A contained *Mycobacterium tuberculosis* mouse infection model predicts active disease and containment in humans

Fergal J Duffy¹, Gregory S. Olson¹, Elizabeth S. Gold¹, Ana Jahn¹, Alan Aderem¹, John Aitchison¹, Alissa C. Rothchild¹,², Alan H. Diercks¹, Johannes Nemeth¹,³

¹Seattle Children’s Research Institute, Seattle, WA, USA
²Current address: Department of Veterinary and Animal Sciences, University of Massachusetts Amherst, Amherst, MA, USA
³Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Switzerland.

*Corresponding author, fergal.duffy@seattlechildrens.org

**Brief Summary**

We previously described the mouse contained TB model, which protects mice against aerosol challenge. Here, we show that blood RNA signatures derived from this model correlate with disease and TB containment in multiple human cohorts.
Abstract

Previous studies have identified whole-blood transcriptional risk and disease signatures for Tuberculosis (TB); however, several lines of evidence suggest that these signatures primarily reflect bacterial burden, which increases prior to symptomatic disease. We found that the peripheral blood transcriptome of mice with contained *Mycobacterium tuberculosis* infection (CMTB) has striking similarities to that of humans with active TB and that a signature derived from these mice predicts human disease with comparable accuracy to signatures derived directly from humans. A set of genes associated with immune defense are upregulated in CMTB mice but not in humans with active TB suggesting that their upregulation is associated with bacterial containment. A signature comprised of these genes predicts both protection from TB disease and successful treatment at early time points where current signatures are not predictive. These results suggest that detailed study of the CMTB mouse model may enable identification of biomarkers for human TB.

**Keywords:** Tuberculosis; signature of risk; cross-species; blood transcription
Introduction

Identification of biomarkers to diagnose stages of Mycobacterium tuberculosis (MTB) infection and disease (active tuberculosis, TB) remains an important clinical and public health goal. Pioneering work has shown that expression of interferon-inducible genes is strongly increased in the peripheral blood of patients with active TB [1]. Additional studies have derived signatures correlated with risk of progression to active disease for latently infected (LTB+) individuals in endemic areas and household contacts of active TB cases [1-7]. Multiple lines of evidence suggest that these biomarkers reflect sub-clinical immune responses correlated with bacterial burden rather than identify individuals pre-disposed to ineffective immune control of TB [8-11] and thus, correlates of protective immunity to TB remain elusive. All of these signatures have been derived directly from clinical or observational studies of human populations. While human studies are essential for proving a biomarker’s clinical utility, they are time-consuming and expensive. In addition, they offer very limited opportunities to experimentally interrogate the target population which is a significant challenge to deciphering the immune mechanisms that underlie a biomarker.

Several recent studies have identified strong similarities between the peripheral blood transcriptomes of MTB-infected humans and those of MTB-infected mice [8-10]. In particular, it has been demonstrated that whole-blood transcriptional signatures derived from mice infected by aerosol at the “conventional dose” (~50-100 bacterial colony-forming units or CFU) or an “ultra-low dose” (~1-3 CFU) can predict the outcome of MTB-infection in humans as accurately as signatures derived directly from human studies [10]. These results suggest that many of the immune mechanisms that are activated by MTB-infection are shared between mice and humans. However, it has long been known that the course of TB disease differs significantly between these species: In humans, the overwhelming majority (90%) of MTB-infections do not progress to clinically diagnosable disease [3, 4, 12, 13]. Non-progressors either clear the infection or successfully contain and control replication-competent bacteria for their lifetimes, often in lymphoid tissue, while remaining asymptomatic [14]. In addition, it has been well documented that humans who resist progression to clinical disease after exposure to MTB have increased protection against subsequent MTB exposure [15, 16]. In contrast, if detectable infection is established in a mouse the animal will ultimately fail to control the bacteria [17]. It is therefore impossible to study the immune impact of contained MTB infection on the peripheral blood transcriptome in mice following aerosol challenge because the infection is never contained. This is a significant limitation of using the mouse model for TB biomarker development as well as for mechanistic studies.

The recent development of a “contained MTB” (CMTB) infection model [18, 19] has overcome many of these limitations. In mice, intradermal inoculation of the ear with ~10,000 CFU of MTB establishes an infection in the draining lymph node that is asymptomatic, stable for at least 1 year, and excluded from the lung [18, 19]. Following the establishment of CMTB, mice are strongly protected against subsequent aerosol challenge, mimicking the well-established protective effect of prior MTB-infection in humans [19]. We hypothesized that CMTB might be a useful model for the development of biomarkers that predict outcomes for asymptomatic MTB infection in humans. In concordance with studies of the ULD model [10], we found that peripheral blood transcriptional responses in CMTB mice are similar to those measured in humans with active TB and are predictive of risk for progression to active disease. We also identified a set of genes strongly upregulated in CMTB mice whose expression is not elevated in humans with active TB. When measured more than
18 months prior to diagnosis, elevated expression of these genes was associated with lower risk of progression to disease. In contrast, the recently developed Correlate of Risk score (COR) [3] is not predictive of progression to active disease this early. Furthermore, when measured at the time of diagnosis, elevated expression of these genes correlated with better treatment outcome.

Methods

Establishment of contained infection in mouse

Intradermal infections to establish CMTB were performed as described previously [18] with the following modifications: 10,000 CFU of MTB (H37Rv) in logarithmic phase growth in 10 μL PBS were injected intradermally using a 10 μL Hamilton syringe into mice anaesthetized with Ketamine.

CMTB whole blood RNAseq

RNA isolation was performed using TRIzol (Invitrogen) with two sequential chloroform extractions, Glycoblue carrier (Thermo Fisher), isopropanol precipitation, and washes with 75% ethanol. RNA was quantified with the Bioanalyzer RNA 6000 Pico Kit (Agilent). cDNA libraries for alveolar macrophages were constructed and amplified using the SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian (Clontech) per the manufacturer’s instructions. cDNA for whole-blood was prepared using the TruSeq Stranded mRNA kit (Illumina, CA, USA). Libraries were amplified and then sequenced on an Illumina NextSeq (2 x 75 base pair, paired-end reads). Stranded paired-end reads of length 76 were preprocessed: For the Pico Input prep, the first three nucleotides of R2 (v2 kit) were removed as described in the SMARTer Stranded Total RNA-Seq Kit - Pico Input Mammalian User Manual (v2: 063017); Read ends consisting of 50 or more of the same nucleotide were removed. The remaining read pairs were aligned to the mouse genome (mm10) + MTB H37Rv genome using the GSNAP aligner [20] (v. 2016-08-24) allowing for novel splicing. Concordantly mapping read pairs (average 10-20 million / sample) that aligned uniquely were assigned to exons using the subRead (v.1.4.6.p4) [21] program and gene definitions from Ensembl Mus_Musculus GRCm38.78 coding and noncoding genes. Differential expression was calculated using the edgeR [22] package from bioconductor.org [23]. False discovery rate was computed with the Benjamini-Hochberg algorithm. Raw and processed data are deposited in GEO (GSE126355).

Preparation of human TB risk and disease cohorts

The following datasets containing whole-blood transcriptomic measurements of human TB risk, disease, and treatment cohorts were downloaded from the NCBI GEO: the human TB disease cohorts GSE107991, GSE107992 [1, 24] and GSE37250 [2]; the ACS and GC6-74 TB risk cohorts GSE79362 [3], GSE94438 [4]; and the Catalysis TB treatment cohort GSE89403 [25]. RNA-seq data was normalized and differential expression calculated using the edgeR package. Similarly, microarray data was background subtracted, quantile normalized and log2 transformed and differential expression calculated using the R limma package. Signatures were translated across species by identification of
gene homologs using the Mouse Genome Informatics database
[http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt].

Calculation of signature scores

CMTB-DS genes were selected as genes that were consistently highly expressed 28 and 42 days post CMTB establishment (FDR < 0.001 and | logFC | > 1.5) that were also detected as expressed in human datasets (Table 1). For each dataset, CMTB-derived signature scores were calculated as the mean per-sample expression of all signature genes.

The ACS-CoR signature scores were calculated using the published pairwise SVM parameters for each junction pair [3]. The signature was adapted for prediction on gene read counts (rather than junction read counts) by substituting the corresponding gene counts for junction counts. Additionally, instead of using a voting threshold of 0.5 to classify pair scores into votes, and calculating the average of pair votes, the average of pair scores was used as the output. This enabled calculation of signature scores on gene-wise RNAseq and microarray studies without reparameterization to account for differing sequencing depths. For RNAseq, signature predictions were made on gene counts, and for microarrays, probe intensity values were used. ROC AUCs and accompanying 95% CIs were calculated using the R pROC package [26].

Results

Contained Mtb-infection in mice induces sustained blood transcriptome changes

To examine the effect of a contained mycobacterial infection on the whole-blood transcriptome in mice, we established contained MTB infections (CMTB) as previously described [19]. We performed RNAseq analysis of blood obtained immediately prior to the establishment of CMTB (day 0), and 10, 28, and 42 days afterwards. 68 genes were strongly differentially expressed for at least one time point compared to day 0 ( | log2(fold-change) | > 1.5, FDR < 0.001) (Figure 1 A-C). These genes consisted of two broad groups: 1) a set of 22 genes that were transiently upregulated at day 10 and 2) a set of 46 genes that were most strongly upregulated at day 28 and remained highly upregulated at day 42 (Figure 1 D). We took the latter set of 46 genes to represent the long-term whole-blood transcriptome associated with low-level, contained MTB-infection in mice.

A gene expression signature derived from CMTB in mice predicts active TB in humans, correlates with treatment response, and discriminates between active and latent TB.

We hypothesized that many of the systemic immune processes that are activated by contained MTB-infection in mice would be similar in humans harboring replicating bacteria and lead to similar transcriptional responses in the peripheral blood. Currently, there are several publicly available datasets containing whole-blood transcriptomic measurements of subjects with active TB (Table 1) (the Berry and Kaforou cohorts GSE107991, GSE107992 [1, 24]), as well as datasets comprising
individuals at risk of progression to TB (the ACS and GC6-74 TB risk cohorts GSE79363 [3], GSE94438 [4]), and responding to TB treatment (the Catalysis cohort GSE89403 [25]). Therefore, we constructed a gene-expression signature from the 46 “long-term” genes and tested its ability to predict TB disease in humans. We determined the human homologs of each of the 46 “long-term” genes and retained those that were consistently expressed at a detectable level in all of the human datasets (23 genes in total, ‘Consistently Detected’ genes in Figure 1 D). The signature score for each subject, which we termed the “CMTB disease score (CMTB-DS)” was defined as the average expression of these 23 genes.

To test the ability of the CMTB-DS to identify TB disease in humans, we analyzed a set of 3 whole-blood transcriptional profiles from patients with active pulmonary TB (ptb), latent TB, and healthy controls [1, 24] (Table 1). Despite being derived from a mouse model, the CMTB-DS was highly effective (AUC: 0.92 [95% CI: 0.86 – 0.97]) at discriminating individuals with active disease from those with latent TB and performed equivalently to the ACS-COR (AUC: 0.93: [95% CI: 0.88 – 0.99]) [3] (Figure 2 A). We also tested the predictive power of the CMTB-DS using the Kaforou [2] dataset consisting of whole-blood expression profiles from both HIV+ and HIV- individuals with active or latent TB in Malawi and South Africa, (Table 1). Again, the CMTB-DS performed comparably to the ACS-COR in discriminating active disease from latent TB in both HIV+ and HIV- individuals (Figure 2 B,C).

One of the most significant recent advances in TB biomarker development was the demonstration that the ACS-CoR score can predict the risk of progression to active TB up to 18 months prior to clinical diagnosis. Given that the CMTB-DS performed equivalently to the ACS-CoR at identifying active disease, we tested whether it also predicted risk of progression. CMTB-DS is significantly predictive of TB progression in the ACS cohort (Table 1) (AUC: 0.65 [95% CI: 0.58 – 0.71]), although with reduced performance vs ACS-CoR (AUC 0.8 [95% CI: 0.74-0.85], Figure 2 D, Figure S1). This reduced performance vs ACS-CoR is not surprising as ACS-CoR was originally developed and trained on these samples. Importantly, we found that this CMTB-DS signature had nearly equivalent predictive power to the ACS-COR for predicting progression to active disease in the independent GC6-74 household contact risk cohort (Table 1) (CMTB-DS AUC: 0.67 [95% CI 0.6 - 0.74] vs ACS-CoR AUC: 0.71 [95% CI 0.64-0.78]), (Figure 2 D, Figure S1).

**A portion of the CMTB disease signature has an opposite expression pattern in humans.**

The CMTB-DS signature was strongly associated with TB disease in humans. However, in mice, CMTB is asymptomatic and does not reduce lifespan [18, 19]. Moreover, CMTB mice are protected against subsequent MTB aerosol challenge [19]. Therefore, we hypothesized that in addition to a component that reflects the presence of replicating bacteria, the whole-blood transcriptome of CMTB mice contains a component that reflects immune mechanisms associated with successful bacterial containment.

We sought to identify genes that differed in their expression pattern between CMTB in mice and active TB in humans and then set out to assess their association with improved TB outcomes in independent human TB risk cohorts. The expression profiles of the 23 mouse CMTB-DS genes were compared with the expression profiles of their human homologs in healthy individuals or individuals...
with latent or active TB (Figure 3). We labelled genes significantly upregulated in CMTB mice vs naive mice and in human TB vs LTB as “concordant”. Genes significantly upregulated in CMTB vs naive that also showed reduced expression in human TB vs LTB were labelled as “opposite”, and all other genes not fitting either of these patterns as “ambiguous”. Thus, the CMTB-DS genes were divided into groups whose expression profiles are either concordant (12/23 genes), opposite (5/23 genes), or ambiguous (6/23 genes) between CMTB in mice at TB disease in humans. None of the five “opposite” genes (CCR5, TPPP3, ZFYVE9, SH2D1B and KLRG1) are included in the ACS-CoR Score, which primarily consists of type I/II interferon responsive genes that are upregulated in active TB. Modular analysis showed that, while “concordant” genes were primarily annotated as members of inflammation, interferon, monocyte and DC modules, none of the “opposite” genes fell into these modules and were instead associated with Cytotoxic/NK Cell and chemokine processes (Table S1). We hypothesized that elevated expression of these five “opposite” genes in CMTB mice was correlated with a protective immune response acting to contain infecting bacteria and sought to determine whether a signature defined by their average expression (“CMTB containment” or CMTB-CT, see Methods) could predict protection from TB disease in humans.

The CMTB-CT signature is associated with long-term disease containment and effective treatment in longitudinal human cohorts

While it is not possible to experimentally test protective immunity in humans, longitudinal studies of high-risk cohorts allowed us to investigate CMTB-CT score variation over time in humans exposed to TB. The performance of the ACS-CoR signature has been shown to decline substantially as the time to diagnosis increases and it is unable to discriminate TB progressors from non-progressors more than 18 months prior to TB diagnosis (Figure 4 A). In contrast, CMTB-CT scores are elevated in non-progressors compared with TB progressors >18 months from TB diagnosis in the ACS cohort (AUC 0.78 [95% CI: 0.59-0.97] Figure 4 A), and is significantly upregulated in TB non-progressors in the GC6-74 cohort (Figure S2). Furthermore, CMTB-CT scores and time to diagnosis in ACS and GC6-74 TB progressors are only weakly correlated (ACS r=0.2, p=0.09, Figure 4 B, Figure S2). In contrast, ACS-CoR scores were strongly correlated with time to disease diagnosis and showed a sharp increase beginning approximately 6 months prior to the onset of symptomatic disease (r=-0.52, p=1.1x10^{-6}, Figure 4 C). Thus, unlike ACS-CoR, CMTB-CT discriminates individuals likely to progress to TB from non-progressors independent of subclinical disease processes. This is consistent with CMTB-CT being associated with a more stable long-term phenotype, not a time-dependent response to TB infection.

Similarly, we hypothesized that TB treatment failure may be linked to reduced natural immunity to MTB and used the data from the Catalysis TB treatment cohort [25] (Table 1) to investigate the relationship between CMTB-CT scores and treatment failure. The CMTB-CT score, measured at the time of diagnosis and prior to treatment initiation, is elevated in individuals who successfully respond to a 24-week course of treatment compared to controls (Figure S3 A) and is significantly predictive of treatment outcome (Figure 4 D). In contrast, ACS-CoR is not predictive of treatment outcome at diagnosis, although it is significantly decreased in treatment successes 24 weeks after treatment initiation (Figure 4 D, Figure S3). Taken together this points to CMTB-CT
having predictive power independent of underlying bacterial burden in humans, thus consistent with natural immunity to TB.

Discussion

In this study, we have used a contained MTB-infection model in mice to develop a peripheral-blood gene-expression signature that is as predictive of human TB disease states as a signature derived directly from human data. The comparable performance of the CMTB-DS and ACS-CoR signatures supports our two core hypotheses: 1) The CMTB mouse model mimics at least a portion of the TB disease spectrum in humans; and 2) Current peripheral-blood transcriptional signatures that predict TB disease state in humans reflect the immune response to live bacteria. Our results are consistent with other recent studies that have uncovered strong similarities between the peripheral-blood transcriptional response to MTB-infection in mice and humans [8-10]. In particular, the CMTB-DS signature is comparable to the recently published ULD signature [10], being a mouse-derived TB phenotype applied to human cohorts. While none of the ULD signature genes overlap with CMTB-DS, both ULD and CMTB-DS signatures contain several genes associated with interferon and inflammation (Table S1). However, to our knowledge, this is the first study to specifically explore the predictive accuracy of a transcriptional signature derived from the CMTB model. This result also suggests that the increased experimental control afforded by a well-defined mouse model offsets the uncertainty that is introduced by translating between species.

It is generally assumed that MTB infection is a spectrum of conditions ranging from asymptomatic exposure and sterilizing immunity, to highly infectious pulmonary TB and clinical data suggest that TB-infected individuals may move back and forth across the disease continuum [27]. Pioneering studies in non-human primates using PET-CT scans have shown that within the same host, some lesions heal while other lesions increase in size [28]. Similar processes have been observed in humans [14, 29]. Thus, it appears that within the same host, progression and containment can take place simultaneously. The expression profile of whole blood is therefore likely to reflect a mixture of immune responses associated with both disease progression and containment, and any whole blood signature will reflect the composition and activation of multiple cell types circulating in the periphery.

Indeed, despite the fact that CMTB mice are protected against aerosol challenge, the CMTB-DS signature, which is composed of the most strongly differentially expressed genes in CMTB mice compared to controls, is elevated in humans with active TB. However, subsequent comparison with a human active TB cohort revealed a subset of genes upregulated in CMTB mice but not in human active TB. We combined these genes into a new signature, the CMTB-CT score that we could then test for association with improved outcomes in human TB risk and treatment cohorts. We confirmed that expression of the CMTB-CT gene signature is elevated in MTB-infected individuals capable of controlling the bacteria compared to individuals who will progress to active TB. Unlike ACS-CoR, CMTB-CT significantly discriminated TB progressors from controls when measured at more than 18 months prior to disease onset. Thus CMTB-CT appears to correlate with a long-term protection or containment phenotype that is not associated with subclinical responses to replicating TB or bacterial burden.
It is tempting to speculate that the CMTB-CT signature correlates with a natural disposition to TB control potentially supporting an effective treatment response. However, these results are limited by the relatively small number of TB progressors >18 months from diagnosis in the ACS cohort (6 total), and the 2-year maximum follow-up time. The Vukuzazi study [30], an ongoing, population-based prospective study of a TB-endemic region in KwaZulu Natal, South Africa, will follow enrolled individuals for 4 years and is powered to both validate and discover long-term correlates of TB immunity independent of short-term TB-disease associated responses.

The CMTB-CT signature includes two genes, KLRG1 and CCR5 that have been investigated in the context of MTB infection. Memory-like NKp46^+CD27^+KLRG1^+ NK cells have been observed to be expanded after BCG vaccination and protective against MTB challenge in a murine model [31]. CCR5 has been associated with control of bacterial migration to lymph nodes in mice [32]. The other three CMTB-CT genes (TPPP3, ZFYVE9 and SH2D1B) have not been investigated in the context of TB, however they are known to play roles in immune and proliferative signaling; ZFYVE9 in the TGFβ pathway [33], TPPP3 in the AKT-STAT3 pathway [34], and SH2D1B, as a regulator of signal transduction in antigen presenting cells [35]. This is in contrast with the ACS-CoR signature which principally consists of genes associated with type I/II interferon responses.

Conclusions

These findings, as well as other recent work [10] demonstrate that well-defined and experimentally accessible mouse MTB-infection models can be used to generate whole-blood transcriptional signatures that are translatable to humans. Because much smaller human cohorts are required to test a proposed predictive signature than are required to discover the same signature, this study highlights an opportunity for more rapid and cost-effective development of biomarkers for TB disease and disease progression.
References

1. Berry MP, Graham CM, McNab FW, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature 2010; 466:973-7.
2. Kaforou M, Wright VJ, Oni T, 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.
3. Zak DE, Penn-Nicholson A, Scriba TJ, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet 2016; 387:2312-22.
4. Suliman S, Thompson E, Sutherland J, et al. Four-gene Pan-African Blood Signature Predicts Progression to Tuberculosis. A systematic comparison of 16 signatures. PLoS Med 2019; 16:e1002786.
5. Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. PLoS Med 2019; 16:e1002786.
6. 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.
7. Duffy FJ, Thompson EG, Scriba TJ, Zak DE. Multinomial modelling of TB/HIV co-infection yields a robust predictive signature and generates hypotheses about the HIV+TB+ disease state. PLoS One 2019; 14:e0219322.
8. Ahmed M, Thirunavukkarasu S, Rosa BA, et al. Immune correlates of tuberculosis disease and risk translate across species. Sci Transl Med 2020; 12.
9. Moreira-Teixeira L, Tabone O, Graham CM, et al. Mouse transcriptome reveals potential signatures of protection and pathogenesis in human tuberculosis. Nat Immunol 2020; 21:464-76.
10. Plumlee CR, Duffy FJ, Gern BH, et al. Ultra-low Dose Aerosol Infection of Mice with Mycobacterium tuberculosis More Closely Models Human Tuberculosis. Cell Host & Microbe 2021; 29:1-15.
11. Scriba TJ, Penn-Nicholson A, Shankar S, et al. Sequential inflammatory processes define human progression from M. tuberculosis infection to tuberculosis disease. PLoS Pathog 2017; 13:e1006687.
12. Houben RM, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLoS Med 2016; 13:e1002152.
13. Behr MA, Edelstein PH, Ramakrishnan L. Is Mycobacterium tuberculosis infection life long? BMJ 2019; 367:i5770.
14. Cadena AM, Fortune SM, Flynn JL. Heterogeneity in tuberculosis. Nat Rev Immunol 2017; 17:691-702.
15. Andrews JR, Noubary F, Walensky RP, Cerda R, Losina E, Horsburgh CR. Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis. Clin Infect Dis 2012; 54:784-91.
16. Blaser N, Zahnd C, Hermans S, et al. Tuberculosis in Cape Town: An age-structured transmission model. Epidemics 2016; 14:54-61.
17. Beamer GL, Turner J. Murine models of susceptibility to tuberculosis. Arch Immunol Ther Exp (Warsz) 2005; 53:469-83.
18. Kupz A, Zedler U, Staber M, Kaufmann SH. A Mouse Model of Latent Tuberculosis Infection to Study Intervention Strategies to Prevent Reactivation. PLoS One 2016; 11:e0158849.
19. Nemeth J, Olson GS, Rothchild AC, et al. Contained Mycobacterium tuberculosis infection induces concomitant and heterologous protection. PLoS Pathog 2020; 16:e1008655.
20. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. Methods Mol Biol 2016; 1418:283-334.
21. Liao Y, Smyth GK. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res 2019; 47:e47.
22. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010; 26:139-40.
23. Reimers M, Carey VJ. Bioconductor: an open source framework for bioinformatics and computational biology. Methods Enzymol 2006; 411:119-34.
24. Singhania A, Verma R, Graham CM, et al. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. Nat Commun 2018; 9:2308.
25. Thompson EG, Du Y, Malherbe ST, et al. Host blood RNA signatures predict the outcome of tuberculosis treatment. Tuberculosis (Edinb) 2017; 107:48-58.
26. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011; 12:77.
27. Barry CE, 3rd, Boshoff HI, Dartois V, et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nat Rev Microbiol 2009; 7:845-55.
28. Lin PL, Ford CB, Coleman MT, et al. Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. Nat Med 2014; 20:75-9.
29. Malherbe ST, Shenai S, Ronacher K, et al. Persisting positron emission tomography lesion activity and Mycobacterium tuberculosis mRNA after tuberculosis cure. Nat Med 2016; 22:1094-100.
30. Ngwenya N, Luthuli M, Gunda R, et al. Participant understanding of informed consent in a multidisease community-based health screening and biobank platform in rural South Africa. Int Health 2020; 12:560-6.
31. Venkatasubramanian S, Cheekatla S, Paidipally P, et al. IL-21-dependent expansion of memory-like NK cells enhances protective immune responses against Mycobacterium tuberculosis. Mucosal Immunol 2017; 10:1031-42.
32. Algood HM, Flynn JL. CCR5-deficient mice control Mycobacterium tuberculosis infection despite increased pulmonary lymphocytic infiltration. J Immunol 2004; 173:3287-96.
33. Lin HK, Bergmann S, Pandolfi PP. Cytoplasmic PML function in TGF-beta signalling. Nature 2004; 431:205-11.
34. Zhou W, Li J, Wang X, Hu R. Stable knockdown of TPPP3 by RNA interference in Lewis lung carcinoma cell inhibits tumor growth and metastasis. Mol Cell Biochem 2010; 343:231-8.
35. Morra M, Lu J, Poy F, et al. Structural basis for the interaction of the free SH2 domain EAT-2 with SLAM receptors in hematopoietic cells. EMBO J 2001; 20:5840-52.
### Tables

| Cohort Name | Description | Samples Analysed |
|-------------|-------------|------------------|
| **Berry [1, 24]** | Active TB patients vs latent and healthy control individuals | 37 TB  
64 LTB  
12 Healthy Controls |
| **Kaforou [2]** | Active TB patients vs latent TB, including HIV+ and HIV- individuals | Malawi:  
HIV- TB: 51, HIV- LTB: 35  
HIV+ TB: 51, HIV- LTB: 36  
South Africa:  
HIV- TB: 46, HIV- LTB: 48  
HIV+ TB: 47, HIV+ LTB 48 |
| **ACS [3]** | Longitudinal study (2 year follow up) of LTB+ adolescents, some of whom progressed to TB | 94 progressor samples  
245 control samples (non-progressors) |
| **GC6-74 [4]** | Longitudinal study (2 year follow up) of household contacts of a TB index case some of whom progressed to TB | 82 progressor samples  
282 control samples (non-progressors) |
| **Catalysis [25]** | Longitudinal study (24 weeks) of TB treatment responses | 78 definite cures  
7 not cured  
4 timepoints each  
[DX, day 7, week 4, week 24] |

Table 1: Previously published RNAseq whole blood profiles used in this study
Figures

**Figure 1**: Establishment of mouse contained infection leads to alterations in blood transcriptional state. A-C Volcano plots of differentially expressed genes comparing post CMTB initiation. (A: d10, B: d28, C: d42) transcriptional states to pre-c.i. (d0). D. Heatmap of significant differential expression changes shown in A-C. Genes showing high expression at d28-42 and consistently detected in multiple human datasets are indicated by the ‘Consistently Detected’ annotation bar above the heatmap.

**Figure 2**: The CMTB-DS signature predicts human TB outcomes. A. ROC curves for the CMTB-DS and ACS-CoR signatures discriminating active from latent TB in the Berry cohort A. and the Kaforou cohort B,C. ROC curves for the CMTB-DS and ACS-CoR signatures TB progressors from non-progressors in the ACS and CG6 cohorts D,E. ROC AUCs and accompanying 95% CIs are shown in the figure legend.

**Figure 3**: CMTB-CT signature genes. Heatmap shows hierarchical clustering of log2 transformed fold changes for c.i. signature genes (columns) for each comparison (rows). Positive fold changes shown in red, negative in blue. Genes are annotated above as ‘concordant, ‘opposite or ‘ambiguous’ based on the consistency of differential expression between CMTB and human TB vs LTB.

**Figure 4**: The CMTB-CT signature is increased in non-progressors in a time-independent manner. A. ROC curves for the CMTB-CT signature discriminating TB non-progressors from TB progressors >18 months from TB diagnosis in the ACS cohort. ROC curves were calculated assuming TB non-progressor scores > progressors for CMTB-CT, and TB progressor scores > non-progressor scores for ACS-CoR. B,C. Signature scores by time to TB for ACS TB progressors for each CMTB-CT and ACS-CoR. Points connected by lines represent multiple samples from the same individual progressor, while individual points represent progressors from whom only a single sample was collected. D. ROC curves for the CMTB-CT and ACS-CoR signature discriminating treatment success (definite cures) from failures (not cured) at the time of TB diagnosis, prior to treatment initiation in the Catalysis TB treatment cohort.
Conflict of Interest
The authors have declared that no conflict of interest exists.

Funding
Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases (Grant number U19AI135976) and the National Institutes of Health (P41 GM109824, R01AI032972 and U19AI35976). JN was supported by Swiss National Science Foundation (SNSF, http://www.snf.ch/en/Pages/default.aspx) grant #P300PB_164742. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Corresponding Authors
Please direct communications to Johannes Nemeth (jo.l.nemeth@gmail.com) or Fergal Duffy (fergal.duffy@seattlechildrens.org)
Figure 1
Figure 2

A. Signatures on Berry Cohort
   Active TB > Latent TB

B. CMTB–DS on Kaforou Cohort
   Active TB > LTBI

C. ACS–CoR on Kaforou Cohort
   Active TB > LTBI

D. Signatures on ACS and GC6–74 Cohorts
   Controls < Progressors
Figure 4

A. ACS >= 18 Mo to TB

B. CMTB-CT signature on ACS progressors

C. ACS-CoR signature on ACS progressors

D. Catalysis cohort, at DX
Definite cure > not cured