nicht so lange durchgehalten hätte. Neben meinem Vater allen voran meine Schwester Nele. Aber auch Jan, Max, Ferdi, Torsten, Laura, Christian, Anissa, Steffi, Kai und Carina haben als meine „längsten“ engen Freunde einen enormen, ihnen vermutlich gar nicht so bewussten Einfluss auf mein Leben genommen, ohne den ich vermutlich nie so weit gekommen wäre. Auch allen anderen, die hier nicht namentlich erwähnt wurden, mir jedoch in den letzten Jahren ans Herz gewachsen sind, gilt mein Dank.

Ohne das große Vertrauen und die enorme Unterstützung aller Beteiligten wäre diese Arbeit niemals möglich gewesen. Vielen Dank!

Einwilligungserklärungen und eidesstaatliche Versicherung

Hiermit erkläre ich, dass ich die vorliegende Arbeit eigenständig und ohne fremde Hilfe angefertigt habe. Alle technischen, personellen und sachlichen Hilfen und Hilfsmittel wurden im Text kenntlich gemacht, darüber hinaus wurden keine weiteren Hilfsmittel verwendet.

Textpassagen, die wörtlich oder dem Sinn nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche kenntlich gemacht. Mit einer Überprüfung unter Zuhilfenahme einer Antiplagiatssoftware erkläre ich mich einverstanden.

Die Dissertation wurde ausschließlich im Rahmen dieser Ausführung an der Universität zu Lübeck eingereicht. Andere Zulassungsanträge oder Einreichungen existieren nicht und wurden nie gestellt.

Dies ist mein erstes Promotionsverfahren.

Ich widerspreche nicht der Zulassung von Zuhörer:innen zur mündlichen Prüfung-

Borstel, 04. März 2022

Ort, Datum	Maja Reimann

Vollständige Publikationsliste

First authorship

Heyckendorf J*, Marwitz S*, Reimann M*, Avsar K, DiNardo AR, Günther G, Hoelscher M, Ibraim E, Kalsdorf B, Kaufmann SHE, Kontsevaya I, van Leth F, Mandalakas AM, Maurer FP, Müller M, Nitschkowski D, Olaru ID, Popa C, Rachow A, Rolling T, Rybniker J, Salzer HJF, Sanchez-Carballo P, Schuhmann M, Schaub D, Spinu V, Suárez I, Terhalle E, Unnewehr M, Weiner J, 3rd, Goldmann T*, Lange C*. Prediction of anti-tuberculosis treatment duration based on a 22-gene transcriptomic model. Eur Respir J 2021: 58(3). * authors contributed equally

Reimann M, Schaub D, Kalsdorf B, Runge C, Carballo PS, Terhalle E, Salzer HJF, Lange C, Heyckendorf J. Cigarette smoking and culture conversion in patients with susceptible and M/XDR-TB. Int J Tuberc Lung Dis 2019: 23(1): 93-98.

Co authorship

Heyckendorf J, Reimann M, Marwitz S, Lange C. Pathogen-free diagnosis of tuberculosis. Lancet Infect Dis 2021: 21(8): 1066.

Chesov E, Chesov D, Maurer FP, Andres S, Utpatel C, Barilar I, Donica A, Reimann M, Niemann S, Lange C, Crudu V, Heyckendorf J, Merker M. Emergence of bedaquiline-resistance in a high-burden country of tuberculosis. Eur Respir J 2021.

Salzer HJF, Massango I, Bhatt N, Machonisse E, Reimann M, Heldt S, Lange C, Hoelscher M, Khosa C, Rachow A. Seroprevalence of Aspergillus-Specific IgG Antibody among Mozambican Tuberculosis Patients. J Fungi (Basel) 2021: 7(8).

Brandenburg J, Marwitz S, Tazoll SC, Waldow F, Kalsdorf B, Vierbuchen T, Scholzen T, Gross A, Goldenbaum S, Hölscher A, Hein M, Linnemann L, Reimann M, Kispert A, Leitges M, Rupp J, Lange C, Niemann S, Behrends J, Goldmann T, Heine H, Schaible UE, Hölscher C, Schwudke D, Reiling N. WNT6/ACC2-induced storage of triacylglycerols in macrophages is exploited by Mycobacterium tuberculosis. J Clin Invest 2021: 131(16).

Chesov D, Heyckendorf J, Alexandru S, Donica A, Chesov E, Reimann M, Crudu V, Botnaru V, Lange C. Impact of bedaquiline on treatment outcomes of multidrug-resistant tuberculosis in a high-burden country. Eur Respir J 2021: 57(6).

Grbbel HP*, Merker M*, Köhler N*, Andres S*, Hoffmann H, Heyckendorf J, Reimann M, Barilar I, Dreyer V, Hillemann D, Kalsdorf B, Kohl TA, Sanchez Carballo P, Schaub D, Todt K, Utpatel C, Maurer

FP*, Lange C*, Niemann S*. Design of Multidrug-Resistant Tuberculosis Treatment Regimens Based on DNA Sequencing. Clin Infect Dis 2021: 73(7): 1194-1202. * authors contributed equally

Luethgen M, Eggeling J, Heyckendorf J, Lange C, Maier C, Reimann M, Sassmann-Schweda A, Shaikh R, Herzmann C. Changes in taste and smell as an early marker for COVID-19. Int J Infect Dis 2020: 99: 8-9.

Jünger C, Reimann M, Krabbe L, Gaede KI, Lange C, Herzmann C, Rüller S. Non-invasive ventilation with pursed lips breathing mode for patients with COPD and hypercapnic respiratory failure: A retrospective analysis. PLoS One 2020: 15(9): e0238619.

Salzer HJF, Prattes J, Flick H, Reimann M, Heyckendorf J, Kalsdorf B, Obersteiner S, Gaede KI, Herzmann C, Johnson GL, Lange C, Hoenigl M. Evaluation of Galactomannan Testing, the Aspergillus-Specific Lateral-Flow Device Test and Levels of Cytokines in Bronchoalveolar Lavage Fluid for Diagnosis of Chronic Pulmonary Aspergillosis. Front Microbiol 2018: 9: 2223.

Reviews

Kontsevaya I, Lange C, Comella-Del-Barrio P, Coarfa C, DiNardo AR, Gillespie SH, Hauptmann M, Leschczyk C, Mandalakas AM, Martinecz A, Merker M, Niemann S, Reimann M, Rzhepishevska O, Schaible UE, Scheu KM, Schurr E, Abel Zur Wiesch P, Heyckendorf J. Perspectives for systems biology in the management of tuberculosis. Eur Respir Rev 2021: 30(160).

Günther G, Heyckendorf J, Zellweger JP, Reimann M, Claassens M, Chesov D, van Leth F. Defining Outcomes of Tuberculosis (Treatment): From the Past to the Future. Respiration 2021: 100(9): 843-852.

Chesov D, Butov D, Reimann M, Heyckendorf J, Myasoedov V, Butov T, Akymenko O, Lange C. Impact of lung function on treatment outcome in patients with TB. Int J Tuberc Lung Dis 2021: 25(4): 277-284.

Günther G, Heyckendorf J, Zellweger JP, Reimann M, Claassens M, Chesov D, van Leth F. Defining Outcomes of Tuberculosis (Treatment): From the Past to the Future. Respiration 2021: 100(9): 843-852.

Lange C, Aarnoutse R, Chesov D, van Crevel R, Gillespie SH, Grobbel HP, Kalsdorf B, Kontsevaya I, van Laarhoven A, Nishiguchi T, Mandalakas A, Merker M, Niemann S, Köhler N, Heyckendorf J, Reimann M, Ruhwald M, Sanchez-Carballo P, Schwudke D, Waldow F, DiNardo AR. Perspective for Precision Medicine for Tuberculosis. Front Immunol 2020: 11: 566608.

Lange C, Aarnoutse RE, Alffenaar JWC, Bothamley G, Brinkmann F, Costa J, Chesov D, van Crevel R, Dedicoat M, Dominguez J, Duarte R, Grobbel HP, Günther G, Guglielmetti L, Heyckendorf J, Kay AW, Kirakosyan O, Kirk O, Koczulla RA, Kudriashov GG, Kuksa L, van Leth F, Magis-Escurra C, Mandalakas AM, Molina-Moya B, Peloquin CA, Reimann M, Rumetshofer R, Schaaf HS, Schön T, Tiberi S, Valda J, Yablonskii PK, Dheda K. Management of patients with multidrug-resistant tuberculosis. Int J Tuberc Lung Dis 2019: 23(6): 645-662.

Lange C, Alghamdi WA, Al-Shaer MH, Brighenti S, Diacon AH, DiNardo AR, Grobbel HP, Gröschel MI, von Groote-Bidlingmaier F, Hauptmann M, Heyckendorf J, Köhler N, Kohl TA, Merker M, Niemann S, Peloquin CA, Reimann M, Schaible UE, Schaub D, Schleusener V, Thye T, Schön T. Perspectives for personalized therapy for patients with multidrug-resistant tuberculosis. J Intern Med 2018. J Intern Med. 2018 May 28. doi: 10.1111/joim.12780.

Poster

Kontsevaja I, Lange C, Reimann M et al. PD-L1 expressing CD4+ and CD8+ T cells as a biomarker of tuberculosis disease and treatment response, ECCMID 2020 (virtual).

Chesiv D, Lange C, Reimann M et al. Treatment outcome in patients with MDR-TB treated with bedaquiline containing regimens, Republic of Moldova

Jünger C, Rüller S, Reimann M et al. Pursed-lip breathing ventilation for the treatment for hypercapnic respiratory faiure in COPD. ERS Respiratory Failure and Mechanical Ventilation Conference 2020

Grobbel H, Lange C, Reimann M et al. The sensitive, the specific and the gold standard – Advanced molecular and phenotypic DST diagnostics in M/XDR Tuberculosis. DZIF Jahrestagung 2019

Jünger, C, Herzmann C, Reimann M et al. Mechanical non-invasive pursed-lip breathing ventilation for hypercapnic patients with COPD. ERS International Congress 2019.

Reimann M et al. Effect of Smoke Cessation in TB Patients. NDI³ 2018

Presentations

Days of Ukraine in Berlin-Brandenburg (UKRAINET), FU Berlin. Ph.D. contest finalist: Titel: TB22 – A 22 gene model for personalizing therapy durations in patients with MDR-TB. 16.09.2021

DZIF Jahrestagung, Bad Nauheim. Titel: RNA based individualized therapy durations and monitoring in patients with tuberculosis, 23.11.2019

1. Cambau E, Drancourt M. Steps towards the discovery of Mycobacterium tuberculosis by Robert Koch, 1882. *Clin Microbiol Infect* 2014: 20(3): 196-201.
2. World Health Organization. WHO Guidelines Approved by the Guidelines Review Committee. WHO consolidated guidelines on tuberculosis: Module 4: Treatment - Drug-resistant tuberculosis treatment. World Health Organization

3. World Health Organization. Global Tuberculosis Report 2018. Geneva: WHO; 2018.
4. World Health Organization. Global Tuberculosis Report 2019. Geneva: WHO; 2019.
5. Centers for Disease Control. Nosocomial transmission of multidrug-resistant tuberculosis to health-care workers and HIV-infected patients in an urban hospital--Florida. *MMWR Morb Mortal Wkly Rep* 1990: 39(40): 718-722.
6. World Health Organization. Global Tuberculosis Report 2020. Geneva: WHO; 2020.
7. Knight GM, McQuaid CF, Dodd PJ, Houben R. Global burden of latent multidrug-resistant tuberculosis: trends and estimates based on mathematical modelling. *Lancet Infect Dis* 2019: 19(8): 903-912.
8. Ahmad N, Ahuja SD, Akkerman OW, Alffenaar JC, Anderson LF, Baghaei P, Bang D, Barry PM, Bastos ML, Behera D, Benedetti A, Bisson GP, Boeree MJ, Bonnet M, Brode SK, Brust JCM, Cai Y, Caumes E, Cegielski JP, Centis R, Chan PC, Chan ED, Chang KC, Charles M, Cirule A, Dalcolmo MP, D'Ambrosio L, de Vries G, Dheda K, Esmail A, Flood J, Fox GJ, Fréchet-Jachym M, Fregona G, Gayoso R, Gegia M, Gler MT, Gu S, Guglielmetti L, Holtz TH, Hughes J, Isaakidis P, Jarlsberg L, Kempker RR, Keshavjee S, Khan FA, Kipiani M, Koenig SP, Koh WJ, Kritski A, Kuksa L, Kvasnovsky CL, Kwak N, Lan Z, Lange C, Laniado-Laborín R, Lee M, Leimane V, Leung CC, Leung EC, Li PZ, Lowenthal P, Maciel EL, Marks SM, Mase S, Mbuagbaw L, Migliori GB, Milanov V, Miller AC, Mitnick CD, Modongo C, Mohr E, Monedero I, Nahid P, Ndjeka N, O'Donnell MR, Padayatchi N, Palmero D, Pape JW, Podewils LJ, Reynolds I, Riekstina V, Robert J, Rodriguez M, Seaworth B, Seung KJ, Schnippel K, Shim TS, Singla R, Smith SE, Sotgiu G, Sukhbaatar G, Tabarsi P, Tiberi S, Trajman A, Trieu L, Udwadia ZF, van der Werf TS, Veziris N, Viiklepp P, Vilbrun SC, Walsh K, Westenhouse J, Yew WW, Yim JJ, Zetola NM, Zignol M, Menzies D. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. *Lancet* 2018: 392(10150): 821-834.
9. Conradie F, Diacon AH, Ngubane N, Howell P, Everitt D, Crook AM, Mendel CM, Egizi E, Moreira J, Timm J, McHugh TD, Wills GH, Bateson A, Hunt R, Van Niekerk C, Li M, Olugbosi M, Spigelman M. Treatment of Highly Drug-Resistant Pulmonary Tuberculosis. *N Engl J Med* 2020: 382(10): 893-902.
10. Burki T. BPaL approved for multidrug-resistant tuberculosis. *Lancet Infect Dis* 2019: 19(10): 1063-1064.
11. Abidi S, Achar J, Assao Neino MM, Bang D, Benedetti A, Brode S, Campbell JR, Casas EC, Conradie F, Dravniece G, du Cros P, Falzon D, Jaramillo E, Kuaban C, Lan Z, Lange C, Li PZ, Makhmudova M, Maug AKJ, Menzies D, Migliori GB, Miller A, Myrzaliev B, Ndjeka N, Noeske J, Parpieva N, Piubello A, Schwoebel V, Sikhondze W, Singla R, Souleymane MB, Trébucq A, Van Deun A, Viney K, Weyer K, Zhang BJ, Ahmad Khan F. Standardised shorter regimens versus individualised longer regimens for rifampin- or multidrug-resistant tuberculosis. *Eur Respir J* 2020: 55(3).
12. Wahid A, Ahmad N, Ghafoor A, Latif A, Saleem F, Khan S, Atif M, Iqbal Q. Effectiveness of Shorter Treatment Regimen in Multidrug-Resistant Tuberculosis Patients in Pakistan: A Multicenter Retrospective Record Review. *Am J Trop Med Hyg* 2021: 104(5): 1784-1791.
13. Anh LTN, A MVK, Ramaswamy G, Htun T, Thanh Hoang Thi T, Hoai Nguyen G, Quelapio M, Gebhard A, Nguyen HB, Nguyen NV. High Levels of Treatment Success and Zero Relapse in Multidrug-Resistant Tuberculosis Patients Receiving a Levofloxacin-Based Shorter Treatment Regimen in Vietnam. *Trop Med Infect Dis* 2020: 5(1).

14. Günther G, Gomez GB, Lange C, Rupert S, van Leth F. Availability, price and affordability of anti-tuberculosis drugs in Europe: a TBNET survey. *Eur Respir J* 2015: 45(4): 1081-1088.
15. Lan Z, Ahmad N, Baghaei P, Barkane L, Benedetti A, Brode SK, Brust JCM, Campbell JR, Chang VWL, Falzon D, Guglielmetti L, Isaakidis P, Kempker RR, Kipiani M, Kuksa L, Lange C, Laniado-Laborín R, Nahid P, Rodrigues D, Singla R, Udwadia ZF, Menzies D. Drug-associated adverse events in the treatment of multidrug-resistant tuberculosis: an individual patient data meta-analysis. *Lancet Respir Med* 2020: 8(4): 383-394.
16. Lange C, Alghamdi WA, Al-Shaer MH, Brighenti S, Diacon AH, DiNardo AR, Grobbel HP, Gröschel MI, von Groote-Bidlingmaier F, Hauptmann M, Heyckendorf J, Köhler N, Kohl TA, Merker M, Niemann S, Peloquin CA, Reimann M, Schaible UE, Schaub D, Schleusener V, Thye T, Schön T. Perspectives for personalized therapy for patients with multidrug-resistant tuberculosis. *J Intern Med* 2018.
17. Pontali E, Raviglione MC, Migliori GB. Regimens to treat multidrug-resistant tuberculosis: past, present and future perspectives. *Eur Respir Rev* 2019: 28(152).
18. Gröschel MI, Walker TM, van der Werf TS, Lange C, Niemann S, Merker M. Pathogen-based precision medicine for drug-resistant tuberculosis. *PLoS Pathog* 2018: 14(10): e1007297.
19. Heyckendorf J, Olaru ID, Ruhwald M, Lange C. Getting personal perspectives on individualized treatment duration in multidrug-resistant and extensively drug-resistant tuberculosis. *Am J Respir Crit Care Med* 2014: 190(4): 374-383.
20. Wanner R, Springer B. Die Tuberkulose, vom Altertum in die Neuzeit. *Bulletin des médecins suisses* 2008: 89: 37.
21. Bange F, Hahn H, Kaufmann S, Lange C, Ulrichs T. Mykobakterien. Medizinische Mikrobiologie und Infektiologie. Springer, Berlin, Heidelberg, 2020; pp. 447-465.
22. Koch R. Die Ätiologie der Tuberkulose (Nach, einem in der Physiologischen Gesellschaft zu Berlin am 24. März 1882 gehaltenen Vortrage). Physiologische Gesellschaft zu Berlin, 1882; pp. 428-445.
23. Migliori GB, Eker B, Richardson MD, Sotgiu G, Zellweger JP, Skrahina A, Ortmann J, Girardi E, Hoffmann H, Besozzi G, Bevilacqua N, Kirsten D, Centis R, Lange C. A retrospective TBNET assessment of linezolid safety, tolerability and efficacy in multidrug-resistant tuberculosis. *Eur Respir J* 2009: 34(2): 387-393.
24. Djelouadji Z, Raoult D, Daffé M, Drancourt M. A single-step sequencing method for the identification of Mycobacterium tuberculosis complex species. *PLoS Negl Trop Dis* 2008: 2(6): e253.
25. Werner H. Mycobacterium. *In:* Werner H, ed. Medizinische Mikrobiologie mit Repetitorium. Walter de Gruyter GmbH & Co KG, Berlin, 2020; pp. 240 - 248.
26. Wirth T, Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T, Kremer K, van Soolingen D, Rüsch-Gerdes S, Locht C, Brisse S, Meyer A, Supply P, Niemann S. Origin, spread and demography of the Mycobacterium tuberculosis complex. *PLoS Pathog* 2008: 4(9): e1000160.
27. Austermann P. Tuberkulose-eine wieder aktuelle Lungenkrankheit. GRIN Verlag, 2005.
28. Houben RM, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLoS Med* 2016: 13(10): e1002152.
29. Mack U, Migliori GB, Sester M, Rieder HL, Ehlers S, Goletti D, Bossink A, Magdorf K, Hölscher C, Kampmann B, Arend SM, Detjen A, Bothamley G, Zellweger JP, Milburn H, Diel R, Ravn P, Cobelens F, Cardona PJ, Kan B, Solovic I, Duarte R, Cirillo DM. LTBI: latent tuberculosis infection or lasting immune responses to M. tuberculosis? A TBNET consensus statement. *Eur Respir J* 2009: 33(5): 956-973.
30. European Centre for Disease Prevention and Control WHO Regional Office for Europe. Tuberculosis surveillance and monitoring in Europe 2021 –2019 data. Copenhagen: WHO; 2021 22.03.2021.
31. Suárez I, Fünger SM, Kröger S, Rademacher J, Fätkenheuer G, Rybniker J. The Diagnosis and Treatment of Tuberculosis. *Dtsch Arztebl Int* 2019: 116(43): 729-735.
32. Bauer T, Häcker B, Otto-Knapp R, Trost C. Welttuberkulosetag 2021: Infektionskontrolle gestern und heute–125 Jahre Deutsches Zentralkomitee zur Bekämpfung der Tuberkulose (DZK). *Epid Bull* 2021: 11: 3-5.

33. Robert Loch-Institut. Aktuelle Daten und Informationen zu Infektionskrankheiten und Public Health. *Epid Bull* 2020: 11: 3-22.
34. Heye T, Stoijkovic M, Kauczor HU, Junghanss T, Hosch W. [Extrapulmonary tuberculosis: radiological imaging of an almost forgotten transformation artist]. *Rofo* 2011: 183(11): 1019-1029.
35. Brodhun B, Altmann D, Hauer B, Fiebig L, Haas W. Bericht zur Epidemiologie der Tuberkulose in Deutschland für 2015. Robert Koch-Institut, 2016.
36. Eum SY, Kong JH, Hong MS, Lee YJ, Kim JH, Hwang SH, Cho SN, Via LE, Barry CE, 3rd. Neutrophils are the predominant infected phagocytic cells in the airways of patients with active pulmonary TB. *Chest* 2010: 137(1): 122-128.
37. Ziegler R, Just HM, Castell S, Diel R, Gastmeier P, Haas W, Hauer B, Loytved G, Mielke M, Moser I, Nienhaus A, Richter E, Rüden H, Rüsch-Gerdes S, Schaberg T, Wischnewski N, Loddenkemper R. [Tuberculosis infection control--recommendations of the DZK]. *Pneumologie* 2012: 66(5): 269-282.
38. World Health Organization. Guidelines on the management of latent tuberculosis infection. World Health Organization, Geneva, 2015.
39. Gualano G, Mencarini P, Lauria FN, Palmieri F, Mfinanga S, Mwaba P, Chakaya J, Zumla A, Ippolito G. Tuberculin skin test - Outdated or still useful for Latent TB infection screening? *Int J Infect Dis* 2019: 80s: S20-s22.
40. Boom WH, Schaible UE, Achkar JM. The knowns and unknowns of latent Mycobacterium tuberculosis infection. *J Clin Invest* 2021: 131(3).
41. Drain PK, Bajema KL, Dowdy D, Dheda K, Naidoo K, Schumacher SG, Ma S, Meermeier E, Lewinsohn DM, Sherman DR. Incipient and Subclinical Tuberculosis: a Clinical Review of Early Stages and Progression of Infection. *Clin Microbiol Rev* 2018: 31(4).
42. Caws M, Marais B, Heemskerk D, Farrar J. Clinical Manifestations. Tuberculosis in Adults and Children. Springer Nature, 2015; p. 66.
43. Fuentes Ferrer M, Gutiérrez Torres L, Ayala Ramírez O, Rumayor Zarzuelo M, del Prado González N. Tuberculosis of the spine. A systematic review of case series. *Int Orthop* 2012: 36(2): 221-231.
44. Rock RB, Olin M, Baker CA, Molitor TW, Peterson PK. Central nervous system tuberculosis: pathogenesis and clinical aspects. *Clin Microbiol Rev* 2008: 21(2): 243-261, table of contents.
45. Ntsekhe M, Mayosi BM. Tuberculous pericarditis with and without HIV. *Heart Fail Rev* 2013: 18(3): 367-373.
46. Ryu YJ. Diagnosis of pulmonary tuberculosis: recent advances and diagnostic algorithms. *Tuberc Respir Dis (Seoul)* 2015: 78(2): 64-71.
47. Walzl G, McNerney R, du Plessis N, Bates M, McHugh TD, Chegou NN, Zumla A. Tuberculosis: advances and challenges in development of new diagnostics and biomarkers. *Lancet Infect Dis* 2018: 18(7): e199-e210.
48. Shapiro AE, Ross JM, Yao M, Schiller I, Kohli M, Dendukuri N, Steingart KR, Horne DJ. Xpert MTB/RIF and Xpert Ultra assays for screening for pulmonary tuberculosis and rifampicin resistance in adults, irrespective of signs or symptoms. *Cochrane Database Syst Rev* 2021: 3(3): Cd013694.
49. Lee JJ, Chong PY, Lin CB, Hsu AH, Lee CC. High resolution chest CT in patients with pulmonary tuberculosis: characteristic findings before and after antituberculous therapy. *Eur J Radiol* 2008: 67(1): 100-104.
50. Nachiappan AC, Rahbar K, Shi X, Guy ES, Mortani Barbosa EJ, Jr., Shroff GS, Ocazionez D, Schlesinger AE, Katz SI, Hammer MM. Pulmonary Tuberculosis: Role of Radiology in Diagnosis and Management. *Radiographics* 2017: 37(1): 52-72.
51. Bhalla AS, Goyal A, Guleria R, Gupta AK. Chest tuberculosis: Radiological review and imaging recommendations. *Indian J Radiol Imaging* 2015: 25(3): 213-225.
52. Bomanji J, Sharma R, Mittal BR, Gambhir S, Qureshy A, Begum SMF, Paez D, Sathekge M, Vorster M, Sobic Saranovic D, Pusuwan P, Mann V, Vinjamuri S, Zumla A, Pascual TNB. PET/CT

features of extrapulmonary tuberculosis at first clinical presentation: a cross-sectional observational 18 F-FDG imaging study across six countries. *Eur Respir J* 2020: 55(2).
53. Goletti D, Lee MR, Wang JY, Walter N, Ottenhoff THM. Update on tuberculosis biomarkers: From correlates of risk, to correlates of active disease and of cure from disease. *Respirology* 2018: 23(5): 455-466.
54. Grobbel HP, Merker M, Köhler N, Andres S, Hoffmann H, Heyckendorf J, Reimann M, Barilar I, Dreyer V, Hillemann D, Kalsdorf B, Kohl TA, Sanchez Carballo P, Schaub D, Todt K, Utpatel C, Maurer FP, Lange C, Niemann S. Design of Multidrug-Resistant Tuberculosis Treatment Regimens Based on DNA Sequencing. *Clin Infect Dis* 2021: 73(7): 1194-1202.
55. Schön T, Miotto P, Köser CU, Viveiros M, Böttger E, Cambau E. Mycobacterium tuberculosis drug-resistance testing: challenges, recent developments and perspectives. *Clin Microbiol Infect* 2017: 23(3): 154-160.
56. Bandodkar B, Shandil R, Bhat J, Balganesh T. Two Decades of TB Drug Discovery Efforts—What Have We Learned? *Applied Sciences* 2020: 10(16): 5704.
57. Barry CE. Lessons from seven decades of antituberculosis drug discovery. *Curr Top Med Chem* 2011: 11(10): 1216-1225.
58. Zumla A, Nahid P, Cole ST. Advances in the development of new tuberculosis drugs and treatment regimens. *Nat Rev Drug Discov* 2013: 12(5): 388-404.
59. Schaberg T, Bauer T, Brinkmann F, Diel R, Feiterna-Sperling C, Haas W, Hartmann P, Hauer B, Heyckendorf J, Lange C, Nienhaus A, Otto-Knapp R, Priwitzer M, Richter E, Rumetshofer R, Schenkel K, Schoch OD, Schönfeld N, Stahlmann R. [Tuberculosis Guideline for Adults - Guideline for Diagnosis and Treatment of Tuberculosis including LTBI Testing and Treatment of the German Central Committee (DZK) and the German Respiratory Society (DGP)]. *Pneumologie* 2017: 71(6): 325-397.
60. Grobbelaar M, Louw GE, Sampson SL, van Helden PD, Donald PR, Warren RM. Evolution of rifampicin treatment for tuberculosis. *Infect Genet Evol* 2019: 74: 103937.
61. Hedberg-Oldfors C, Mensch A, Visuttijai K, Stoltenburg G, Stoevesandt D, Kraya T, Oldfors A, Zierz S. Polyglucosan myopathy and functional characterization of a novel GYG1 mutation. *Acta Neurol Scand* 2018: 137(3): 308-315.
62. Jain K, Desai M, Solanki R, Dikshit RK. Treatment outcome of standardized regimen in patients with multidrug resistant tuberculosis. *J Pharmacol Pharmacother* 2014: 5(2): 145-149.
63. Imperial MZ, Nahid P, Phillips PPJ, Davies GR, Fielding K, Hanna D, Hermann D, Wallis RS, Johnson JL, Lienhardt C, Savic RM. A patient-level pooled analysis of treatment-shortening regimens for drug-susceptible pulmonary tuberculosis. *Nat Med* 2018: 24(11): 1708-1715.
64. Lange C, Aarnoutse R, Chesov D, van Crevel R, Gillespie SH, Grobbel HP, Kalsdorf B, Kontsevaya I, van Laarhoven A, Nishiguchi T, Mandalakas A, Merker M, Niemann S, Köhler N, Heyckendorf J, Reimann M, Ruhwald M, Sanchez-Carballo P, Schwudke D, Waldow F, DiNardo AR. Perspective for Precision Medicine for Tuberculosis. *Front Immunol* 2020: 11: 566608.
65. Nora D, Salluh J, Martin-Loeches I, Póvoa P. Biomarker-guided antibiotic therapy-strengths and limitations. *Ann Transl Med* 2017: 5(10): 208.
66. Califf RM. Biomarker definitions and their applications. *Exp Biol Med (Maywood)* 2018: 243(3): 213-221.
67. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982: 143(1): 29-36.
68. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol* 2008: 56(1): 45-50.
69. Alene KA, Viney K, Yi H, McBryde ES, Yang K, Bai L, Gray DJ, Xu Z, Clements ACA. Comparison of the validity of smear and culture conversion as a prognostic marker of treatment outcome in patients with multidrug-resistant tuberculosis. *PLoS One* 2018: 13(5): e0197880.
70. Olaru ID, Heyckendorf J, Grossmann S, Lange C. Time to culture positivity and sputum smear microscopy during tuberculosis therapy. *PLoS One* 2014: 9(8): e106075.
71. Sabiiti W, Azam K, Farmer ECW, Kuchaka D, Mtafya B, Bowness R, Oravcova K, Honeyborne I, Evangelopoulos D, McHugh TD, Khosa C, Rachow A, Heinrich N, Kampira E, Davies G, Bhatt N, Ntinginya EN, Viegas S, Jani I, Kamdolozi M, Mdolo A, Khonga M, Boeree MJ, Phillips

PPJ, Sloan D, Hoelscher M, Kibiki G, Gillespie SH. Tuberculosis bacillary load, an early marker of disease severity: the utility of tuberculosis Molecular Bacterial Load Assay. *Thorax* 2020: 75(7): 606-608.
72. Martin C, Castaigne C, Vierasu I, Garcia C, Wyndham-Thomas C, de Wit S. Prospective Serial FDG PET/CT During Treatment of Extrapulmonary Tuberculosis in HIV-Infected Patients: An Exploratory Study. *Clin Nucl Med* 2018: 43(9): 635-640.
73. Esmail H, Lai RP, Lesosky M, Wilkinson KA, Graham CM, Coussens AK, Oni T, Warwick JM, Said-Hartley Q, Koegelenberg CF, Walzl G, Flynn JL, Young DB, Barry Iii CE, O'Garra A, Wilkinson RJ. Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-18F fluoro-D-glucose positron emission and computed tomography. *Nat Med* 2016: 22(10): 1090-1093.
74. Heo EY, Chun EJ, Lee CH, Kim YW, Han SK, Shim YS, Lee HJ, Yim JJ. Radiographic improvement and its predictors in patients with pulmonary tuberculosis. *Int J Infect Dis* 2009: 13(6): e371-376.
75. Labugger I, Heyckendorf J, Dees S, Häussinger E, Herzmann C, Kohl TA, Richter E, Rivera-Milla E, Lange C. Detection of transrenal DNA for the diagnosis of pulmonary tuberculosis and treatment monitoring. *Infection* 2017: 45(3): 269-276.
76. Heyckendorf J, Labugger I, van der Merwe L, Garcia-Basteiro AL, Diacon AH, Lange C. Serial measurements of transrenal mycobacterial DNA as indicators of the early bactericidal activity (EBA) of antituberculosis drugs. *Tuberculosis (Edinb)* 2017: 102: 31-33.
77. Guérardel Y, Maes E, Briken V, Chirat F, Leroy Y, Locht C, Strecker G, Kremer L. Lipomannan and lipoarabinomannan from a clinical isolate of Mycobacterium kansasii: novel structural features and apoptosis-inducing properties. *J Biol Chem* 2003: 278(38): 36637-36651.
78. Bulterys MA, Wagner B, Redard-Jacot M, Suresh A, Pollock NR, Moreau E, Denkinger CM, Drain PK, Broger T. Point-Of-Care Urine LAM Tests for Tuberculosis Diagnosis: A Status Update. *J Clin Med* 2019: 9(1).
79. MacLean E, Broger T, Yerlikaya S, Fernandez-Carballo BL, Pai M, Denkinger CM. A systematic review of biomarkers to detect active tuberculosis. *Nat Microbiol* 2019: 4(5): 748-758.
80. Kawasaki M, Echiverri C, Raymond L, Cadena E, Reside E, Gler MT, Oda T, Ito R, Higashiyama R, Katsuragi K, Liu Y. Lipoarabinomannan in sputum to detect bacterial load and treatment response in patients with pulmonary tuberculosis: Analytic validation and evaluation in two cohorts. *PLoS Med* 2019: 16(4): e1002780.
81. Rudolf F, Lemvik G, Abate E, Verkuilen J, Schön T, Gomes VF, Eugen-Olsen J, Østergaard L, Wejse C. TBscore II: refining and validating a simple clinical score for treatment monitoring of patients with pulmonary tuberculosis. *Scand J Infect Dis* 2013: 45(11): 825-836.
82. Heyckendorf J, van Leth F, Avsar K, Glattki G, Günther G, Kalsdorf B, Müller M, Olaru ID, Rolling T, Salzer HJF, Schuhmann M, Terhalle E, Lange C. Treatment responses in multidrug-resistant tuberculosis in Germany. *Int J Tuberc Lung Dis* 2018: 22(4): 399-406.
83. Santos VS, Goletti D, Kontogianni K, Adams ER, Molina-Moya B, Dominguez J, Crudu V, Martins-Filho PRS, Ruhwald M, Lawson L, Bimba JS, Garcia-Basteiro AL, Petrone L, Kabeer BS, Reither K, Cuevas LE. Acute phase proteins and IP-10 as triage tests for the diagnosis of tuberculosis: systematic review and meta-analysis. *Clin Microbiol Infect* 2019: 25(2): 169-177.
84. Wilson D, Moosa MS, Cohen T, Cudahy P, Aldous C, Maartens G. Evaluation of Tuberculosis Treatment Response With Serial C-Reactive Protein Measurements. *Open Forum Infect Dis* 2018: 5(11): ofy253.
85. Nikolayevskyy V, Balabanova Y, Kontsevaya I, Ignatyeva O, Skenders G, Vasiliauskiene E, Bockel DV, Drobniewski F. Biomarkers of treatment success in fully sensitive pulmonary tuberculosis patients: a multicenter longitudinal study. *Biomark Med* 2020: 14(15): 1439-1452.
86. Kumar NP, Moideen K, Banurekha VV, Nair D, Babu S. Plasma Proinflammatory Cytokines Are Markers of Disease Severity and Bacterial Burden in Pulmonary Tuberculosis. *Open Forum Infect Dis* 2019: 6(7): ofz257.
87. Ambreen A, Khaliq A, Naqvi SZH, Tahir A, Mustafa M, Chaudhary SU, Mirza S, Mustafa T. Host biomarkers for monitoring therapeutic response in extrapulmonary tuberculosis. *Cytokine* 2021: 142: 155499.

88. Kumar NP, Moideen K, Nancy A, Viswanathan V, Shruthi BS, Sivakumar S, Natarajan M, Kornfeld H, Babu S. Plasma chemokines are biomarkers of disease severity, higher bacterial burden and delayed sputum culture conversion in pulmonary tuberculosis. *Sci Rep* 2019: 9(1): 18217.
89. Ahmed M. Exploring sputum independent host biomarkers for detection of Tuberculosis disease and monitoring treatment. München: Ludwig-Maximilian Univesität München; 2019.
90. Liu C, Zhao Z, Fan J, Lyon CJ, Wu HJ, Nedelkov D, Zelazny AM, Olivier KN, Cazares LH, Holland SM, Graviss EA, Hu Y. Quantification of circulating Mycobacterium tuberculosis antigen peptides allows rapid diagnosis of active disease and treatment monitoring. *Proc Natl Acad Sci U S A* 2017: 114(15): 3969-3974.
91. Kaewseekhao B, Roytrakul S, Yingchutrakul Y, Salao K, Reechaipichitkul W, Faksri K. Proteomic analysis of infected primary human leucocytes revealed PSTK as potential treatment-monitoring marker for active and latent tuberculosis. *PLoS One* 2020: 15(4): e0231834.
92. Nahid P, Dorman SE, Alipanah N, Barry PM, Brozek JL, Cattamanchi A, Chaisson LH, Chaisson RE, Daley CL, Grzemska M, Higashi JM, Ho CS, Hopewell PC, Keshavjee SA, Lienhardt C, Menzies R, Merrifield C, Narita M, O'Brien R, Peloquin CA, Raftery A, Saukkonen J, Schaaf HS, Sotgiu G, Starke JR, Migliori GB, Vernon A. Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. *Clin Infect Dis* 2016: 63(7): e147-e195.
93. Fitzgerald BL, Islam MN, Graham B, Mahapatra S, Webb K, Boom WH, Malherbe ST, Joloba ML, Johnson JL, Winter J, Walzl G, Belisle JT. Elucidation of a Human Urine Metabolite as a Seryl-Leucine Glycopeptide and as a Biomarker of Effective Anti-Tuberculosis Therapy. *ACS Infect Dis* 2019: 5(3): 353-364.
94. Combrink M, du Preez I, Ronacher K, Walzl G, Loots DT. Time-Dependent Changes in Urinary Metabolome Before and After Intensive Phase Tuberculosis Therapy: A Pharmacometabolomics Study. *Omics* 2019: 23(11): 560-572.
95. Tientcheu LD, Maertzdorf J, Weiner J, Adetifa IM, Mollenkopf HJ, Sutherland JS, Donkor S, Kampmann B, Kaufmann SH, Dockrell HM, Ota MO. Differential transcriptomic and metabolic profiles of M. africanum- and M. tuberculosis-infected patients after, but not before, drug treatment. *Genes Immun* 2015: 16(5): 347-355.
96. du Preez I, Luies L, Loots DT. The application of metabolomics toward pulmonary tuberculosis research. *Tuberculosis (Edinb)* 2019: 115: 126-139.
97. Casamassimi A, Federico A, Rienzo M, Esposito S, Ciccodicola A. Transcriptome Profiling in Human Diseases: New Advances and Perspectives. *Int J Mol Sci* 2017: 18(8).
98. Brix J, Koch H-G, Heinrich PC. Transkription und Prozessierung der RNA. Löffler/Petrides Biochemie und Pathobiochemie. Springer, 2014; pp. 567-587.
99. Lekka E, Hall J. Noncoding RNAs in disease. *FEBS Lett* 2018: 592(17): 2884-2900.
100. Bao Z, Yang Z, Huang Z, Zhou Y, Cui Q, Dong D. LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res* 2019: 47(D1): D1034-d1037.
101. Xi X, Li T, Huang Y, Sun J, Zhu Y, Yang Y, Lu ZJ. RNA Biomarkers: Frontier of Precision Medicine for Cancer. *Noncoding RNA* 2017: 3(1).
102. Mantione KJ, Kream RM, Kuzelova H, Ptacek R, Raboch J, Samuel JM, Stefano GB. Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq. *Med Sci Monit Basic Res* 2014: 20: 138-142.
103. Vlahou A, Magni F, Mischak H, Zoidakis J. Integration of omics approaches and systems biology for clinical applications. John Wiley & Sons, 2018.
104. Dong Z, Chen Y. Transcriptomics: advances and approaches. *Sci China Life Sci* 2013: 56(10): 960-967.
105. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015: 43(7): e47.
106. Quinn TP, Erb I, Gloor G, Notredame C, Richardson MF, Crowley TM. A field guide for the compositional analysis of any-omics data. *Gigascience* 2019: 8(9).

107. Honeyborne I, McHugh TD, Phillips PP, Bannoo S, Bateson A, Carroll N, Perrin FM, Ronacher K, Wright L, van Helden PD, Walzl G, Gillespie SH. Molecular bacterial load assay, a culture-free biomarker for rapid and accurate quantification of sputum Mycobacterium tuberculosis bacillary load during treatment. *J Clin Microbiol* 2011: 49(11): 3905-3911.
108. Honeyborne I, Mtafya B, Phillips PP, Hoelscher M, Ntinginya EN, Kohlenberg A, Rachow A, Rojas-Ponce G, McHugh TD, Heinrich N. The molecular bacterial load assay replaces solid culture for measuring early bactericidal response to antituberculosis treatment. *J Clin Microbiol* 2014: 52(8): 3064-3067.
109. Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. *PLoS Med* 2019: 16(4): e1002786.
110. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, Mahomed H, Erasmus M, Whatney W, Hussey GD, Abrahams D, Kafaar F, Hawkridge T, Verver S, Hughes EJ, Ota M, Sutherland J, Howe R, Dockrell HM, Boom WH, Thiel B, Ottenhoff THM, Mayanja-Kizza H, Crampin AC, Downing K, Hatherill M, Valvo J, Shankar S, Parida SK, Kaufmann SHE, Walzl G, Aderem A, Hanekom WA. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* 2016: 387(10035): 2312-2322.
111. Darboe F, Mbandi SK, Naidoo K, Yende-Zuma N, Lewis L, Thompson EG, Duffy FJ, Fisher M, Filander E, van Rooyen M, Bilek N, Mabwe S, McKinnon LR, Chegou N, Loxton A, Walzl G, Tromp G, Padayatchi N, Govender D, Hatherill M, Karim SA, Zak DE, Penn-Nicholson A, Scriba TJ. Detection of Tuberculosis Recurrence, Diagnosis and Treatment Response by a Blood Transcriptomic Risk Signature in HIV-Infected Persons on Antiretroviral Therapy. *Front Microbiol* 2019: 10: 1441.
112. Turner CT, Gupta RK, Tsaliki E, Roe JK, Mondal P, Nyawo GR, Palmer Z, Miller RF, Reeve BW, Theron G, Noursadeghi M. Blood transcriptional biomarkers for active pulmonary tuberculosis in a high-burden setting: a prospective, observational, diagnostic accuracy study. *Lancet Respir Med* 2020: 8(4): 407-419.
113. Warsinske HC, Rao AM, Moreira FMF, Santos PCP, Liu AB, Scott M, Malherbe ST, Ronacher K, Walzl G, Winter J, Sweeney TE, Croda J, Andrews JR, Khatri P. Assessment of Validity of a Blood-Based 3-Gene Signature Score for Progression and Diagnosis of Tuberculosis, Disease Severity, and Treatment Response. *JAMA Netw Open* 2018: 1(6): e183779.
114. Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir Med* 2016: 4(3): 213-224.
115. Ho J, Bokil NJ, Nguyen PTB, Nguyen TA, Liu MY, Hare N, Fox GJ, Saunders BM, Marks GB, Britton WJ. A transcriptional blood signature distinguishes early tuberculosis disease from latent tuberculosis infection and uninfected individuals in a Vietnamese cohort. *J Infect* 2020: 81(1): 72-80.
116. Hoang LT, Jain P, Pillay TD, Tolosa-Wright M, Niazi U, Takwoingi Y, Halliday A, Berrocal-Almanza LC, Deeks JJ, Beverley P, Kon OM, Lalvani A. Transcriptomic signatures for diagnosing tuberculosis in clinical practice: a prospective, multicentre cohort study. *Lancet Infect Dis* 2021: 21(3): 366-375.
117. Penn-Nicholson A, Mbandi SK, Thompson E, Mendelsohn SC, Suliman S, Chegou NN, Malherbe ST, Darboe F, Erasmus M, Hanekom WA, Bilek N, Fisher M, Kaufmann SHE, Winter J, Murphy M, Wood R, Morrow C, Van Rhijn I, Moody B, Murray M, Andrade BB, Sterling TR, Sutherland J, Naidoo K, Padayatchi N, Walzl G, Hatherill M, Zak D, Scriba TJ. RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response. *Sci Rep* 2020: 10(1): 8629.
118. Thompson EG, Du Y, Malherbe ST, Shankar S, Braun J, Valvo J, Ronacher K, Tromp G, Tabb DL, Alland D, Shenai S, Via LE, Warwick J, Aderem A, Scriba TJ, Winter J, Walzl G, Zak DE. Host blood RNA signatures predict the outcome of tuberculosis treatment. *Tuberculosis (Edinb)* 2017: 107: 48-58.
119. Sivakumaran D, Jenum S, Vaz M, Selvam S, Ottenhoff THM, Haks MC, Malherbe ST, Doherty TM, Ritz C, Grewal HMS. Combining host-derived biomarkers with patient characteristics improves signature performance in predicting tuberculosis treatment outcomes. *Commun Biol* 2020: 3(1): 359.

120. Ronacher K, Chegou NN, Kleynhans L, Djoba Siawaya JF, du Plessis N, Loxton AG, Maasdorp E, Tromp G, Kidd M, Stanley K, Kriel M, Menezes A, Gutschmidt A, van der Spuy GD, Warren RM, Dietze R, Okwera A, Thiel B, Belisle JT, Cliff JM, Boom WH, Johnson JL, van Helden PD, Dockrell HM, Walzl G. Distinct serum biosignatures are associated with different tuberculosis treatment outcomes. *Tuberculosis (Edinb)* 2019: 118: 101859.
121. Sepulveda JL. Using R and Bioconductor in Clinical Genomics and Transcriptomics. *J Mol Diagn* 2020: 22(1): 3-20.
122. Friedman J. The elements of statistical learning: Data mining, inference, and prediction. Springer open, 2017.
123. Gentleman R. R programming for bioinformatics. CRC Press Taylor & Francis Group, Seattle, 2008.
124. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 2005: 21(16): 3439-3440.
125. Huppmann D, Gidden M, Fricko O, Kolp P, Orthofer C, Pimmer M, Kushin N, Vinca A, Mastrucci A, Riahi K. The MESSAGEix Integrated Assessment Model and the ix modeling platform (ixmp): An open framework for integrated and cross-cutting analysis of energy, climate, the environment, and sustainable development. *Environ Model Softw* 2019: 112: 143-156.
126. Smirnov P, Safikhani Z, El-Hachem N, Wang D, She A, Olsen C, Freeman M, Selby H, Gendoo DM, Grossmann P, Beck AH, Aerts HJ, Lupien M, Goldenberg A, Haibe-Kains B. PharmacoGx: an R package for analysis of large pharmacogenomic datasets. *Bioinformatics* 2016: 32(8): 1244-1246.
127. Kim JH, Shin DM, Lee YS. Effect of local background intensities in the normalization of cDNA microarray data with a skewed expression profiles. *Exp Mol Med* 2002: 34(3): 224-232.
128. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, Smyth GK. A comparison of background correction methods for two-colour microarrays. *Bioinformatics* 2007: 23(20): 2700-2707.
129. Scherer A. Batch effects and noise in microarray experiments: sources and solutions. John Wiley & Sons, United Kingdom, 2009.
130. Kerr MK, Churchill GA. Statistical design and the analysis of gene expression microarray data. *Genet Res* 2001: 77(2): 123-128.
131. Fan J, Li R. Statistical challenges with high dimensionality. In: Proceedings of the international Congress of Mathematicians. Madrid, Spain; 2006; 2006.
132. Manbari Z, AkhlaghianTab F, Salavati C. Hybrid fast unsupervised feature selection for high-dimensional data. *Expert Syst Appl* 2019: 124: 97-118.
133. Zimek A, Schubert E, Kriegel H. A survey on unsupervised outlier detection in high-dimensional numerical data. *Stat Anal Data Min* 2012: 5(5): 363-387.
134. Chum O, Matas J. Unsupervised discovery of co-occurrence in sparse high dimensional data. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2010: IEEE; 2010. p. 3416-3423.
135. Trezzi J, Hiller K, Mollenhauer B. The importance of an independent validation cohort for metabolomics biomarker studies. *Mov Disord* 2018: 33(5): 856-856.
136. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 2016: 69: 245-247.
137. Musoro JZ, Zwinderman AH, Puhan MA, ter Riet G, Geskus RB. Validation of prediction models based on lasso regression with multiply imputed data. *BMC Med Res Methodol* 2014: 14: 116.
138. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol* 2012: 8(2): e1002375.
139. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014: 15(12): 550.
140. Noble WS. How does multiple testing correction work? *Nat Biotechnol* 2009: 27(12): 1135-1137.
141. Hawkins DM. The problem of overfitting. *J Chem Inf Comput Sci* 2004: 44(1): 1-12.

142. Heyckendorf J, van Leth F, Kalsdorf B, Olaru ID, Günther G, Salzer HJF, Terhalle E, Rolling T, Glattki G, Müller M, Schuhmann M, Avsar K, Lange C. Relapse-free cure from multidrug-resistant tuberculosis in Germany. *Eur Respir J* 2018: 51(2).
143. Heyckendorf J, Marwitz S, Reimann M, Avsar K, DiNardo AR, Günther G, Hoelscher M, Ibraim E, Kalsdorf B, Kaufmann SHE, Kontsevaya I, van Leth F, Mandalakas AM, Maurer FP, Müller M, Nitschkowski D, Olaru ID, Popa C, Rachow A, Rolling T, Rybniker J, Salzer HJF, Sanchez-Carballo P, Schuhmann M, Schaub D, Spinu V, Suárez I, Terhalle E, Unnewehr M, Weiner J, 3rd, Goldmann T, Lange C. Prediction of anti-tuberculosis treatment duration based on a 22-gene transcriptomic model. *Eur Respir J* 2021: 58(3).
144. Schaberg T, Bauer T, Castell S, Dalhoff K, Detjen A, Diel R, Greinert U, Hauer B, Lange C, Magdorf K, Loddenkemper R. [Recommendations for therapy, chemoprevention and chemoprophylaxis of tuberculosis in adults and children. German Central Committee against Tuberculosis (DZK), German Respiratory Society (DGP)]. *Pneumologie* 2012: 66(3): 133-171.
145. World Health Organization. WHO treatment guidelines for drug-resistant tuberculosis, 2016 update, Geneva, 2016.
146. Günther G, Lange C, Alexandru S, Altet N, Avsar K, Bang D, Barbuta R, Bothamley G, Ciobanu A, Crudu V, Danilovits M, Dedicoat M, Duarte R, Gualano G, Kunst H, de Lange W, Leimane V, Magis-Escurra C, McLaughlin AM, Muylle I, Polcová V, Popa C, Rumetshofer R, Skrahina A, Solodovnikova V, Spinu V, Tiberi S, Viiklepp P, van Leth F. Treatment Outcomes in Multidrug-Resistant Tuberculosis. *N Engl J Med* 2016: 375(11): 1103-1105.
147. World Health Organization. Definitions and reporting framework for tuberculosis–2013 revision: updated December 2014 and January 2020. Geneva: World Health Organization; 2013. Report No.: 9241505346.
148. Bothamley G. The Tuberculosis Network European Trials Group (TBNET): new directions in the management of tuberculosis. *Breathe (Sheff)* 2017: 13(3): e65-e71.
149. Patrick E, Schramm SJ, Ormerod JT, Scolyer RA, Mann GJ, Mueller S, Yang JY. A multi-step classifier addressing cohort heterogeneity improves performance of prognostic biomarkers in three cancer types. *Oncotarget* 2017: 8(2): 2807-2815.
150. Bogdan M, Ghosh J, Tokdar S. A comparison of the Benjamini-Hochberg procedure with some Bayesian rules for multiple testing. Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen. Institute of Mathematical Statistics, 2008; pp. 211-230.
151. Morey LC, Blashfield RK, Skinner HA. A Comparison of Cluster Analysis Techniques Withing a Sequential Validation Framework. *Multivariate Behav Res* 1983: 18(3): 309-329.
152. Benesty J, Chen J, Huang Y, Cohen I. Pearson correlation coefficient. Noise reduction in speech processing. Springer, Berlin, 2009; pp. 1-4.
153. Held U. Tücken von Korrelationen: die Korrelationskoeffizienten von Pearson und Spearman. In: Swiss Med Forum; 2010: EMH Swiss Medical Publishers; 2010. p. 652-653.
154. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J. g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res* 2016: 44(W1): W83-89.
155. Akaike H. A new look at the statistical model identification. *IEEE Trans Control Syst Technol* 1974: 19(6): 716-723.
156. Burnham K, Anderson D. A practical information-theoretic approach, 2002.
157. Liaw A, Wiener M. Classification and regression by randomForest. *R news* 2002: 2(3): 18-22.
158. Breiman L. Random forests. *Mach learn* 2001: 45(1): 5-32.
159. Archer KJ, Kimes RV. Empirical characterization of random forest variable importance measures. *Computational statistics & data analysis* 2008: 52(4): 2249-2260.
160. Tuszynski J. caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc. *R package version* 2008: 1.
161. Bradley A. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognit* 1997: 30(7): 1145-1159.

162. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986: 42(1): 121-130.
163. Akter S, Xu D, Nagel SC, Bromfield JJ, Pelch K, Wilshire GB, Joshi T. Machine Learning Classifiers for Endometriosis Using Transcriptomics and Methylomics Data. *Front Genet* 2019: 10: 766.
164. Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, Flanagan JM, Vineis P. DNA methylation as a long-term biomarker of exposure to tobacco smoke. *Epidemiology* 2013: 24(5): 712-716.
165. De Meyer G, Shapiro F, Vanderstichele H, Vanmechelen E, Engelborghs S, De Deyn PP, Coart E, Hansson O, Minthon L, Zetterberg H, Blennow K, Shaw L, Trojanowski JQ. Diagnosis-independent Alzheimer disease biomarker signature in cognitively normal elderly people. *Arch Neurol* 2010: 67(8): 949-956.
166. Dervieux T, Greenstein N, Kremer J. Pharmacogenomic and metabolic biomarkers in the folate pathway and their association with methotrexate effects during dosage escalation in rheumatoid arthritis. *Arthritis Rheum* 2006: 54(10): 3095-3103.
167. Zhang Z. Reverse regression: a method for joint analysis of multiple endpoints in randomized clinical trials. *Stat Sin* 2014: 24(4): 1753-1769.
168. Wang H, Li G, Jiang G. Robust regression shrinkage and consistent variable selection through the LAD-Lasso. *J Bus Econ Stat* 2007: 25(3): 347-355.
169. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997: 16(4): 385-395.
170. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol* 1996: 58(1): 267-288.
171. Urminsky O, Hansen C, Chernozhukov V. Using double-lasso regression for principled variable selection. Booth School of Business & Department of Economics and Center for Statistics MIT, Chicago, 2016.
172. Belloni A, Chernozhukov V, Hansen C. Inference on treatment effects after selection among high-dimensional controls. *Rev Econ Stud* 2014: 81(2): 608-650.
173. Belloni A, Chernozhukov V, Kato K. Valid post-selection inference in high-dimensional approximately sparse quantile regression models. *J Am Stat Assoc* 2019: 114(526): 749-758.
174. Heyckendorf J, Reimann M, Marwitz S, Lange C. Pathogen-free diagnosis of tuberculosis. *Lancet Infect Dis* 2021: 21(8): 1066.
175. Anderson ST, Kaforou M, Brent AJ, Wright VJ, Banwell CM, Chagaluka G, Crampin AC, Dockrell HM, French N, Hamilton MS, Hibberd ML, Kern F, Langford PR, Ling L, Mlotha R, Ottenhoff THM, Pienaar S, Pillay V, Scott JAG, Twahir H, Wilkinson RJ, Coin LJ, Heyderman RS, Levin M, Eley B. Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Engl J Med* 2014: 370(18): 1712-1723.
176. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, Wilkinson KA, Banchereau R, Skinner J, Wilkinson RJ, Quinn C, Blankenship D, Dhawan R, Cush JJ, Mejias A, Ramilo O, Kon OM, Pascual V, Banchereau J, Chaussabel D, O'Garra A. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010: 466(7309): 973-977.
177. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, Banwell CM, Brent AJ, Crampin AC, Dockrell HM, Eley B, Heyderman RS, Hibberd ML, Kern F, Langford PR, Ling L, Mendelson M, Ottenhoff TH, Zgambo F, Wilkinson RJ, Coin LJ, Levin M. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med* 2013: 10(10): e1001538.
178. Laux da Costa L, Delcroix M, Dalla Costa ER, Prestes IV, Milano M, Francis SS, Unis G, Silva DR, Riley LW, Rossetti ML. A real-time PCR signature to discriminate between tuberculosis and other pulmonary diseases. *Tuberculosis (Edinb)* 2015: 95(4): 421-425.
179. Maertzdorf J, McEwen G, Weiner J, 3rd, Tian S, Lader E, Schriek U, Mayanja-Kizza H, Ota M, Kenneth J, Kaufmann SH. Concise gene signature for point-of-care classification of tuberculosis. *EMBO Mol Med* 2016: 8(2): 86-95.

180. Sambarey A, Devaprasad A, Baloni P, Mishra M, Mohan A, Tyagi P, Singh A, Akshata JS, Sultana R, Buggi S, Chandra N. Meta-analysis of host response networks identifies a common core in tuberculosis. *NPJ Syst Biol Appl* 2017: 3: 4.
181. Singhania A, Verma R, Graham CM, Lee J, Tran T, Richardson M, Lecine P, Leissner P, Berry MPR, Wilkinson RJ, Kaiser K, Rodrigue M, Woltmann G, Haldar P, O'Garra A. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. *Nat Commun* 2018: 9(1): 2308.
182. Suliman S, Thompson EG, Sutherland J, Weiner J, 3rd, Ota MOC, Shankar S, Penn-Nicholson A, Thiel B, Erasmus M, Maertzdorf J, Duffy FJ, Hill PC, Hughes EJ, Stanley K, Downing K, Fisher ML, Valvo J, Parida SK, van der Spuy G, Tromp G, Adetifa IMO, Donkor S, Howe R, Mayanja-Kizza H, Boom WH, Dockrell HM, Ottenhoff THM, Hatherill M, Aderem A, Hanekom WA, Scriba TJ, Kaufmann SHE, Zak DE, Walzl G. Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis. *Am J Respir Crit Care Med* 2018: 197(9): 1198-1208.
183. Sutherland JS, Loxton AG, Haks MC, Kassa D, Ambrose L, Lee JS, Ran L, van Baarle D, Maertzdorf J, Howe R, Mayanja-Kizza H, Boom WH, Thiel BA, Crampin AC, Hanekom W, Ota MO, Dockrell H, Walzl G, Kaufmann SH, Ottenhoff TH. Differential gene expression of activating Fcγ receptor classifies active tuberculosis regardless of human immunodeficiency virus status or ethnicity. *Clin Microbiol Infect* 2014: 20(4): O230-238.
184. Ilievska-Poposka B, Metodieva M, Zakoska M, Vragoterova C, Trajkov D. Latent Tuberculosis Infection - Diagnosis and Treatment. *Open Access Maced J Med Sci* 2018: 6(4): 651-655.
185. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012: 28(6): 882-883.
186. Weiner 3rd J, Domaszewska T. tmod: an R package for general and multivariate enrichment analysis. *PeerJ Preprints* 2016: 4.
187. Du J, Yuan Z, Ma Z, Song J, Xie X, Chen Y. KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Mol Biosyst* 2014: 10(9): 2441-2447.
188. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, Stichweh D, Blankenship D, Li L, Munagala I, Bennett L, Allantaz F, Mejias A, Ardura M, Kaizer E, Monnet L, Allman W, Randall H, Johnson D, Lanier A, Punaro M, Wittkowski KM, White P, Fay J, Klintmalm G, Ramilo O, Palucka AK, Banchereau J, Pascual V. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* 2008: 29(1): 150-164.
189. Rodríguez Del Águila M, González-Ramírez A. Sample size calculation. *Allergol Immunopathol (Madr)* 2014: 42(5): 485-492.
190. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol* 2010: 5(9): 1315-1316.
191. Bloom CI, Graham CM, Berry MP, Wilkinson KA, Oni T, Rozakeas F, Xu Z, Rossello-Urgell J, Chaussabel D, Banchereau J, Pascual V, Lipman M, Wilkinson RJ, O'Garra A. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. *PLoS One* 2012: 7(10): e46191.
192. Blankley S, Graham CM, Turner J, Berry MP, Bloom CI, Xu Z, Pascual V, Banchereau J, Chaussabel D, Breen R, Santis G, Blankenship DM, Lipman M, O'Garra A. The Transcriptional Signature of Active Tuberculosis Reflects Symptom Status in Extra-Pulmonary and Pulmonary Tuberculosis. *PLoS One* 2016: 11(10): e0162220.
193. Bloom CI, Graham CM, Berry MP, Rozakeas F, Redford PS, Wang Y, Xu Z, Wilkinson KA, Wilkinson RJ, Kendrick Y, Devouassoux G, Ferry T, Miyara M, Bouvry D, Valeyre D, Gorochov G, Blankenship D, Saadatian M, Vanhems P, Beynon H, Vancheeswaran R, Wickremasinghe M, Chaussabel D, Banchereau J, Pascual V, Ho LP, Lipman M, O'Garra A. Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PLoS One* 2013: 8(8): e70630.
194. Sambarey A, Devaprasad A, Mohan A, Ahmed A, Nayak S, Swaminathan S, D'Souza G, Jesuraj A, Dhar C, Babu S, Vyakarnam A, Chandra N. Unbiased Identification of Blood-based

Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks. *EBioMedicine* 2017: 15: 112-126.
195. Maertzdorf J, Ota M, Repsilber D, Mollenkopf HJ, Weiner J, Hill PC, Kaufmann SH. Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis. *PLoS One* 2011: 6(10): e26938.
196. Lange C, Mori T. Advances in the diagnosis of tuberculosis. *Respirology* 2010: 15(2): 220-240.
197. Nguyen MH, Levy NS, Ahuja SD, Trieu L, Proops DC, Achkar JM. Factors Associated With Sputum Culture-Negative vs Culture-Positive Diagnosis of Pulmonary Tuberculosis. *JAMA Netw Open* 2019: 2(2): e187617.
198. Hamada Y, Paulos L, Baruch NG, Cronin W. Proposed Approach for 4-Month Treatment of Culture-Negative Pulmonary Tuberculosis in Adults. *Ann Am Thorac Soc* 2016: 13(9): 1657-1658.
199. Mandal N, Anand PK, Gautam S, Das S, Hussain T. Diagnosis and treatment of paediatric tuberculosis: An insight review. *Crit Rev Microbiol* 2017: 43(4): 466-480.
200. Umair M, Siddiqui SA, Farooq MA. Diagnostic Accuracy of Sputum Microscopy in Comparison With GeneXpert in Pulmonary Tuberculosis. *Cureus* 2020: 12(11): e11383.
201. Acharya B, Acharya A, Gautam S, Ghimire SP, Mishra G, Parajuli N, Sapkota B. Advances in diagnosis of Tuberculosis: an update into molecular diagnosis of Mycobacterium tuberculosis. *Mol Biol Rep* 2020: 47(5): 4065-4075.
202. Zifodya JS, Kreniske JS, Schiller I, Kohli M, Dendukuri N, Schumacher SG, Ochodo EA, Haraka F, Zwerling AA, Pai M, Steingart KR, Horne DJ. Xpert Ultra versus Xpert MTB/RIF for pulmonary tuberculosis and rifampicin resistance in adults with presumptive pulmonary tuberculosis. *Cochrane Database Syst Rev* 2021: 2: Cd009593.
203. Vonasek B, Ness T, Takwoingi Y, Kay AW, van Wyk SS, Ouellette L, Marais BJ, Steingart KR, Mandalakas AM. Screening tests for active pulmonary tuberculosis in children. *Cochrane Database Syst Rev* 2021: 6(6): Cd013693.
204. Gillespie SH, Crook AM, McHugh TD, Mendel CM, Meredith SK, Murray SR, Pappas F, Phillips PP, Nunn AJ. Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis. *N Engl J Med* 2014: 371(17): 1577-1587.
205. Merle CS, Sismanidis C, Sow OB, Gninafon M, Horton J, Lapujade O, Lo MB, Mitchinson DA, Perronne C, Portaels F, Odhiambo J, Olliaro P, Rustomjee R, Lienhardt C, Fielding K. A pivotal registration phase III, multicenter, randomized tuberculosis controlled trial: design issues and lessons learnt from the Gatifloxacin for TB (OFLOTUB) project. *Trials* 2012: 13: 61.
206. Nunn A. An international multicentre controlled clinical trial to evaluate high-dose rifapentine and a quinolone in the treatment of pulmonary tuberculosis (RIFAQUIN). In: TM & IH; 2010; 2010. p. S22-S23.
207. Dorman SE, Nahid P, Kurbatova EV, Phillips PPJ, Bryant K, Dooley KE, Engle M, Goldberg SV, Phan HTT, Hakim J, Johnson JL, Lourens M, Martinson NA, Muzanyi G, Narunsky K, Nerette S, Nguyen NV, Pham TH, Pierre S, Purfield AE, Samaneka W, Savic RM, Sanne I, Scott NA, Shenje J, Sizemore E, Vernon A, Waja Z, Weiner M, Swindells S, Chaisson RE. Four-Month Rifapentine Regimens with or without Moxifloxacin for Tuberculosis. *N Engl J Med* 2021: 384(18): 1705-1718.
208. Van Deun A, Salim MA, Das AP, Bastian I, Portaels F. Results of a standardised regimen for multidrug-resistant tuberculosis in Bangladesh. *Int J Tuberc Lung Dis* 2004: 8(5): 560-567.
209. Trébucq A, Decroo T, Van Deun A, Piubello A, Chiang CY, Koura KG, Schwoebel V. Short-Course Regimen for Multidrug-Resistant Tuberculosis: A Decade of Evidence. *J Clin Med* 2019: 9(1).
210. Moodley R, Godec TR. Short-course treatment for multidrug-resistant tuberculosis: the STREAM trials. *Eur Respir Rev* 2016: 25(139): 29-35.
211. Udwadia ZF, Tornheim JA, Ganatra S, DeLuca A, Rodrigues CS, Gupta A. Few eligible for the newly recommended short course MDR-TB regimen at a large Mumbai private clinic. *BMC Infect Dis* 2019: 19(1): 94.

212. Caminero JA, Piubello A, Scardigli A, Migliori GB. Proposal for a standardised treatment regimen to manage pre- and extensively drug-resistant tuberculosis cases. *Eur Respir J* 2017: 50(1).
213. den Hertog AL, Montero-Martín M, Saunders RL, Blakiston M, Menting S, Sherchand JB, Lawson L, Oladimeji O, Abdurrahman ST, Cuevas LE, Anthony RM. Cytokine Kinetics in the First Week of Tuberculosis Therapy as a Tool to Confirm a Clinical Diagnosis and Guide Therapy. *PLoS One* 2015: 10(6): e0129552.
214. Gebremicael G, Kassa D, Quinten E, Alemayehu Y, Gebreegziaxier A, Belay Y, van Baarle D, Ottenhoff THM, Cliff JM, Haks MC. Host Gene Expression Kinetics During Treatment of Tuberculosis in HIV-Coinfected Individuals Is Independent of Highly Active Antiretroviral Therapy. *J Infect Dis* 2018: 218(11): 1833-1846.
215. Clifford V, He Y, Zufferey C, Connell T, Curtis N. Interferon gamma release assays for monitoring the response to treatment for tuberculosis: A systematic review. *Tuberculosis (Edinb)* 2015: 95(6): 639-650.
216. Denkinger CM, Pai M, Patel M, Menzies D. Gamma interferon release assay for monitoring of treatment response for active tuberculosis: an explosion in the spaghetti factory. *J Clin Microbiol* 2013: 51(2): 607-610.
217. Ahmed MIM, Ziegler C, Held K, Dubinski I, Ley-Zaporozhan J, Geldmacher C, von Both U. The TAM-TB Assay-A Promising TB Immune-Diagnostic Test With a Potential for Treatment Monitoring. *Front Pediatr* 2019: 7: 27.
218. Zhou Y, Du J, Hou HY, Lu YF, Yu J, Mao LY, Wang F, Sun ZY. Application of ImmunoScore Model for the Differentiation between Active Tuberculosis and Latent Tuberculosis Infection as Well as Monitoring Anti-tuberculosis Therapy. *Front Cell Infect Microbiol* 2017: 7: 457.
219. Essone PN, Kalsdorf B, Chegou NN, Loxton AG, Kriel M, Preyer R, Ernst M, Walzl G, Lange C. Bifunctional T-cell-derived cytokines for the diagnosis of tuberculosis and treatment monitoring. *Respiration* 2014: 88(3): 251-261.
220. Kontsevaya I, Lange C, Comella-Del-Barrio P, Coarfa C, DiNardo AR, Gillespie SH, Hauptmann M, Leschczyk C, Mandalakas AM, Martinecz A, Merker M, Niemann S, Reimann M, Rzhepishevska O, Schaible UE, Scheu KM, Schurr E, Abel Zur Wiesch P, Heyckendorf J. Perspectives for systems biology in the management of tuberculosis. *Eur Respir Rev* 2021: 30(160).
221. Tucci P, González-Sapienza G, Marin M. Pathogen-derived biomarkers for active tuberculosis diagnosis. *Front Microbiol* 2014: 5: 549.
222. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 2005: 21(9): 2067-2075.
223. Ranstam J, Cook J. LASSO regression. *BJS* 2018: 105(10): 1348-1348.
224. Backes C, Ludwig N, Leidinger P, Huwer H, Tenzer S, Fehlmann T, Franke A, Meese E, Lenhof HP, Keller A. Paired proteomics, transcriptomics and miRNomics in non-small cell lung cancers: known and novel signaling cascades. *Oncotarget* 2016: 7(44): 71514-71525.
225. Harcourt JL, Offermann MK. Multiple signaling cascades are differentially involved in gene induction by double stranded RNA in interferon-alpha-primed cells. *Eur J Biochem* 2001: 268(5): 1373-1381.
226. Reunanen J. Overfitting in making comparisons between variable selection methods. *J Mach Learn Res* 2003: 3(Mar): 1371-1382.
227. Schaffer C. Overfitting avoidance as bias. *Mach learn* 1993: 10(2): 153-178.
228. Liu H, Wu X, Zhang S. Feature selection using hierarchical feature clustering. In: Proceedings of the 20th ACM international conference on Information and knowledge management; 2011; New York: Association for Computing Machinery; 2011. p. 979-984.
229. Zhou N, Wang L. A modified T-test feature selection method and its application on the HapMap genotype data. *Genomics Proteomics Bioinformatics* 2007: 5(3-4): 242-249.
230. Ma S, Huang J. Penalized feature selection and classification in bioinformatics. *Brief Bioinform* 2008: 9(5): 392-403.
231. Chong I, Jun C. Performance of some variable selection methods when multicollinearity is present. *Chemometr Intell Lab Syst* 2005: 78(1-2): 103-112.

232. Wang H, Lengerich BJ, Aragam B, Xing EP. Precision Lasso: accounting for correlations and linear dependencies in high-dimensional genomic data. *Bioinformatics* 2019: 35(7): 1181-1187.
233. Fan L, Chen S, Li Q, Zhu Z. Variable selection and model prediction based on lasso, adaptive lasso and elastic net. In: 2015 4th International Conference on Computer Science and Network Technology (ICCSNT); 2015; Harbin, China: IEEE; 2015. p. 579-583.
234. Karamizadeh S, Abdullah S, Manaf A, Zamani M, Hooman A. An overview of principal component analysis. *JSIP* 2013: 4(3B): 173-175.
235. Saelens W, Cannoodt R, Saeys Y. A comprehensive evaluation of module detection methods for gene expression data. *Nat Commun* 2018: 9(1): 1090.
236. Ogutu JO, Schulz-Streeck T, Piepho HP. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. *BMC Proc* 2012: 6 Suppl 2(Suppl 2): S10.
237. Omranian N, Eloundou-Mbebi JM, Mueller-Roeber B, Nikoloski Z. Gene regulatory network inference using fused LASSO on multiple data sets. *Sci Rep* 2016: 6: 20533.
238. Loher P, Karathanasis N. Machine Learning Approaches Identify Genes Containing Spatial Information From Single-Cell Transcriptomics Data. *Front Genet* 2020: 11: 612840.
239. Gupta N, Jorwal P. Treatment Outcomes Associated with Multidrug-resistant Tuberculosis. *J Glob Infect Dis* 2018: 10(3): 125-128.
240. Leong S, Zhao Y, Joseph NM, Hochberg NS, Sarkar S, Pleskunas J, Hom D, Lakshminarayanan S, Horsburgh CR, Jr., Roy G, Ellner JJ, Johnson WE, Salgame P. Existing blood transcriptional classifiers accurately discriminate active tuberculosis from latent infection in individuals from south India. *Tuberculosis (Edinb)* 2018: 109: 41-51.
241. Meng Z, Wang M, Guo S, Zhou Y, Lyu M, Hu X, Bai H, Wu Q, Tao C, Ying B. Novel Long Non-coding RNA and LASSO Prediction Model to Better Identify Pulmonary Tuberculosis: A Case-Control Study in China. *Front Mol Biosci* 2021: 8: 632185.
242. Waldmann P, Mészáros G, Gredler B, Fuerst C, Sölkner J. Evaluation of the lasso and the elastic net in genome-wide association studies. *Front Genet* 2013: 4: 270.
243. Richards SA, Whittingham MJ, Stephens PA. Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework. *Behav Ecol Sociobiol* 2011: 65(1): 77-89.
244. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics* 2007: 8: 25.
245. Grochtmann M, Grimm K. Classification trees for partition testing. *STVR* 1993: 3(2): 63-82.
246. Kirasich K, Smith T, Sadler B. Random forest vs logistic regression: binary classification for heterogeneous datasets. *SMU Data Science Review* 2018: 1(3): 9.
247. Hengl T, Nussbaum M, Wright MN, Heuvelink GBM, Gräler B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* 2018: 6: e5518.
248. Lindner C. Automated image interpretation using statistical shape models. Statistical Shape and Deformation Analysis. Elsevier, 2017; pp. 3-32.
249. Muschelli J, Betz J, Varadhan R. Binomial regression in R. Handbook of Statistics. Elsevier, 2014; pp. 257-308.
250. Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics* 2018: 19(1): 270.
251. Burel JG, Babor M, Pomaznoy M, Lindestam Arlehamn CS, Khan N, Sette A, Peters B. Host Transcriptomics as a Tool to Identify Diagnostic and Mechanistic Immune Signatures of Tuberculosis. *Front Immunol* 2019: 10: 221.
252. MacDougall D, Crummett WB. Guidelines for data acquisition and data quality evaluation in environmental chemistry. *Anal Chem* 1980: 52(14): 2242-2249.
253. Scriba TJ, Fiore-Gartland A, Penn-Nicholson A, Mulenga H, Kimbung Mbandi S, Borate B, Mendelsohn SC, Hadley K, Hikuam C, Kaskar M, Musvosvi M, Bilek N, Self S, Sumner T, White RG, Erasmus M, Jaxa L, Raphela R, Innes C, Brumskine W, Hiemstra A, Malherbe ST, Hassan-Moosa R,

Tameris M, Walzl G, Naidoo K, Churchyard G, Hatherill M. Biomarker-guided tuberculosis preventive therapy (CORTIS): a randomised controlled trial. *Lancet Infect Dis* 2021: 21(3): 354-365.
254. Scriba TJ, Penn-Nicholson A, Shankar S, Hraha T, Thompson EG, Sterling D, Nemes E, Darboe F, Suliman S, Amon LM, Mahomed H, Erasmus M, Whatney W, Johnson JL, Boom WH, Hatherill M, Valvo J, De Groote MA, Ochsner UA, Aderem A, Hanekom WA, Zak DE. Sequential inflammatory processes define human progression from M. tuberculosis infection to tuberculosis disease. *PLoS Pathog* 2017: 13(11): e1006687.
255. Alam A, Imam N, Ahmed MM, Tazyeen S, Tamkeen N, Farooqui A, Malik MZ, Ishrat R. Identification and Classification of Differentially Expressed Genes and Network Meta-Analysis Reveals Potential Molecular Signatures Associated With Tuberculosis. *Front Genet* 2019: 10: 932.
256. Barber DL, Mayer-Barber KD, Feng CG, Sharpe AH, Sher A. CD4 T cells promote rather than control tuberculosis in the absence of PD-1-mediated inhibition. *J Immunol* 2011: 186(3): 1598-1607.
257. Dupnik KM, Bean JM, Lee MH, Jean Juste MA, Skrabanek L, Rivera V, Vorkas CK, Pape JW, Fitzgerald DW, Glickman M. Blood transcriptomic markers of Mycobacterium tuberculosis load in sputum. *Int J Tuberc Lung Dis* 2018: 22(8): 950-958.
258. Chen Y, Cao S, Sun Y, Li C. Gene expression profiling of the TRIM protein family reveals potential biomarkers for indicating tuberculosis status. *Microb Pathog* 2018: 114: 385-392.
259. Hibbert L, Pflanz S, De Waal Malefyt R, Kastelein RA. IL-27 and IFN-alpha signal via Stat1 and Stat3 and induce T-Bet and IL-12Rbeta2 in naive T cells. *J Interferon Cytokine Res* 2003: 23(9): 513-522.
260. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS One* 2019: 14(11): e0224365.
261. Kristman V, Manno M, Côté P. Loss to follow-up in cohort studies: how much is too much? *Eur J Epidemiol* 2004: 19(8): 751-760.
262. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ, Jr. Selection Bias Due to Loss to Follow Up in Cohort Studies. *Epidemiology* 2016: 27(1): 91-97.
263. Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: an overview. *Ann Intern Med* 2001: 135(2): 112-123.
264. Chesov D, Alexandru S, Crudu V, Ciobanu N, Botnaru V, Heyckendorf J, Lange C. Failing treatment of multidrug-resistant tuberculosis: a matter of definition. *Int J Tuberc Lung Dis* 2019: 23(4): 522-524.
265. Naftalin CM, Verma R, Gurumurthy M, Hee KH, Lu Q, Yeo BCM, Tan KH, Lin W, Yu B, Seng KY, Lee LS, Paton NI. Adjunctive use of celecoxib with anti-tuberculosis drugs: evaluation in a whole-blood bactericidal activity model. *Sci Rep* 2018: 8(1): 13491.
266. Hamilton CD, Stout JE, Goodman PC, Mosher A, Menzies R, Schluger NW, Khan A, Johnson JL, Vernon AN. The value of end-of-treatment chest radiograph in predicting pulmonary tuberculosis relapse. *Int J Tuberc Lung Dis* 2008: 12(9): 1059-1064.
267. Xie YL, de Jager VR, Chen RY, Dodd LE, Paripati P, Via LE, Follmann D, Wang J, Lumbard K, Lahouar S, Malherbe ST, Andrews J, Yu X, Goldfeder LC, Cai Y, Arora K, Loxton AG, Vanker N, Duvenhage M, Winter J, Song T, Walzl G, Diacon AH, Barry CE, 3rd. Fourteen-day PET/CT imaging to monitor drug combination activity in treated individuals with tuberculosis. *Sci Transl Med* 2021: 13(579).
268. Roe JK, Thomas N, Gil E, Best K, Tsaliki E, Morris-Jones S, Stafford S, Simpson N, Witt KD, Chain B, Miller RF, Martineau A, Noursadeghi M. Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. *JCI Insight* 2016: 1(16): e87238.
269. Gjøen JE, Jenum S, Sivakumaran D, Mukherjee A, Macaden R, Kabra SK, Lodha R, Ottenhoff THM, Haks MC, Doherty TM, Ritz C, Grewal HMS. Novel transcriptional signatures for sputum-independent diagnostics of tuberculosis in children. *Sci Rep* 2017: 7(1): 5839.
270. Gliddon HD, Kaforou M, Alikian M, Habgood-Coote D, Zhou C, Oni T, Anderson ST, Brent AJ, Crampin AC, Eley B, Heyderman R, Kern F, Langford PR, Ottenhoff THM, Hibberd ML, French N, Wright VJ, Dockrell HM, Coin LJ, Wilkinson RJ, Levin M. Identification of Reduced Host

Transcriptomic Signatures for Tuberculosis Disease and Digital PCR-Based Validation and Quantification. *Front Immunol* 2021: 12: 637164.
271. Huang HH, Liu XY, Liang Y, Chai H, Xia LY. Identification of 13 blood-based gene expression signatures to accurately distinguish tuberculosis from other pulmonary diseases and healthy controls. *Biomed Mater Eng* 2015: 26 Suppl 1: S1837-1843.
272. de Araujo LS, Vaas LA, Ribeiro-Alves M, Geffers R, Mello FC, de Almeida AS, Moreira AD, Kritski AL, Lapa ESJR, Moraes MO, Pessler F, Saad MH. Transcriptomic Biomarkers for Tuberculosis: Evaluation of DOCK9. EPHA4, and NPC2 mRNA Expression in Peripheral Blood. *Front Microbiol* 2016: 7: 1586.
273. Qian Z, Lv J, Kelly GT, Wang H, Zhang X, Gu W, Yin X, Wang T, Zhou T. Expression of nuclear factor, erythroid 2-like 2-mediated genes differentiates tuberculosis. *Tuberculosis (Edinb)* 2016: 99: 56-62.
274. Rajan JV, Semitala FC, Mehta T, Seielstad M, Montalvo L, Andama A, Asege L, Nakaye M, Katende J, Mwebe S, Kamya MR, Yoon C, Cattamanchi A. A Novel, 5-Transcript, Whole-blood Gene-expression Signature for Tuberculosis Screening Among People Living With Human Immunodeficiency Virus. *Clin Infect Dis* 2019: 69(1): 77-83.
275. Roe J, Venturini C, Gupta RK, Gurry C, Chain BM, Sun Y, Southern J, Jackson C, Lipman MC, Miller RF, Martineau AR, Abubakar I, Noursadeghi M. Blood Transcriptomic Stratification of Short-term Risk in Contacts of Tuberculosis. *Clin Infect Dis* 2020: 70(5): 731-737.
276. Walter ND, Miller MA, Vasquez J, Weiner M, Chapman A, Engle M, Higgins M, Quinones AM, Rosselli V, Canono E, Yoon C, Cattamanchi A, Davis JL, Phang T, Stearman RS, Datta G, Garcia BJ, Daley CL, Strong M, Kechris K, Fingerlin TE, Reves R, Geraci MW. Blood Transcriptional Biomarkers for Active Tuberculosis among Patients in the United States: a Case-Control Study with Systematic Cross-Classifier Evaluation. *J Clin Microbiol* 2016: 54(2): 274-282.
277. Verhagen LM, Zomer A, Maes M, Villalba JA, Del Nogal B, Eleveld M, van Hijum SA, de Waard JH, Hermans PW. A predictive signature gene set for discriminating active from latent tuberculosis in Warao Amerindian children. *BMC Genomics* 2013: 14: 74.
278. Darboe F, Mbandi SK, Thompson EG, Fisher M, Rodo M, van Rooyen M, Filander E, Bilek N, Mabwe S, Hatherill M, Zak DE, Penn-Nicholson A, Scriba TJ. Diagnostic performance of an optimized transcriptomic signature of risk of tuberculosis in cryopreserved peripheral blood mononuclear cells. *Tuberculosis (Edinb)* 2018: 108: 124-126.
279. Kulkarni V, Queiroz ATL, Sangle S, Kagal A, Salvi S, Gupta A, Ellner J, Kadam D, Rolla VC, Andrade BB, Salgame P, Mave V. A Two-Gene Signature for Tuberculosis Diagnosis in Persons With Advanced HIV. *Front Immunol* 2021: 12: 631165.
280. Jenum S, Bakken R, Dhanasekaran S, Mukherjee A, Lodha R, Singh S, Singh V, Haks MC, Ottenhoff TH, Kabra SK, Doherty TM, Ritz C, Grewal HM. BLR1 and FCGR1A transcripts in peripheral blood associate with the extent of intrathoracic tuberculosis in children and predict treatment outcome. *Sci Rep* 2016: 6: 38841.
281. Day CL, Abrahams DA, Bunjun R, Stone L, de Kock M, Walzl G, Wilkinson RJ, Burgers WA, Hanekom WA. PD-1 Expression on Mycobacterium tuberculosis-Specific CD4 T Cells Is Associated With Bacterial Load in Human Tuberculosis. *Front Immunol* 2018: 9: 1995.
282. Hu JF, Zhang W, Zuo W, Tan HQ, Bai W. Inhibition of the PD-1/PD-L1 signaling pathway enhances innate immune response of alveolar macrophages to mycobacterium tuberculosis in mice. *Pulm Pharmacol Ther* 2020: 60: 101842.
283. Tezera LB, Bielecka MK, Ogongo P, Walker NF, Ellis M, Garay-Baquero DJ, Thomas K, Reichmann MT, Johnston DA, Wilkinson KA, Ahmed M, Jogai S, Jayasinghe SN, Wilkinson RJ, Mansour S, Thomas GJ, Ottensmeier CH, Leslie A, Elkington PT. Anti-PD-1 immunotherapy leads to tuberculosis reactivation via dysregulation of TNF-α. *Elife* 2020: 9.
284. Pan L, Wei N, Jia H, Gao M, Chen X, Wei R, Sun Q, Gu S, Du B, Xing A, Zhang Z. Genome-wide transcriptional profiling identifies potential signatures in discriminating active tuberculosis from latent infection. *Oncotarget* 2017: 8(68): 112907-112916.
285. Tarui M, Shindou H, Kumagai K, Morimoto R, Harayama T, Hashidate T, Kojima H, Okabe T, Nagano T, Nagase T, Shimizu T. Selective inhibitors of a PAF biosynthetic enzyme lysophosphatidylcholine acyltransferase 2. *J Lipid Res* 2014: 55(7): 1386-1396.

286. Morimoto R, Shindou H, Tarui M, Shimizu T. Rapid production of platelet-activating factor is induced by protein kinase Cα-mediated phosphorylation of lysophosphatidylcholine acyltransferase 2 protein. *J Biol Chem* 2014: 289(22): 15566-15576.
287. Singh A, Yamamoto M, Kam SH, Ruan J, Gauvreau GM, O'Byrne PM, FitzGerald JM, Schellenberg R, Boulet LP, Wojewodka G, Kanagaratham C, De Sanctis JB, Radzioch D, Tebbutt SJ. Gene-metabolite expression in blood can discriminate allergen-induced isolated early from dual asthmatic responses. *PLoS One* 2013: 8(7): e67907.
288. Wang J, Teng JL, Zhao D, Ge P, Li B, Woo PC, Liu CH. The ubiquitin ligase TRIM27 functions as a host restriction factor antagonized by Mycobacterium tuberculosis PtpA during mycobacterial infection. *Sci Rep* 2016: 6: 34827.
289. Smith CL, Dickinson P, Forster T, Craigon M, Ross A, Khondoker MR, France R, Ivens A, Lynn DJ, Orme J, Jackson A, Lacaze P, Flanagan KL, Stenson BJ, Ghazal P. Identification of a human neonatal immune-metabolic network associated with bacterial infection. *Nat Commun* 2014: 5: 4649.
290. Izquierdo-Bouldstridge A, Bustillos A, Bonet-Costa C, Aribau-Miralbés P, García-Gomis D, Dabad M, Esteve-Codina A, Pascual-Reguant L, Peiró S, Esteller M, Murtha M, Millán-Ariño L, Jordan A. Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats. *Nucleic Acids Res* 2017: 45(20): 11622-11642.
291. Martino F, Pal M, Muñoz-Hernández H, Rodríguez CF, Núñez-Ramírez R, Gil-Carton D, Degliesposti G, Skehel JM, Roe SM, Prodromou C, Pearl LH, Llorca O. RPAP3 provides a flexible scaffold for coupling HSP90 to the human R2TP co-chaperone complex. *Nat Commun* 2018: 9(1): 1501.
292. Kurachi M, Barnitz RA, Yosef N, Odorizzi PM, DiIorio MA, Lemieux ME, Yates K, Godec J, Klatt MG, Regev A, Wherry EJ, Haining WN. The transcription factor BATF operates as an essential differentiation checkpoint in early effector CD8+ T cells. *Nat Immunol* 2014: 15(4): 373-383.
293. Guler R, Mpotje T, Ozturk M, Nono JK, Parihar SP, Chia JE, Abdel Aziz N, Hlaka L, Kumar S, Roy S, Penn-Nicholson A, Hanekom WA, Zak DE, Scriba TJ, Suzuki H, Brombacher F. Batf2 differentially regulates tissue immunopathology in Type 1 and Type 2 diseases. *Mucosal Immunol* 2019: 12(2): 390-402.
294. Maertzdorf J, Kaufmann SH, Weiner J, 3rd. Toward a unified biosignature for tuberculosis. *Cold Spring Harb Perspect Med* 2014: 5(1): a018531.
295. Thakkinstian A, McEvoy M, Chakravarthy U, Chakrabarti S, McKay GJ, Ryu E, Silvestri G, Kaur I, Francis P, Iwata T, Akahori M, Arning A, Edwards AO, Seddon JM, Attia J. The association between complement component 2/complement factor B polymorphisms and age-related macular degeneration: a HuGE review and meta-analysis. *Am J Epidemiol* 2012: 176(5): 361-372.
296. Sargent CA, Young C, Marsh S, Ferguson-Smith MA, Affara NA. The glycerol kinase gene family: structure of the Xp gene, and related intronless retroposons. *Hum Mol Genet* 1994: 3(8): 1317-1324.
297. Larrouy-Maumus G, Biswas T, Hunt DM, Kelly G, Tsodikov OV, de Carvalho LP. Discovery of a glycerol 3-phosphate phosphatase reveals glycerophospholipid polar head recycling in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* 2013: 110(28): 11320-11325.
298. Diamond MS, Farzan M. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat Rev Immunol* 2013: 13(1): 46-57.
299. Mandalakas AM, Detjen AK, Hesseling AC, Benedetti A, Menzies D. Interferon-gamma release assays and childhood tuberculosis: systematic review and meta-analysis. *Int J Tuberc Lung Dis* 2011: 15(8): 1018-1032.
300. Reich NC. A death-promoting role for ISG54/IFIT2. *J Interferon Cytokine Res* 2013: 33(4): 199-205.
301. Ranjbar S, Haridas V, Jasenosky LD, Falvo JV, Goldfeld AE. A Role for IFITM Proteins in Restriction of Mycobacterium tuberculosis Infection. *Cell Rep* 2015: 13(5): 874-883.
302. Yang G, Xu Y, Chen X, Hu G. IFITM1 plays an essential role in the antiproliferative action of interferon-gamma. *Oncogene* 2007: 26(4): 594-603.

303. Causeret F, Sumia I, Pierani A. Kremen1 and Dickkopf1 control cell survival in a Wnt-independent manner. *Cell Death Differ* 2016: 23(2): 323-332.
304. Tornheim JA, Madugundu AK, Paradkar M, Fukutani KF, Queiroz ATL, Gupte N, Gupte AN, Kinikar A, Kulkarni V, Balasubramanian U, Sreenivasamurthy S, Raja R, Pradhan N, Shivakumar S, Valvi C, Hanna LE, Andrade BB, Mave V, Pandey A, Gupta A. Transcriptomic Profiles of Confirmed Pediatric Tuberculosis Patients and Household Contacts Identifies Active Tuberculosis, Infection, and Treatment Response Among Indian Children. *J Infect Dis* 2020: 221(10): 1647-1658.
305. Pai SG, Carneiro BA, Mota JM, Costa R, Leite CA, Barroso-Sousa R, Kaplan JB, Chae YK, Giles FJ. Wnt/beta-catenin pathway: modulating anticancer immune response. *J Hematol Oncol* 2017: 10(1): 101.
306. Lin SL, Li B, Rao S, Yeo EJ, Hudson TE, Nowlin BT, Pei H, Chen L, Zheng JJ, Carroll TJ, Pollard JW, McMahon AP, Lang RA, Duffield JS. Macrophage Wnt7b is critical for kidney repair and regeneration. *Proc Natl Acad Sci U S A* 2010: 107(9): 4194-4199.
307. Burgin AB, Magnusson OT, Singh J, Witte P, Staker BL, Bjornsson JM, Thorsteinsdottir M, Hrafnsdottir S, Hagen T, Kiselyov AS, Stewart LJ, Gurney ME. Design of phosphodiesterase 4D (PDE4D) allosteric modulators for enhancing cognition with improved safety. *Nat Biotechnol* 2010: 28(1): 63-70.
308. Subbian S, Tsenova L, O'Brien P, Yang G, Koo MS, Peixoto B, Fallows D, Dartois V, Muller G, Kaplan G. Phosphodiesterase-4 inhibition alters gene expression and improves isoniazid-mediated clearance of Mycobacterium tuberculosis in rabbit lungs. *PLoS Pathog* 2011: 7(9): e1002262.
309. Liu H, Palmer D, Jimmo SL, Tilley DG, Dunkerley HA, Pang SC, Maurice DH. Expression of phosphodiesterase 4D (PDE4D) is regulated by both the cyclic AMP-dependent protein kinase and mitogen-activated protein kinase signaling pathways. A potential mechanism allowing for the coordinated regulation of PDE4D activity and expression in cells. *J Biol Chem* 2000: 275(34): 26615-26624.
310. Yang C, Song C, Jung S, Lee K, Kim S, Lee J, Shin A, Oh J, Kwon Y, Kim H. The Phospholipase-Protein Kinase C-MEK-ERK Pathway is Essential in Mycobacteria-induced CCL3 and CCL4 Expression in Human Monocytes. *Immune Netw* 2005: 5(4): 237-246.
311. Shenoy AR, Wellington DA, Kumar P, Kassa H, Booth CJ, Cresswell P, MacMicking JD. GBP5 promotes NLRP3 inflammasome assembly and immunity in mammals. *Science* 2012: 336(6080): 481-485.
312. Fakruddin JM, Lempicki RA, Gorelick RJ, Yang J, Adelsberger JW, Garcia-Pineres AJ, Pinto LA, Lane HC, Imamichi T. Noninfectious papilloma virus-like particles inhibit HIV-1 replication: implications for immune control of HIV-1 infection by IL-27. *Blood* 2007: 109(5): 1841-1849.
313. Pflanz S, Timans JC, Cheung J, Rosales R, Kanzler H, Gilbert J, Hibbert L, Churakova T, Travis M, Vaisberg E, Blumenschein WM, Mattson JD, Wagner JL, To W, Zurawski S, McClanahan TK, Gorman DM, Bazan JF, de Waal Malefyt R, Rennick D, Kastelein RA. IL-27, a heterodimeric cytokine composed of EBI3 and p28 protein, induces proliferation of naive CD4+ T cells. *Immunity* 2002: 16(6): 779-790.
314. Abdalla AE, Li Q, Xie L, Xie J. Biology of IL-27 and its role in the host immunity against Mycobacterium tuberculosis. *Int J Biol Sci* 2015: 11(2): 168-175.
315. Charlot-Rabiega P, Bardel E, Dietrich C, Kastelein R, Devergne O. Signaling events involved in interleukin 27 (IL-27)-induced proliferation of human naive CD4+ T cells and B cells. *J Biol Chem* 2011: 286(31): 27350-27362.
316. Cheng C, Wang Q, Zhu M, Liu K, Zhang Z. Integrated analysis reveals potential long non-coding RNA biomarkers and their potential biological functions for disease free survival in gastric cancer patients. *Cancer Cell Int* 2019: 19: 123.
317. Schoenfeld AK, Lahrsen E, Alban S. Regulation of Complement and Contact System Activation via C1 Inhibitor Potentiation and Factor XIIa Activity Modulation by Sulfated Glycans - Structure-Activity Relationships. *PLoS One* 2016: 11(10): e0165493.

318. Weiner J, Domaszewska T, Donkor S, Kaufmann SHE, Hill PC, Sutherland JS. Changes in Transcript, Metabolite, and Antibody Reactivity During the Early Protective Immune Response in Humans to Mycobacterium tuberculosis Infection. *Clin Infect Dis* 2020: 71(1): 30-40.
319. Son M, Diamond B, Santiago-Schwarz F. Fundamental role of C1q in autoimmunity and inflammation. *Immunol Res* 2015: 63(1-3): 101-106.
320. Liu B, Liao J, Rao X, Kushner SA, Chung CD, Chang DD, Shuai K. Inhibition of Stat1-mediated gene activation by PIAS1. *Proc Natl Acad Sci U S A* 1998: 95(18): 10626-10631.
321. Yao K, Chen Q, Wu Y, Liu F, Chen X, Zhang Y. Unphosphorylated STAT1 represses apoptosis in macrophages during Mycobacteriumtuberculosis infection. *J Cell Sci* 2017: 130(10): 1740-1751.
322. Manca C, Tsenova L, Freeman S, Barczak AK, Tovey M, Murray PJ, Barry C, Kaplan G. Hypervirulent M. tuberculosis W/Beijing strains upregulate type I IFNs and increase expression of negative regulators of the Jak-Stat pathway. *J Interferon Cytokine Res* 2005: 25(11): 694-701.
323. Li X, Wang F, Xu X, Zhang J, Xu G. The Dual Role of STAT1 in Ovarian Cancer: Insight Into Molecular Mechanisms and Application Potentials. *Front Cell Dev Biol* 2021: 9: 636595.
324. Esquivel-Solís H, Quiñones-Falconi F, Zarain-Herzberg A, Amieva-Fernández RI, López-Vidal Y. Impaired activation of Stat1 and c-Jun as a possible defect in macrophages of patients with active tuberculosis. *Clin Exp Immunol* 2009: 158(1): 45-54.
325. Sugawara I, Yamada H, Mizuno S. STAT1 knockout mice are highly susceptible to pulmonary mycobacterial infection. *Tohoku J Exp Med* 2004: 202(1): 41-50.
326. Pan G, Bauer JH, Haridas V, Wang S, Liu D, Yu G, Vincenz C, Aggarwal BB, Ni J, Dixit VM. Identification and functional characterization of DR6, a novel death domain-containing TNF receptor. *FEBS Lett* 1998: 431(3): 351-356.
327. Bidère N, Su HC, Lenardo MJ. Genetic disorders of programmed cell death in the immune system. *Annu Rev Immunol* 2006: 24: 321-352.
328. Zeng Q, Subramaniam VN, Wong SH, Tang BL, Parton RG, Rea S, James DE, Hong W. A novel synaptobrevin/VAMP homologous protein (VAMP5) is increased during in vitro myogenesis and present in the plasma membrane. *Mol Biol Cell* 1998: 9(9): 2423-2437.
329. Ikezawa M, Tajika Y, Ueno H, Murakami T, Inoue N, Yorifuji H. Loss of VAMP5 in mice results in duplication of the ureter and insufficient expansion of the lung. *Dev Dyn* 2018: 247(5): 754-762.
330. Stow JL, Manderson AP, Murray RZ. SNAREing immunity: the role of SNAREs in the immune system. *Nat Rev Immunol* 2006: 6(12): 919-929.