HHS Public Access

Author manuscript

Nat Rev Immunol. Author manuscript; available in PMC 2019 May 01.

Published in final edited form as:

Nat Rev Immunol. 2018 May ; 18(5): 289–290. doi:10.1038/nri.2018.26.

Discovering transcriptional signatures of disease for diagnosis versus mechanism

Julie G. Burel and **Bjoern Peters***

La Jolla Institute for Allergy and Immunology, Department of Vaccine Discovery, La Jolla, CA, USA.

Abstract

Transcriptional signatures of disease can be used for diagnosis or to gain insight into disease mechanisms. This Comment article discusses the different sets of criteria that should be considered for the optimal design of investigations addressing these two purposes, using examples from the study of tuberculosis.

Transcriptional signatures of diseases have the potential to be used for two distinct purposes: to help diagnose the disease status and/or prognosis of a given patient and thus guide treatment decisions; or to gain insights into disease mechanisms and thus guide the design of therapeutic or prophylactic interventions. Here, we discuss factors to consider for the design and interpretation of these two types of transcriptomic study, which we have found to not be obvious for researchers moving into the field as well as for more senior scientists focusing solely on one of the two applications. For mechanistic insights, it is necessary to determine the biological meaning of observed changes, such as which cells are responsible for the signature observed. By contrast, it is not necessary to understand the underlying mechanism of an observed signature for it to be used as a diagnostic tool. It is only important that the signature has high discriminatory power and is easy to obtain in a clinical setting.

We illustrate our recommendations using the example of infection with *Mycobacterium tuberculosis*. Most individuals control the pathogen at the asymptomatic stage of latent infection, but 10% of individuals progress to active tuberculosis (TB), which has high levels of morbidity and mortality. Host transcriptomics could improve current diagnostic tools to better characterize the spectrum of TB disease (in particular, to identify individuals with latent infection who are at risk of developing active disease) and could provide a better mechanistic understanding of TB pathology to develop novel therapeutic interventions.

Studies that aim to discover transcriptional signatures of human diseases share four key steps: enrolling cohorts of patients, sample collection, data generation and data analysis. As

* bpeters@lji.org.

Author contributions

J.G.B. and B.P. researched data for the article, made substantial contributions to discussions of the content, wrote the article and reviewed and/or edited the manuscript before submission.

Competing interests

The authors declare no competing interests.

we discuss below, how each of these steps is implemented affects the usefulness of the generated transcriptional signature.

Enrolling cohorts

Diagnostic studies should include not only healthy individuals in the control cohort but also individuals with other diseases to ensure the specificity of gene signatures identified. For example, a seminal study of TB[1] showed the feasibility of distinguishing subjects with active disease from uninfected individuals based on whole-blood transcriptomics. However, follow−up studies revealed that the same transcriptional signature is found in sarcoidosis[2], a disease that also results in lung granulomas. Cofactors such as other diseases prevalent in the target population are particularly important to consider for diagnostic studies. Specifically, transcriptomic signatures for active TB derived from HIV-negative cohorts did not reproduce well in HIV-positive cohorts[3]. Finally, diversifying the geographic location and ethnicity of disease cohorts is crucial to ensure that diagnostic signatures are relevant for the entire target population. Large cohorts are necessary to cover a multitude of disease states and cofactors to develop a robust diagnostic signature.

By contrast, mechanistic studies that aim to identify targets for intervention can be carried out in restricted, homogenous cohorts comparing individuals with and without TB, while excluding individuals with cofactors that could obscure underlying mechanisms. Smaller cohorts in turn enable carrying out more expensive experiments for each individual studied.

Sample collection

In diagnostic studies, ease of sample acquisition is crucial. Blood is readily accessible and — in contrast to urine or saliva — is rich in cells and RNA. Conversely, for mechanistic studies, access to disease-relevant tissues is more important, even if they are hard to obtain. Small sample sizes can still generate mechanistic insights. For example, a study of lung granulomas from three patients with TB identified several immune-related pathways that are dysregulated between patients and controls[4].

Diagnostic tests require a robust workflow with minimal processing steps. Thus, unfractionated samples, such as whole blood, are preferred. By contrast, mechanistic studies should identify the cell types responsible for transcriptional signatures, which can be achieved by studying cell subsets isolated by fluorescence-activated cell sorting (FACS). For example, we have recently discovered novel markers of latent TB by comparing the transcriptome of sorted memory $CD4^+$ T cells in infected versus non-infected subjects[5], which provides mechanistic insights into how these cells control the infection, but is not a practical approach for a diagnostic tool.

Data generation

An ideal diagnostic test is straightforward and cheap, to ensure technical reproducibility and applicability to low-income areas. This is achieved by, for example, PCR assays for a limited panel of genes. However, diagnostic studies need to first identify discriminatory gene candidates based on unbiased analyses, and then proceed to validation at the individual gene

level[6,7]. For mechanistic studies, the objective is typically to generate as many data as possible, particularly when samples are hard to obtain. In this case, whole-transcriptome analyses are preferred, and sophisticated techniques, such as single-cell RNA-sequencing, can provide comprehensive insights into the signatures associated with disease, but cannot realistically be used as a diagnostic tool in a low-resource setting.

In mechanistic studies, generating transcriptomic data on in vitro-stimulated cells has advantages over ex vivo analysis. Antigen stimulation activates cells responsible for combating TB, thereby increasing the likelihood of discovering signatures with disease relevance[8,9]. Although direct ex vivo analysis is usually preferred for diagnostic studies for simplicity, antigen-specific stimulation can remove convoluting signals (such as co−infections and non-disease-specific inflammatory processes) by focusing on disease-relevant antigens, such as ESAT6 and CFP10 in the interferon−γ release assay for diagnosing latent TB.

Data analysis

Diagnostic studies aim to identify genes whose transcription discriminates between disease states. The preferred approach for diagnostic gene selection is differential expression analysis to identify genes with high discriminatory power. Machine learning methods can identify concise sets of classifier genes, which translate to simple assays that are well suited for clinical assessment. Examples of promising diagnostic tools to discriminate between active and latent TB include a three-gene signature[10], and *BATF2* gene expression[6] in whole blood. More recently, a 16−gene signature in whole blood was reported to predict risk of active disease in individuals with latent TB[7]. These signatures may include genes that are dysregulated far downstream of the initial causal event and thus are poor targets for therapeutic intervention.

Conversely, mechanistic studies provide biological interpretation of the disease signature, including the underlying molecular mechanisms and their causal relationships. Knowledge of these relationships can guide the development of therapeutics to intervene with upstream molecular targets. Differential gene expression analysis should be carried out in a less stringent manner than for diagnostic studies, because small changes in the expression of regulatory genes can have a large effect on cell states. Modular analysis can be used to identify co−expression patterns and gene clusters associated with disease that have regulatory genes at their centre, such as the association of IL−32 with host defence mechanisms in TB[9]. Finally, to identify upstream regulators that are the most promising targets for therapeutic intervention, transcriptomic signatures of disease should consider gene dysregulations as a network to tease apart causality and distinguish primary versus secondary effects.

Future directions

Host transcriptomics is an extremely useful tool to tackle diagnostic and mechanistic challenges associated with diseases such as TB. Studies have identified candidate genes for diagnostic and prognostic tests, and have also improved our knowledge of TB−specific

immune mechanisms, which provides potential new areas for intervention. We have identified factors for the optimal design and analysis of future transcriptomic studies to address the outstanding needs in the TB field and beyond. Future diagnostic studies should identify gene expression signatures that reliably distinguish TB from other diseases, and that predict which patients with latent infection are at risk of progressing to active TB. For mechanistic studies, the identification of disease-relevant cell subsets and network analysis should facilitate the identification of key dysregulated molecules as promising candidates for therapeutic intervention.

Acknowledgements

The authors thank M. Babor, M. Pomaznoy, N. Khan, A. Sette and C. S. Lindestam Arlehamn from the Department of Vaccine Discovery of La Jolla Institute for Allergy and Immunology for their contribution to the literature review on which this work is based.

References

1. Berry MP et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature 466, 973–977 (2010). [PubMed: 20725040]
2. Maertzdorf J et al. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. Proc. Natl Acad. Sci. USA 109, 7853–7858 (2012). [PubMed: 22547807]
3. Kaforou M et al. Detection of tuberculosis in HIV-infected and-uninfected African adults using whole blood RNA expression signatures: a case-control study. PLoS Med 10, e1001538 (2013). [PubMed: 24167453]
4. Kim MJ et al. Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism. EMBO Mol. Med 2, 258–274 (2010). [PubMed: 20597103]
5. Burel JG et al. Transcriptomic analysis of $CD4^+$ T cells reveals novel immune signatures of latent tuberculosis. J. Immunol 10.4049/jimmunol.1800118 (2018).
6. Roe JK et al. Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. JCI Insight 1, e87238 (2016). [PubMed: 27734027]
7. Zak DE et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet 387, 2312–2322 (2016). [PubMed: 27017310]
8. Cliff JM et al. Excessive cytolytic responses predict tuberculosis relapse after apparently successful treatment. J. Infect. Dis 213, 485–495 (2016). [PubMed: 26351358]
9. Montoya D et al. IL−32 is a molecular marker of a host defense network in human tuberculosis. Sci. Transl Med 6, 250ra114 (2014).
10. Sweeney TE et al. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. Lancet Respir. Med 4, 213–224 (2016). [PubMed: 26907218]