

Review

High-throughput and computational approaches for diagnostic and prognostic host tuberculosis biomarkers



January Weiner*, Stefan H.E. Kaufmann

Max Planck Institute for Infection Biology, Department of Immunology, 10117 Berlin, Germany

ARTICLE INFO

Article history:

Received 20 October 2016

Accepted 20 October 2016

Corresponding Editor: Eskild Petersen, Aarhus, Denmark

Keywords:

Tuberculosis

Biomarker

High-throughput

Gene set enrichment analysis

SUMMARY

High-throughput techniques strive to identify new biomarkers that will be useful for the diagnosis, treatment, and prevention of tuberculosis (TB). However, their analysis and interpretation pose considerable challenges. Recent developments in the high-throughput detection of host biomarkers in TB are reported in this review.

© 2016 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

There is a widely acknowledged need for novel biomarkers of tuberculosis (TB), for all levels of TB diagnosis, treatment, and prevention.^{1–3} Much effort in this direction has been devoted to host biomarkers, because progression towards clinical disease can be detected by specific changes that are evoked by the pathological processes in the host organism. In TB, this is a particular advantage, as the diagnosis of direct symptoms of TB (e.g., by auscultation or chest X-ray (CXR), through detection of the causative agent, acid-fast bacteria in sputum by microscopy, or positive bacterial cultures) may sometimes be problematic. Sputum samples are difficult to obtain from neonates, who moreover frequently suffer from extrapulmonary TB.

However, there are further reasons to focus on the host response. The onset of active TB disease is frequently delayed for years, and the time span between the first TB symptoms and diagnosis has been estimated to range from 5 days to as long as 162 days.⁴ Thus, TB may exist without apparent symptoms, although the molecular processes underlying TB pathology have already commenced. Likewise, TB may persist in a subclinical stage after drug treatment and may later relapse. Positron emission computed tomography (PET/CT) has revealed hallmarks of active TB in patients who have been treated successfully.⁵ Host

biomarkers may provide a sensitive and specific approach to detect subclinical manifestations of clinical or subclinical TB.

The early detection of TB is another important area for biomarker research. Of two billion *Mycobacterium tuberculosis*-infected individuals, most remain healthy but infected (latent TB infection, LTBI) and only a fraction of 5–7% will develop clinical TB during their lifetime. Although *M. tuberculosis* infection can be determined reliably by interferon gamma release assay (IGRA), this test cannot be used to diagnose or determine the prognosis of active TB.⁶ Thus, the identification of biomarkers of TB risk and early stage of progression to active TB would allow screening for individuals at risk. This would allow preventive drug therapy, and also interruption of transmission, with a marked influence on treatment success. Practically, the treatment outcome cannot be assessed in a point-of-care setting. Although PET/CT has predictive value for the treatment outcome,⁷ simpler and more accessible tests have thus far failed. For example, although CXR allows a reliable diagnosis of TB, it has limited predictive value for the treatment outcome.⁸ Early and personalized treatment adjustment, as well as prediction of the treatment outcome in new drug trials is a major concern in the face of increasing incidences of drug-resistant TB.

2. Computational approaches to high-throughput biomarkers

High-throughput techniques such as transcriptomics allow the inspection of tens of thousands of variables (such as gene expression, protein or metabolite levels) in one step (A glossary

* Corresponding author.

E-mail address: january@mpiib-berlin.mpg.de (J. Weiner).

Table 1
Glossary

Biomarker	A measurable indicator of the organism state.
Signature	A set of individual biomarkers, corresponding values, and specific machine learning models, which act together as an indicator of the state of an organism.
Predictive vs. prospective	Biomarkers that allow the prediction of the likely natural course of the untreated disease in the individual are termed 'prospective biomarkers'. Biomarkers that allow the prediction of the outcome of treatment are termed 'predictive biomarkers'.
Machine learning (ML)	Methods in computer science that allow the construction of a model of reality based on automatic inspection of data. In 'supervised ML', a model of reality is first derived from a training data set, and subsequently validated by application to a test data set. For example, a model can be trained on gene expression data from TB patients and healthy controls. Its performance will then be evaluated by applying the model to a separate validation set.
ROC curve	A curve describing the predictive ability of a supervised ML model, showing all possible combinations of specificity and sensitivity that can be obtained from that model.
Random forests	A type of supervised ML in which a large number of partially randomized decision trees is generated. When applied to a sample, each tree casts a vote, and the model then decides on the classification of the sample by majority rule.
Variable importance	A measure that determines the relative importance of different variables for correctly classifying a sample by a machine learning model.
Gene set enrichment analysis	Genes (or other variables) can be grouped into functional categories such as gene ontology sets, co-expression modules, or sets of genes that are up- or down-regulated in a particular condition or are specific for a given cell subtype. Gene set enrichment analysis can take advantage of such a classification by testing whether a particular category of genes (e.g., interferon inducible genes or monocyte surface proteins) are enriched in genes that are strongly regulated in a given comparison (e.g., TB vs. healthy controls).

of the terms used in this article is given in Table 1). However, the large number of variables (compared to the number of samples analysed) is a two-edged sword. The obvious advantage of such an approach is the comparatively unbiased acquisition of a large number of potential candidates. On the other hand, if the number of variables is much larger than the number of samples utilized, sophisticated and careful statistical analyses are necessary. Most importantly, the statistical power for detecting a single or a few suitable biomarkers amongst the thousands of variables analysed decreases profoundly, thus correct signals are often hidden in a deluge of false-positives. Moreover, given that the number of functionally characterized protein-coding genes remains insufficient, and only a few microRNAs have been functionally characterized, the interpretation of results may pose an additional obstacle.

Data mining tools such as supervised and unsupervised learning have been employed successfully in a number of biomarker studies.^{2,9} Supervised machine learning algorithms include both established methods (such as linear discrimination analysis,¹⁰ k-nearest neighbour algorithm,¹¹ and random forests^{12,13}) and novel, unique approaches. Zak et al.¹⁴ constructed a new classification method by combining k-top-scoring pairs¹⁵ with support vector machines (SVMs), taking advantage of relatively simple interpretability of the k-top-scoring pairs approach with the flexibility of SVMs. Kaforou et al. defined a new metrics termed the 'disease risk score' (DSR), defined as the sum of signed absolute intensities of discriminatory biomarkers, combined with a TB/no TB threshold.¹⁶ Despite the computational simplicity of DSR, it was shown to perform well in discriminating TB patients both from healthy individuals and from patients suffering from other diseases.

A disease signature is only superficially a compilation of variables (e.g., genes) that differ between two conditions. Firstly, as a minimum these variables are linked to particular values (e.g., gene expression in healthy individuals and in TB patients, as in the k-nearest neighbour algorithm) or more complex structures (e.g., decision trees). Secondly, most machine learning algorithms provide a score, which subsequently is compared to an arbitrarily chosen threshold. This latter step, however, depends on a given context, because modifying the threshold optimizes either specificity or sensitivity. As a solution, results of such biomarker analyses are frequently shown as so-called receiver operating characteristic (ROC) curves—all possible sensitivity/specificity combinations for a given signature (Figure 1A).

The interpretation of signatures is increasingly confounded by the size and complexity of the model. While biological functions to

which a four-gene signature is related may be glimpsed with relative ease, it is much harder to gain an overview in more complex cases. However, machine learning algorithms often allow the calculation of a 'variable importance' (VI) measure. VI can be used to rank genes according to their contribution to the model, which in turn can be used by adapting a gene set enrichment analysis framework such as GSEA²¹, piano,²² or tmod.²⁰ In the case of a shrunken model based on a subset of genes, the subset itself can be tested for enrichment in relevant classes of genes.

Note that all statistical approaches are based on assumptions, which incompletely fit the biological reality. Moreover, the large number of variables tested in a high-throughput setting increases the risk of false-positives, even when strictly adhering to standards in statistical methodology, e.g. by using a suitable method for family-wise error correction. It has been estimated that at $p < 0.05$, as many as 30% of the rejected hypotheses may be false-positives²³ irrespective of using a correction for multiple testing, which may be one of the reasons for the much debated 'reproducibility crisis' in science. The point here is that high-throughput analyses are especially vulnerable to these problems.

Three not mutually exclusive approaches are suggested here, which do not require additional statistical assumptions or novel techniques. Firstly, because unblinded studies overestimate the actual observed effect size,²⁴ any biomarker study in future should consider separating ('locking') a randomly chosen subset of samples for a blinded, post-hoc validation of the findings, and studies should be evaluated by adherence to this rule. Secondly, an independent analysis by several statisticians (both as study authors and reviewers) would greatly increase confidence in the findings. Thirdly, biomarker studies need to be validated in various settings and cohorts, and using independent experimental approaches. This would facilitate the process of translating the high-throughput to practical clinical applications.

3. High-throughput biomarkers in TB

3.1. Transcriptomic profiling

High-throughput-derived transcriptomic biomarkers have been studied for almost a decade in TB, with the first studies appearing in 2007.^{2,10,25} The broadly studied differences in gene expression between TB patients and healthy (infected or uninfected) controls thus far have been investigated in a total of over a thousand individuals on four continents. Kaforou et al. included over 500 individuals in two cohorts, not only TB patients and healthy controls (both HIV-negative and HIV-positive), but also

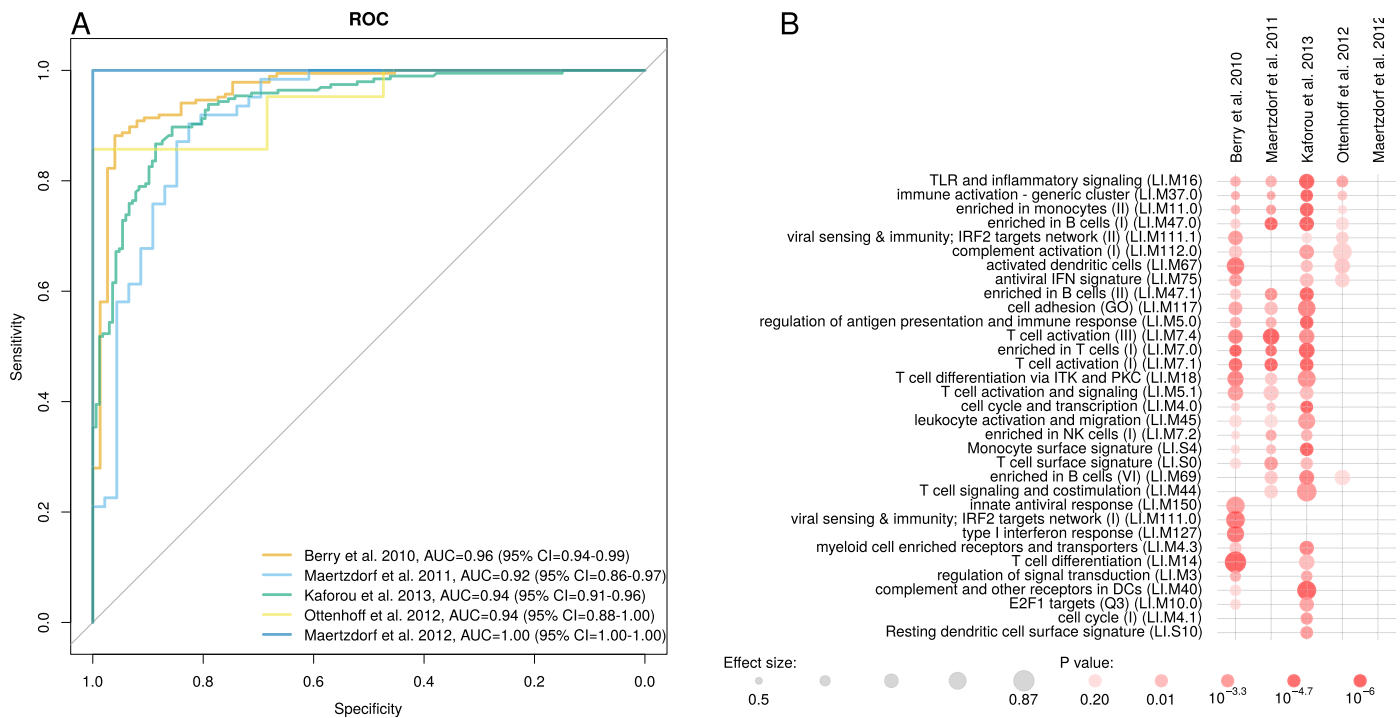


Figure 1. Classification results of random forest training for transcriptomic samples from TB patients compared to healthy controls. (A) Receiver operating characteristic (ROC) curves showing the relative performance of the transcriptomic signatures in distinguishing between the two groups. (B) Results of the gene set enrichment analysis as applied to genes, sorted by their importance in the random forest models. The size of the points indicates the effect size of the gene set enrichment (AUC), and bolder colours are used for lower q -values. The transcriptomic profiles were derived from five independent studies^{11,16,17–19} and the enrichment was calculated using the tmod package.²⁰

patients who were suspected of having TB but who had been clinically diagnosed with other diseases.¹⁶ The biomarkers identified include components of the interferon gamma response (such as CD64, identified by Jacobsen et al.²⁵), neutrophil-driven interferon signature,^{11,17,19} down-regulation of T-cell- and B-cell-related genes, and others.^{26,27} These initial signatures were defined as sets of differentially regulated genes characteristic of gene expression in the blood of TB patients and involved up to several hundred genes. Despite the universality of qualitative findings, the balance in the extent of regulation found to occur in different areas of the host response may differ between the studies. For example, random forest models based on data derived from Berry et al.¹¹ show a strong interferon response dominating the signature (in concordance with the main conclusions of the authors), while the data from Kaforou et al.¹⁶ are dominated by changes in expression of T-cell- and B-cell-related genes (see Figure 1B). It is difficult to decide by purely computational criteria whether these shifts in the different parts of the immune system may be due to different technical platforms or whether they reflect real biological differences.

Meta-analyses combined with a rigorous variable selection process allow the size of transcriptomic biosignatures to be reduced to as few as three or four genes (Maertzdorf et al.,¹³ Sweeney et al.²⁸). In fact, a concise (shrunken) four-gene signature based on comparison between healthy controls and TB patients turned out to be more specific for TB than larger signatures involving 15 or more genes, and even allowed the discrimination between TB and other pulmonary diseases when applied to independent data sets and cohorts.¹³ It is possible that a broader signature involves genes that capture the general aspects of TB disease (such as inflammation or other unspecific innate responses) shared with other diseases. Note that these concise transcriptomic signatures are remarkably universal, and the models built can be successfully validated on data sets derived from other cohorts and technical platforms used to obtain the data.

Recently, transcriptomic profiling has been applied in a longitudinal study with the goal of obtaining a predictive signature for active TB disease. Zak and colleagues followed healthy adolescents from South Africa for 2 years, collecting blood samples every 6 months.¹⁴ Out of several thousand study participants, 46 individuals were eventually diagnosed with TB. Transcriptomic profiles were obtained from blood samples of these individuals, collected prior to the time point of TB diagnosis, and were compared to profiles of those individuals who remained healthy throughout the study. Indeed, these profiles (which were all collected from apparently healthy individuals) were able to discriminate between the two groups in the study design, with statistical significance, even though the statistical performance (as it was to be expected) was not comparable to the power of transcriptomic profiles in discriminating between healthy individuals and TB patients. The results of the study have been confirmed using an independent set of samples obtained from another longitudinal cohort collected in the Grand Challenges GC6–74 effort.²⁹

There were two further notable findings in this study. Firstly, the biomarkers identified largely coincided with the biomarkers for clinical TB diagnosis, including CD64 (identified by Jacobsen et al.²⁵) and several interferon inducible genes.¹¹ In other words, the prognostic or predictive signature of TB obtained overlaps with the diagnostic signature of clinical TB. Secondly, there was a clear time dependence relative to time point of clinical diagnosis: samples obtained within the 12 to 18 months prior to clinical diagnosis produced a predictive signature, but samples obtained earlier did not. These findings suggest that the biomarkers identified do not correspond to a persistent TB risk (or, reciprocally, an inherent protectivity), but more likely are indicative of an incipient, subclinical form of TB. This is in line with recent findings that apparently healthy individuals after successful drug treatment show an ongoing TB process that can be captured with PET.⁵

This hypothesis will prompt further studies in other or larger cohorts.

The elephant in the room when it comes to blood transcriptome studies is the fact that it is generally impossible to reliably distinguish between bona fide gene regulation within a cell and changes or differences in the composition of the cell populations constituting the analysed samples. While differential cell counts sometimes accompany blood transcriptomic data, this is likely insufficient if, for example, the observed effects are due to changes in the migration pattern of T lymphocytes or specific subtypes thereof. Future analyses involving single-cell RNA sequencing (scRNASeq) may shed further light on fine differences in the state of the individual cells and cell compositions of the investigated tissue.

3.2. Further high-throughput approaches

Transcriptomic analyses have been followed by large-scale proteomic analyses. De Groote et al. used a highly multiplexed proteomics approach to analyse serum from TB patients before and after treatment and identified a number of potential biomarkers, including C-reactive protein, the metalloproteinase inhibitor TIMP2, thrombospondin 4 (THBS4), and serum amyloid A (SAA), as well as a number of involved pathways, including microbial pattern recognition, coagulation and complement system, as well as, to lesser extent, the interferon gamma pathway.³⁰ Similar results were obtained in a large-scale comparison of serum protein biomarkers from TB patients and healthy controls.³¹

Several studies have analysed the differences in microRNA profiles between TB patients and healthy controls, both in RNA extracted from peripheral blood cells^{18,32–35} and freely circulating microRNAs found in patient serum or plasma samples.^{36–38} However, our incomplete knowledge about microRNA functions makes it hard to reliably interpret mere lists of identifiers in terms of biological functionality. In-depth computational and experimental assessment of candidate biomarkers shows that microRNAs can play an important role in regulating the immune response,³⁹ for example by influencing the recruitment of neutrophils to the lung.³⁴

Metabolic profiling of serum metabolites by mass spectrometry has demonstrated excellent performance in discriminating TB patients from healthy controls.¹² The analyses revealed changes in lysophosphatidylcholines, amino acids (notably glutamine and glutamate), bile acids, and fibrinopeptides, and the top biomarkers included inosine, cortisol, and kynurenine. Kynurenine is known to correlate with increased expression of indoleamine 2,3-dioxygenase upon contact with *M. tuberculosis*, while adenosine deaminase, which enzymatically converts adenosine to inosine, has been identified as a potential serum proteomic biomarker for TB.⁴⁰ Zhou et al. confirmed some of these findings in a nuclear magnetic resonance (NMR) rather than mass spectrometry based approach.⁴¹ Another study also showed changes in metabolic profiles upon TB drug treatment; however, the study design did not allow for an unambiguous annotation of the unique molecular features collected.⁴² Frediani et al. identified metabolites in the plasma samples of 17 TB patients and found, in addition to some of the previous findings, a higher abundance of resolvins and compounds that may directly be derived from *M. tuberculosis* cell wall lysis.⁴³

Few investigators have considered epigenetic modifications as potential TB biomarkers. Esterhuysen et al. simultaneously collected DNA methylation data, transcriptomic profiles (including microRNAs), and proteomic profiles from monocytes and neutrophils isolated from TB patients and healthy controls.³⁹

At the bottom line, these studies demonstrate that the deep impact of TB on the host can reliably be acquired by high-

throughput techniques at all levels tested to date. Significant changes in TB can be observed for virtually any tissue and molecule type tested. Universal, but TB-specific patterns emerge, including the interferon response or changes in host metabolism. In spite of this, the data are both too rich and too poor.

Firstly, studies comparing TB to healthy controls are abundant and have generated a diverse landscape of data sets. Meta-analyses (such as the one performed on transcriptomic data by Joosten et al.²⁶ and that by Sweeney et al.²⁸) are now a key task for computational analyses. Unfortunately, while primary transcriptomic readouts (i.e., signal intensities for microarrays or read counts for RNASeq) are usually readily available from the GEO database, other types of data are less frequently disclosed upon publication.

On the other hand, most of the aforementioned study designs have primarily considered the comparison between TB patients and healthy controls or TB patients before and after treatment. Although transcriptomic profiling demonstrates that even signatures derived purely from such designs can be used to reliably discriminate between TB and other diseases with similar symptoms,¹³ the inclusion of patients with other diseases in the study design of metabolic, epigenetic, and proteomic profiling is a necessary next step.

4. Outlook

High-throughput techniques such as transcriptomics have demonstrated their ability to not only distinguish TB patients from healthy individuals, but also to discriminate between TB and other diseases, monitoring treatment outcome and even predicting the onset of active TB months before a clinical diagnosis can be performed. Although current biosignatures are composed of dozens, if not hundreds of variables, first attempts to reduce the number of transcripts in transcriptomic profiling of TB show that not only is the information contained in such large biosignatures redundant, but that more specific signatures can be derived by a more selective approach.¹³

Biomarkers in TB are by no means limited to the transcriptome, and several studies have shown that TB manifests itself on different levels. Given that even within blood, transcriptomic profiles (derived only from peripheral blood cells) are not necessarily fully correlated to profiles of serum molecules (derived from other cells as well as peripheral blood cells), combining biomarkers from these different platforms may improve the overall performance of biosignatures.

Most importantly, new, independent studies in different cohorts are needed to allow for meta-analyses and the construction of concise, universal, and predictive TB biosignatures.

Acknowledgements

SEK acknowledges support from The European Union's Seventh Framework Programme (EU FP7) ADITEC (HEALTH-F4-2011-280873); the EU Horizon 2020 project TBVAC 2020 (grant number 643381); The Bill & Melinda Gates Foundation (BMGF) GC6-2013, #OPP 1055806 and #OPP 1065330; the Bundesministerium für Bildung und Forschung (BMBF) project "Infect Control 2020".

Conflict of interest: None.

References

- Walzl G, Ronacher K, Hanekom W, Scriba TJ, Zumla A. Immunological biomarkers of tuberculosis. *Nat Rev Immunol* 2011;11:343–54.
- Maertzdorf J, Kaufmann SH, Weiner J. Toward a unified biosignature for tuberculosis. *Cold Spring Harb Perspect Med* 2014;5:a018531.

3. Goletti D, Petruccioli E, Joosten SA, Ottenhoff TH. Tuberculosis biomarkers: from diagnosis to protection. *Infect Dis Rep* 2016;**8**:6568.
4. Sreeramareddy CT, Panduru KV, Menten J, Van den Ende J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. *BMC Infect Dis* 2009;**9**:1.
5. Esmail H, Lai RP, Lesosky M, Wilkinson KA, Graham CM, Coussens AK, et al. Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[18F] fluoro-D-glucose positron emission and computed tomography. *Nat Med* 2016;**22**:1090–3.
6. Rangaka MX, Wilkinson KA, Glynn JR, Ling D, Menzies D, Mwansa-Kambafwe J, et al. Predictive value of interferon- γ release assays for incident active tuberculosis: a systematic review and meta-analysis. *Lancet Infect Dis* 2012;**12**:45–55.
7. Chen RY, Dodd LE, Lee M, Paripati P, Hammoud DA, Mountz JM, et al. PET/CT imaging correlates with treatment outcome in patients with multidrug-resistant tuberculosis. *Sci Transl Med* 2014;**6**:265ra166–266ra.
8. Kriel M, Lotz JW, Kidd M, Walzl G. Evaluation of a radiological severity score to predict treatment outcome in adults with pulmonary tuberculosis. *Int J Tuberc Lung Dis* 2015;**19**:1354–60.
9. Weiner III J, Kaufmann SH. Recent advances towards tuberculosis control: vaccines and biomarkers. *J Intern Med* 2014;**275**:467–80.
10. Mistry R, Cliff JM, Clayton CL, Beyers N, Mohamed YS, Wilson PA, et al. Gene-expression patterns in whole blood identify subjects at risk for recurrent tuberculosis. *J Infect Dis* 2007;**195**:357–65.
11. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010;**466**:973–7.
12. Weiner 3rd J, Parida SK, Maertzdorf J, Black GF, Replibser D, Telaar A, et al. Biomarkers of inflammation, immunosuppression and stress with active disease are revealed by metabolomic profiling of tuberculosis patients. *PLoS One* 2012;**7**:e40221.
13. Maertzdorf J, McEwen G, Weiner J, Tian S, Lader E, Schriek U, et al. Concise gene signature for point-of-care classification of tuberculosis. *EMBO Mol Med* 2016;**8**:86–95.
14. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* 2016;**387**:2312–22.
15. Tan AC, Naiman DQ, Xu L, Winslow RL, Geman D. Simple decision rules for classifying human cancers from gene expression profiles. *Bioinformatics* 2005;**21**:3896–904.
16. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, et al. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med* 2013;**10**:e1001538.
17. Maertzdorf J, Ota M, Replibser D, Mollenkopf HJ, Weiner J, Hill PC, et al. Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis. *PLoS One* 2011;**6**:e26938.
18. Maertzdorf J, Weiner J, Mollenkopf HJ, TBoNotTB Network, Bauer T, Prasse A, et al. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proc Natl Acad Sci U S A* 2012;**109**:7853–8.
19. Ottenhoff TH, Dass RH, Yang N, Zhang MM, Wong HE, Sahiratmadja E, et al. Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis. *PLoS One* 2012;**7**:e45839.
20. Weiner 3rd J, Domaszewska T. tmod: an R package: for general and multivariate enrichment analysis. *PeerJ Preprints* 2016;**4**:1–9.
21. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-p: a desktop application for gene set enrichment analysis. *Bioinformatics* 2007;**23**:3251–3.
22. Våremo L, Nielsen J, Nookaew I. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res* 2013;**41**:4378–91.
23. Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. *Open Sci* 2014;**1**:140216.
24. Crossley NA, Sena E, Goehler J, Horn J, van der Worp B, Bath PM, et al. Empirical evidence of bias in the design of experimental stroke studies: a metaepidemiologic approach. *Stroke* 2008;**39**:929–34.
25. Jacobsen M, Replibser D, Gutschmidt A, Neher A, Feldmann K, Mollenkopf HJ, et al. Candidate biomarkers for discrimination between infection and disease caused by *Mycobacterium tuberculosis*. *J Mol Med* 2007;**85**:613–21.
26. Joosten SA, Fletcher HA, Ottenhoff TH. A helicopter perspective on TB biomarkers: pathway and process based analysis of gene expression data provides new insight into TB pathogenesis. *PLoS One* 2013;**8**:e73230.
27. Cliff JM, Kaufmann SH, McShane H, Helden P, O'Garra A. The human immune response to tuberculosis and its treatment: a view from the blood. *Immunol Rev* 2015;**264**:88–102.
28. Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir Med* 2016;**4**:213–24.
29. Kaufmann SH, Parida SK. Tuberculosis in Africa: learning from pathogenesis for biomarker identification. *Cell Host Microbe* 2008;**4**:219–28.
30. De Groot MA, Nahid P, Jarlsberg L, Johnson JL, Weiner M, Muzanyi G, et al. Elucidating novel serum biomarkers associated with pulmonary tuberculosis treatment. *PLoS One* 2013;**8**:e61002.
31. Achkar JM, Cortes L, Croteau P, Yanofsky C, Mentinova M, Rajotte I, et al. Host protein biomarkers identify active tuberculosis in HIV uninfected and co-infected individuals. *EBioMedicine* 2015;**2**:1160–8.
32. Wu J, Lu C, Diao N, Zhang S, Wang S, Wang F, et al. Analysis of microRNA expression profiling identifies miR-155 and miR-155* as potential diagnostic markers for active tuberculosis: a preliminary study. *Hum Immunol* 2012;**73**:31–7.
33. Spinelli SV, Diaz A, D'Attilio L, Marchesini MM, Bogue C, Bay ML, et al. Altered microRNA expression levels in mononuclear cells of patients with pulmonary and pleural tuberculosis and their relation with components of the immune response. *Mol Immunol* 2013;**53**:265–9.
34. Dorhoi A, Iannaccone M, Farinacci M, Faè KC, Schreiber J, Moura-Alves P, et al. MicroRNA-223 controls susceptibility to tuberculosis by regulating lung neutrophil recruitment. *J Clin Invest* 2013;**123**:4836–48.
35. Iannaccone M, Dorhoi A, Kaufmann SH. Host-directed therapy of tuberculosis: what is in it for microRNA? *Exp Opin Ther Targets* 2014;**18**:491–4.
36. Yi Z, Fu Y, Wu X, Li J, Xu F. Circulating microRNAs in patients with active pulmonary tuberculosis. *J Clin Microbiol* 2011;**49**:4246–51.
37. Qi Y, Cui L, Ge Y, Shi Z, Zhao K, Guo X, et al. Altered serum microRNAs as biomarkers for the early diagnosis of pulmonary tuberculosis infection. *BMC Infect Dis* 2012;**12**:1.
38. Abd-El-Fattah AA, Sadik NA, Shaker OG, Aboulfouh ML. Differential microRNAs expression in serum of patients with lung cancer, pulmonary tuberculosis, and pneumonia. *Cell Biochem Biophys* 2013;**67**:875–84.
39. Esterhuysen MM, Weiner J, Caron E, Loxton AG, Iannaccone M, Wagman C, et al. Epigenetics and proteomics join transcriptomics in the quest for tuberculosis biomarkers. *MBio* 2015;**6**:e01187–215.
40. Pandey R, Tamrakar D, Jaiswal S, Sharma A, Koju S, Duwal SR. Serum adenosine deaminase: a novel biomarker tool for the diagnosis of tuberculosis. *Biosciences Biotechnology Research Asia* 2016;**13**:551–6.
41. Zhou A, Ni J, Xu Z, Wang Y, Lu S, Sha W, et al. Application of 1H NMR spectroscopy-based metabolomics to sera of tuberculosis patients. *J Proteome Res* 2013;**12**:4642–9.
42. Mahapatra S, Hess AM, Johnson JL, Eisenach KD, DeGroot MA, Gitta P, et al. A metabolic biosignature of early response to anti-tuberculosis treatment. *BMC Infect Dis* 2014;**14**:1.
43. Frediani JK, Jones DP, Tukvadze N, Uppal K, Sanikidze E, Kipiani M, et al. Plasma metabolomics in human pulmonary tuberculosis disease: a pilot study. *PLoS One* 2014;**9**:e108854.