Ver 2.0 last edited by RJW on 31 AUGUST 2021

Transcriptomic characterization of tuberculous sputum reveals a host Warburg effect and microbial cholesterol catabolism

Rachel PJ Lai[1,3], Teresa Cortes[5], Suzaan Marais[2], Neesha Rockwood[2,3,4], Melissa L Burke[1,7], Acely Garza-Garcia[1], Stuart Horswell[1], Abdul K Sesay[1,8], Anne O'Garra[1,9], Douglas B Young[1,5,10] and Robert J Wilkinson[1,2,3,10,11]

1. The Francis Crick Institute, London NW1 1AT, United Kingdom
2. Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Disease and Molecular Medicine and Department of Medicine, University of Cape Town, Republic of South Africa
3. Department of Infectious Disease, Imperial College London, W12 0NN, United Kingdom
4. Present address Department of Microbiology, University of Colombo, Colombo 8, Sir Lanka
5. MRC Centre for Molecular Bacteriology and Infection, Imperial College London, SW7 2AZ, United Kingdom
6. Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, WC1E 7HT, United Kingdom
7. Present address: Australian BioCommons, Australia; Research Computing Centre, The University of Queensland, Queensland, Australia; and Queensland Cyber Infrastructure Foundation, Queensland, Australia
8. Present address: Genomics Core, MRC Unit The Gambia at LSHTM, Serrekunda, The Gambia
9. National Heart & Lung Institute, Imperial College London, London W2 1PG, United Kingdom
10. These authors contributed equally and jointly supervised the work
11. Correspondence and requests for materials should be addressed to Robert J Wilkinson (r.j.wilkinson@imperial.ac.uk). Tel: +27 214066084. Fax: +27 214066796

Abstract

The crucial transmission phase of tuberculosis (TB) relies on infectious sputum yet cannot easily be modelled. We applied one-step RNA-Sequencing to sputum from infectious TB patients to investigate the host and microbial environments underlying transmission of *Mycobacterium tuberculosis* (*Mtb*). In such TB sputa, compared to non-TB controls, transcriptional upregulation of inflammatory responses including an interferon-driven proinflammatory response and a metabolic shift towards glycolysis was observed in the host. Amongst all bacterial sequences in the sputum, approximately 1.5% originated from *Mtb* and its transcript abundance was lower in HIV-1 coinfected patients. Commensal bacterial abundance was reduced in the presence of *Mtb* infection. Direct alignment to the genomes of the predominant microbiota species also reveals differential adaptation, whereby firmicutes (e.g. *Streptococci*) displayed a non-replicating phenotype with reduced transcription of ribosomal proteins and reduced activities of ATP synthases, while *Neisseria* and *Prevotella* were less affected by comparison. The transcriptome of sputum *Mtb* more closely resembled aerobic replication and shared similarity in carbon metabolism to *in vitro* and *in vivo* models with significantly upregulation of genes associated with cholesterol metabolism and the downstream propionate detoxification pathways. In addition, and counter to previous reports on intracellular *Mtb* infection *in vitro*, *Mtb* in sputum was zinc, but not iron, deprived and the phoP loci were also significantly downregulated, suggesting the pathogen is likely to be extracellular in location.

Importance

Although a few studies have described the microbiome composition of TB sputa based on 16S ribosomal DNA, these studies did not compare to non-TB samples and the nature of the method does not allow any functional inference. This is the first study to apply such technology on clinical specimens and obtained functional transcriptional data on all three aspects simultaneously. We anticipate that an improved understanding on the biological interactions in the respiratory tract may also allow novel interventions, such as those involving microbiome manipulation or inhibitor targeting disease-specific metabolic pathways.

Keywords: Host-pathogen, RNA-Seq, *Mycobacterium tuberculosis*, Warburg effect, cholesterol

Introduction

Concerted efforts over the last two decades have widened availability of therapy for tuberculosis (TB). While this has saved millions of lives, the incidence of disease has declined by only 1.5% annually (1). The host-pathogen interaction in TB is complex, thus hindering the development of diagnostic tests and effective new treatments. Studies on TB rely heavily on *in vitro* or *in vivo* experimental models, or blood from TB patients, as lung sampling is invasive. While these approaches provide insights into TB immune responses and the development of tuberculous lesions at a cellular and molecular level, the events following bacterial release from liquefied lung cavities into the airways remain poorly understood.

As TB is spread by aerosol generated mainly through coughing, understanding the physiological state of *Mycobacterium tuberculosis* (*Mtb*) and its interaction with the host in the nasopharyngeal environment may bring insights on new treatment or preventive therapy strategies. Sputum is routinely collected for TB diagnosis and has been proposed as a surrogate for bronchoalveolar lavage for monitoring the transcriptional profiles of *Mtb* in patients (2). While several studies in the past have characterised the transcriptomes of sputum *Mtb* using microarray and/or targeted quantitative PCR (qPCR), they lacked simultaneous profiling of the host response. We reasoned that a comprehensive RNA sequence-based analysis that yields dual host-pathogen transcriptomes would provide important insight to improve understanding of the biology of *Mtb* transmission and pathogenesis. Technical difficulties and the overwhelming eukaryotic content have limited conventional sequencing approaches either to the host or to a pathogen that has been physically separated or independently enriched, but dual RNA-Seq allows comprehensive and simultaneous survey of gene expression of both the host and the pathogen in one step. To date, there has been increasing success in dual RNA-Seq where the technology was successfully applied to profile gene expression of *Salmonella enterica* in infected HeLa cells (3), *Haemophilus influenzae* colonized primary mucosal epithelium (4) and murine Peyer's patch infected with *Yersinia psedotuberculosis* (5). Non-one-step dual RNA-Seq has also been used to study *Mycobacterium paratuberculosis* and *Mycobacterium bovis* Bacillus Calmette-

Guerin (BCG) infected cells *in vitro* but with limited success despite separate microbial enrichment (6, 7). Most recently, dual RNA-Seq on *Mtb*-infected mice indicated that alveolar and interstitial macrophages utilised different mechanisms to sustain or restrict intracellular *Mtb* growth (8). In this study, we applied one-step dual RNA-Seq to sputa collected directly from patients with and without active TB to survey the global transcription profiles of the host and *Mtb*. Transcriptional signature of TB-infected host displayed characteristic of the Warburg effect, while cholesterol catabolism and zinc-deprivation were identified in sputum *Mtb*.

Results

Dual RNA-Seq and the host transcriptome

RNA was extracted from 17 sputum samples from South African patients with untreated active TB (9 HIV-uninfected and 8 HIV-infected, referred to as TB-only and TB-HIV, respectively) and 9 samples from persons with respiratory symptoms but no evidence of active TB (referred to as non-TB) (**Table S1**). No physical separation or microbial enrichment was performed to avoid technical error or bias. An average of 1.7×10^8 reads were generated per sample. Sequence reads were first quality filtered then aligned to the human genome, with unaligned reads extracted for microbiome taxonomy classification and species mapping (**Fig. 1a**). Regardless of HIV-1 status, human reads accounted for an average of 74(±17)% and bacteria for 13(±13)% of all sequenced reads in tuberculous samples (**Fig. 1b**). In contrast, non-TB sputa generated significantly fewer human reads (44±20%, p=0.0007) and a non-statistically significant higher number of bacterial reads (24±21%). Unassigned reads may have arisen from incomplete filtering of human sequences and from fungal and unidentified bacterial genomes missing from the database.

We first examined the impact of *Mtb* and HIV-1 infections on the host transcriptome. We identified 21 genes that, when compared to HIV-1 uninfected patients, were differentially expressed in HIV-1 co-infected TB sputa (**Table S2**), including upregulation of T-cell markers such as CD8A/B, LAG3 and CRTAM. This observation was consistent with that from nonhuman primates with TB, in which co-infection with simian immunodeficiency virus significantly induced LAG3 expression (9), suggesting that T-cell recruitment to TB sputum is quantitatively and qualitatively affected by HIV-1 co-infection. The presence of *Mtb* had a significant impact on the host transcriptome in the respiratory tract, with total segregation between TB and non-TB samples in Principal Component Analysis (**Supplementary Fig. S1**). One of the non-TB samples (SP321) was a conspicuous outlier and was omitted from further analysis. Comparison between TB sputa (regardless of HIV-1 status) and non-TB controls identified 5843 genes that were differentially expressed (log2FoldChange > ±0.5, *p-adjusted* < 0.05; **Table S3**). Gene set enrichment analysis of these 5843 genes identified 11 significant gene sets, of which 9 were positively enriched in TB sputum and 2 were negatively enriched in non-TB (**Fig. 1c**).

The TB enriched pathways consisted of inflammatory responses mediated by interferon-gamma (IFNγ), tumor necrosis factor alpha (TNF-α) and, to a lesser extent, by type I interferon (IFNα/β) (**Fig. 1d**). The enhanced transcription of these inflammatory mediators is consistent with elevated cytokine concentrations previously reported in TB sputum when compared to pneumonia controls (10). Significant transcriptional changes associated with T helper cell activation and differentiation, including T-bet, GATA3, RORγt and FOXP3 transcriptional regulators, were also detected despite lymphocytes typically accounting for less than 1% of the total cellular composition in TB sputum (10) (**Fig. 1e**). Expression of IL-18 was significantly downregulated in TB sputum while its neutralizing binding protein (IL18BP) was significantly upregulated, suggesting that the increased IFNγ-mediated response may be driven by IL-12 without IL-18 synergy (11, 12). Furthermore, increased expression of Th17 and the Foxp3$^+$ Treg subsets in TB sputa was consistent with significantly enhanced transcription of transforming growth factor beta (TGF-β). Together, the host transcriptome in sputum shares both similarities and key differences compared to whole blood (13) and reveals a significant and specific anti-mycobacterial response in the airways not found in non-TB respiratory conditions.

In parallel with the inflammatory response there was a striking change in host central metabolism in TB sputa, with evidence of a switch from oxidative phosphorylation to glycolysis (**Table S3**). Expression of genes involved in the tricarboxylic acid (TCA) cycle was significantly downregulated (**Fig. 1f**) and broken after citrate, with reduced transcription of aconitase (ACO1) and elevated transcription of aconitate decarboxylase (ACOD1/IRG1) (14) (**Supplementary Fig. S2**). The electron transport chain (ETC) (**Fig. 1g**) was also significantly downregulated in TB sputa, including genes encoding NADH dehydrogenase, cytochrome c oxidase, ubiquinol-cytochrome c reductase and mitochondrial ATP (F_0F_1) synthase (**Table S3**). In contrast, there was an enhanced expression of glucose transporter GLUT1 (encoded by SLC2A1) and lactate exporter

MCT4 (encoded by SLC16A3) (**Fig. 1h**), along with a significant increase in the ratio of LDHA to LDHB (lactate dehydrogenase A and B) (**Fig. 1h**) indicative of increased conversion from pyruvate to lactate (15). Increased transcription of genes involved in the oxidative branch of the pentose phosphate pathway (PPP) was consistent with production of NAPDH in association with generation of reactive oxygen species (ROS) (**Fig. 1i** and **Supplementary Fig. S3**), though transcripts associated with alternative NADPH-generating pathways (cytoplasmic malate dehydrogenase (MDH1), malic enzyme (ME1) and isocitrate dehydrogenase (IDH1)) were found at higher abundance in non-TB sputum. Together, these data support the notion that there is an overall reprogramming of host central metabolism during *Mtb* infection towards increased glycolysis, either as a positive feedback mechanism to maintain a fully activated immune response (16), or to produce glycolytic intermediates required for cell proliferation as part of antimicrobial defense (17).

Microbiome landscape and its adaptation to *Mtb* infection

The inflammatory response revealed by direct transcriptional profiling of sputum samples shares key features common to responses to *Mtb* infection previously documented in cell culture models and infected human and animal tissues. We anticipated that if this transcription profile was translated into a functional antimicrobial response, it may disrupt the ecology of the commensal respiratory microbiota. To test this hypothesis, we compared overall microbiome taxonomy and the transcriptional profile of dominant commensal bacterial species between TB and non-TB sputum.

Taxonomic classification of the bacterial reads identified 30 phyla, 613 genera and 1331 species (**Table S4**). Reads mapping to sequenced bacterial genomes ranged from 10^6 to 10^8 and the overall taxonomic composition of our TB sputa was similar to that previously reported using 16S DNA (18), with *Streptococcus*, *Neisseria*, *Prevotella*, *Haemophilus* and *Veillonella* being the most represented genera (**Fig. 2a**). Non-TB sputa had significantly higher microbiome species richness than TB sputa (p<0.01 for both operational taxonomic units (OTUs) and Chao1 estimator) (**Fig. 2b**), but there was no difference in species diversity (Shannon and Simpson indices) (**Fig. 2c**), indicating that the distribution of species dominance and evenness was not affected by *Mtb* infection. In keeping with published literature, similar lung and oral microbiome diversity in HIV-uninfected and HIV-infected patients (19) , species richness or diversity in TB sputa was unaffected by HIV-1 co-infection (**Fig. 2d**).

Transcriptional profiling of sputum *Mtb*

Reads mapping to *Mtb* accounted for only 0.85±2% of total mapped bacterial reads (**Fig. 3a**), ranging from 10^3 to 10^5. Consistent with evidence of lower transmission from HIV-1 co-infected patients (20), there was a significantly higher percentage of *Mtb* reads in TB-only, compared to the TB-HIV sputa (mean: 1.55% *vs.* 0.06%, respectively; p=0.027) (**Fig. 3a**).

Seven samples (6 TB-only and 1 TB-HIV) had sufficient read coverage (>$4x10^4$ reads) to quantify transcript abundance for >50% of the *Mtb* genome. Three of the samples were identified as belonging to Lineage 2, one to Lineage 3, and three to Lineage 4 (**Table S1**). In the obvious absence of a comparative control from non-TB sputa, we compared the sputum *Mtb* transcriptome to exponential and stationary phase liquid laboratory cultures of *Mtb* strain H37Rv. Plotting expression data as a correlation matrix demonstrated that the sputum profiles formed a closely related cluster that shared greater similarity to exponential than to stationary phase culture (**Fig. 3b**). Expression analysis identified 198 genes as differentially expressed between sputum and exponential culture (*p-adjusted* < 0.05; **Table S5**), and 392 genes between sputum and stationary phase (*p-adjusted* < 0.05; **Table S6**).

Transcript abundance across the ATP synthase operon in sputum was closer to stationary phase than to exponential culture (**Fig. 3c**), whereas transcription of the main ribosomal protein operons more closely resembled the exponential reference (**Fig. 3d**). A striking feature of the ribosomal protein gene profile in sputum was high abundance of transcripts for a set of four alternative ribosomal proteins characteristic of growth in a low zinc environment (**Fig. 3e**). Additional zinc-regulated genes (21) including the putative chaperone Rv0106, methyltransferase Rv2990c, and the ESX-3 operon were also significantly increased in sputum compared to laboratory culture (**Tables S5 and S6**). The ESX-3 operon is under dual control of zinc-responsive Zur and iron-responsive IdeR repressors; induction of *ppe3*, which lies upstream of the IdeR site and downstream of a Zur site, provides further indication of zinc deprivation (**Fig. 3e**). Expression of the DosR stress regulon in sputum more closely resembled the exponential than the stationary phase reference (**Tables S5 and S6**), with significantly higher expression of DosR genes in sputum samples infected with Lineage 2 compared to Lineage 4 isolates (**Fig. 3f**). Inspection of expression profiles showed that this reflected an increase in *dosR* transcripts originating from a SNP-generated constitutive start site internal to Rv3134c in Lineage 2, rather than from the stress-inducible start site upstream of Rv3134c (22-24) (**Supplementary Fig. S3**).

Thirty-four members of the KstR and KstR2 regulons involved in degradation of cholesterol side chain and ABCD rings (25), and genes involved in downstream propionate metabolism by the methylcitrate cycle (26) and methylmalonate pathways (27) were consistently higher in sputum than laboratory culture (**Fig. 3g**). This is similar to previous descriptions of the induction of *Mtb* cholesterol catabolism genes in macrophage and mouse models (28, 29). PhoP plays an important role in transcriptional regulation during *Mtb* infection and analysis by chromatin-immunoprecipitation has identified a set of genes that are regulated by binding of PhoP to upstream sites (30). Twenty PhoP-regulated transcripts, including small RNA *mcr7*, were differentially expressed in sputum compared to laboratory culture; in all but one case the sputum profile was consistent with a decrease in PhoP binding (**Table S5**).

We validated 15 differentially expressed genes using NanoString methodology and compared transcript levels in three sputum samples against an independent *Mtb* H37Rv reference culture. These included representative upregulated (KstR, Zur, propionate) and downregulated (ATP and mycobactin synthesis) genes. All genes showed the same pattern of differential expression (**Table S7**) and validated the use of dual RNA-Seq in studying *Mtb* transcriptome despite its minor representation among the microbial community.

Discussion

Mycobacterium tuberculosis spends most of its life sequestered in lesions within tissues, but in order to transmit to a new host it has to move into the respiratory tract prior to release in the form of infectious aerosol droplets (31). The transmission phase is difficult to model in experimental systems and is poorly understood. We reasoned that sputum samples could be exploited to obtain additional information about conditions in the respiratory tract that may influence the efficiency of TB transmission. We generated RNA sequence data directly from sputum and analyzed these with respect to host, pathogen and microbiome transcripts to provide a comprehensive overview of the entire ecosystem. This is the first report that such strategy can be successfully applied to pathological specimens, with manifest implications for the study of other human infectious diseases to complement *in vitro* and animal models.

Comparison of host transcript profiles from TB patient sputum with *Mtb*-negative sputum revealed wholesale changes characteristic of the innate and adaptive immune inflammatory response. Given the unpromising physical appearance of sputum as a heterogeneous mixture of cell debris and mucoid secretions, the homogeneity and clarity of the transcriptional response is striking and may reflect elimination of signal from dead cells by mRNA degradation. As in previous clinical studies using whole blood (32), we detected a strong type I/II interferon-mediated cytokine responses in sputum, but a strong T-cell activation and differentiation signature detected in sputum is not seen in blood; likely reflecting sequestration of these cells at the site of disease. These changes were accompanied by a metabolic shift towards glycolysis with a reduction in oxidative phosphorylation and a broken TCA cycle (33). The Warburg effect in mycobacterial infection is IFNγ-dependent (34) and probably results from a functional change in the mitochondria from energy generation to production of ROS. Upregulation of superoxide dismutase, myeloperoxidase, and glutathione peroxidase were identified in TB sputa (**Table S3**), implicating a shift in the role of host mitochondria towards bactericidal activity. A switch to glycolysis, which allows rapid production of ATP, would therefore compensate for energy loss and maintain the mitochondrial membrane potential, while upholding antimicrobial defense mechanisms.

While the majority of microbiome studies focus on the intestine, there is increasing interest in respiratory microbiota (35). Only a few studies have examined the microbiome in TB (18, 36, 37). The bacterial species detected by sputum RNA sequencing in our cohort are similar to those reported in other studies of the oral cavity and respiratory tract, reflecting the inevitable mixing associated with coughing and expectoration, and include a combination of aerobic and anaerobic members of firmicute, bacteroidetes and proteobacterial phyla. As reported in previous studies of the lung microbiome, we did not observe any major impact of HIV-1 infection on taxonomic distribution (19). In a recent 16S rDNA based analysis of tuberculous and non-tuberculous sputa, no association between the sputum microbiota composition and TB disease, or variation throughout anti-TB treatment, could be found in three different settings (38). The authors suggested transcriptomic approaches may provide greater power and in this single centre study we did find a significant reduction in species richness in TB compared to non-TB sputum. Intriguingly, despite having active disease, *Mtb* only accounted for a very small percentage of total bacterial reads measured and was very small in those with HIV-1.

It is likely the change in pattern of metabolism in tuberculous sputum we describe is majorly contributed to by neutrophils as these cells are the predominant infected phagocytic cells in the airways of patients with active pulmonary TB (39). It is recognised that even minimal tuberculous lesions can be sensitively detected by uptake of the false substrate [^{18}F]-fluorodeoxyglucose, most likely by neutrophils (40). We did not perform cell counts on sputum and single cell RNA sequencing analysis of sputum would likely be highly demanding. Thus our ability to deconvolute the cellular origin of the host sputum transcriptome is limited. We did detect the simultaneous over-representation of type I and II interferon pathways in sputum recapitulating findings in peripheral blood (13), and more recently found in the lungs of mice in conjunction with increased glycolysis (41). We also acknowledge that the total read counts detected for

Mtb is low for typical differential gene expression analysis. This is due to the one-step protocol in which no bacterial enrichment was performed in order to accurately assess the abundance of *Mtb* in its natural environment and to avoid induction of transcriptomic changes during the enrichment process. Despite the low read counts and its scarce representation amongst total bacterial population, there was an overwhelming upregulation of genes associated with cholesterol catabolism (29, 42). The ability of *Mtb* to utilize cholesterol is unique amongst the major species in the respiratory microbiome as *Mtb* can shunt the toxic by-product (propionate) into the methylcitrate cycle and the methylmalonyl pathway, which may be of a crucial adaptive significance. The *Mtb* sputum transcriptome also reveals evidence of zinc deprivation. This is of particular interest in light of evidence that the bacteria face the opposite challenge of zinc intoxication when phagocytosed by activated macrophages (43). Neutrophil-derived calprotectin may restrict the availability of zinc in the respiratory tract, and competition with commensals for free zinc may represent a vulnerability of *Mtb* in sputum. It has been proposed that zinc limitation defines a population of *Mtb* with anticipatory adaptations against impending immune attack, based on the evidence that Zinc-limited Mtb are more resistant to oxidative stress and exhibit increased survival and induce more severe pulmonary granulomas in mice (44). Similarly, contrasting with results in macrophage culture (45), the *Mtb* sputum transcriptome is characterized by reduced activation of the PhoP regulon in comparison to exponential culture. Several studies have partially characterized the transcriptome of *Mtb* from sputum or bronchoalveolar lavage using whole-genome probed-based qPCR or microarray (2, 46-49). There is significant common ground in energy metabolism, ATP synthesis, iron response and PhoP regulon when comparing our data to these studies, but with key differences in the DosR regulon. Expression of DosR genes in sputum *Mtb* has been described to resemble hypoxic non-replicating laboratory cultures (47, 49), or distinctive from both aerobic and hypoxic cultures (2), and found in lower abundance in HIV-1 coinfected patient samples when lineage was controlled (50). The discrepancies could be due to geographic location and lineage of the samples collected, sample preparation, the technology used for quantification and the growth conditions and origin of the laboratory cultures used for comparison. Finally, it will be important to determine the ratio of extracellular to intracellular *Mtb* in sputum; while there is clearly recruitment of an activated population of inflammatory cells in TB sputum, it is possible that they are engaged in phagocytosis of commensal bacteria rather than *Mtb*.

Conclusions

The overall aim of our research was to identify interventions that will reduce the viability of *Mtb* in the respiratory tract in order to reduce the efficiency of infection and transmission. We anticipate that this could involve vaccination to prime effective T cell responses and opsonizing antibodies, targeted antibody or small molecule therapies to optimize host responses, and nutritional or antibiotic interventions that alter the respiratory microbiome. Comprehensive mapping of the transcriptional landscape of both the host and the *Mtb* described here provides a crucial framework for further study.

Materials and Methods

Ethical statement. The Human Research Ethics Committee of the University of Cape Town approved the study (HREC References: 031/2012 and 568/2012) and written informed consent was obtained from all participants.

Data Accession. The RNA-Seq data reported in this paper have been deposited in the European Nucleotide Archive with the study number ERP012221 and accession number PRJEB10919.

Patient cohort and sample collection. The study was conducted at the Ubuntu Clinic, an integrated HIV/TB outpatient facility in Khayelitsha Site B, Cape Town. Adult (≥ 18years old) patients starting TB treatment for confirmed pulmonary TB as evidenced by a sputum sample that was 1) smear positive for acid-fast bacilli, or 2) positive for *Mtb* by Xpert® *Mtb*/Rif (Cepheid) testing, were recruited for the study. Additional sputum samples from respiratory symptomatic non-TB patients were collected subsequently. TB disease was excluded when patients did not meet the two above criteria and had no radiographic evidence of TB. Demographic data (age and sex), HIV status, CD4 count and antiretroviral therapy (ART) prescription (if HIV-1 infected) are recorded in **Table S1**. Spontaneously produced sputum was collected from each patient recruited prior to treatment initiation. Sputum samples were collected in a 40 ml specimen jar and TRIzol reagent (Life Technologies) was added in a 2:1 ratio (i.e. 2ml of TRIzol to 1ml of sputum) with a pipette. The specimen jar was then closed and shaken to homogenize the sputum. Samples were stored at -80°C until use.

Bacterial strains and growth conditions. *M. tuberculosis* H37Rv (SysteMTb strain) was grown in Middlebrook 7H9 medium (Sigma-Aldrich) with 10% albumin dextrose catalase supplement (Sigma-Aldrich), 0.2% glycerol and 0.05% Tween 80. Exponential phase mycobacterial cultures were grown to OD_{600} between 0.6-0.8 in roller bottle at 37°C and 2rpm. Stationary phase cultures were grown for 4 weeks after OD_{600} reached 1.0. Bacteria were harvested by centrifugation at room temperature for 5min at 2000g.

TRIzol reagent was immediately added to the bacterial pellet in 2:1 ratio, followed by vigorous vortexing for homogenization. Samples were stored at -80°C until use.

RNA Extraction. 2mL of TRIzol preserved sputum or H37Rv cultures were thawed immediately before RNA extraction. Samples were ribolyzed twice with 0.1mm silica spheres (MPBio) with a setting of 6m/s for 45 seconds. Ribolyzed samples were immediately placed on ice and centrifuge briefly. Chloroform (200µl) was added to each millilitre of lysed sample and vortexed for 1 min before centrifugation at 10000g for 1min. The aqueous phase was carefully transferred to a new Eppendorf tube and mixed rigorously with equal volume of Chloroform:Isoamyl alcohol 24:1 and centrifuged at 10000g for 5 min. The aqueous phase was carefully transferred to a new Eppendorf tube and mixed with an equal volume of 100% ethanol. The mixture was then passed through a Zymo-Spin IC column (Zymo Research) where nucleic acids were captured in the membrane. The column was treated twice with 10U of TURBO DNase (Thermo Fisher Scientific) and 100U of RNase inhibitor (Takara Clontech) at 37°C for 30min until DNA-free. The DNase-treated RNA was then purified using the RNA Clean and Concentrator-5 kit (Zymo Research) and eluted in nuclease-free water. RNA was extracted from 26 sputum samples and from 4 culture samples and their quantity and quality were determined by Qubit fluorometer (Thermo Fisher Scientific), NanoDrop spectrophotometer (Thermo Fisher Scientific) and Caliper LabChip systems (Perkin Elmer).

Library preparation and RNA-Seq. RNA-Seq libraries for the 26 sputum samples and 4 culture samples were prepared with 200ng of corresponding RNA using the Ovation Human FFPE RNA-Seq Multiplex System (NuGen), which includes proprietary oligos for removal of human rRNA and customized oligos to remove rRNA of *Mtb*. The cDNA was sheared to approximately 200bp with a Covaris E220 ultrasonicator (Covaris) prior to adaptor ligation and amplification. All cDNA libraries were quantified using Qubit fluorometer and quality checked using the DNA-1000 kit (Agilent) on a 2100 Bioanalyzer. Each sputum library was loaded onto a single lane in a flow cell and sequenced with a Hi-Seq 2500 instrument (Illumina). With the exception of 4 samples

(Rv_E1, Rv_S1, SP55 and SP61) where only ~100 million 100bp-single-end reads were obtained, all other sputum samples and laboratory cultures (Rv_E2 and Rv_S2) generated ~200million 100bp-single-end reads.

Read mapping and read counting. The quality of the Illumina-produced fastq files was assessed using FastQC and poor quality reads were trimmed using the SolexaQA package (51) using default parameters, trimming bases with confidence P> 0.05 and removing reads <25 bases. The good quality reads were mapped to human genome (NCBI GRCh38 build) using Tophat2 using default parameters (52). The non-human reads were then exported for taxonomic classification using Kraken (see section below) and subsequently aligned to reference genomes of *Mtb* and commensal bacteria (see **Table S6** for accession numbers and references) as single-end data using BWA v 0.7.12 (53) and genome coverage was calculated using BEDTools (54). Lineage of the sputum *Mtb* was determined using the KvarQ algorithm (55) and scanned with the SNPs testsuite.

Read count normalization and differential gene expression analysis. Date were analyzed in R ver 3.5.2. Read count normalization was done using DESeq2 (56), which is based on a negative binomial distribution model. DESeq2 also determined the fold change between sputum *Mtb* and H37Rv cultures or between TB and non-TB samples (results are shown as log2FoldChange). Statistical significance was calculated and adjusted using the Benjamini Hochberg multiple testing method with a false discovery rate of 10% (shown as p-adjusted). Differentially expressed genes in *Mtb* that are statistically significant were used to generate a correlation matrix using corrplot with hierarchical clustering (57). A heatmap was created for the differentially expressed genes in the host transcriptome using gplots. Pathway analysis of differentially expressed genes was performed using Gene Set Enrichment Analysis (58) and IPA Ingenuity (QIAGEN) for human data and KEGG pathway (59) for bacterial data.

Taxonomic classification of sequenced reads*.* For each of the 26 sputum samples (17 samples from patients with untreated active TB and 9 additional samples from patients who were non-TB respiratory symptomatic), the set of non-human reads was used for taxonomic classification using Kraken (60) screening against the reference MiniKraken database representing complete bacterial, archaeal, and viral genomes in RefSeq. Classification results were visualized using Krona (61). The percent representation of *Mtb* was calculated relative to the total bacterial sequences identified and statistical difference between TB and TB-HIV groups was calculated using the nonparametric Mann-Whitney *U*-test in Prism 6 software.

Taxonomic diversity and comparative analysis among sputum samples. Taxonomic reports derived from Kraken were imported into QIIME (62) for the comparative analysis of microbiome species richness and diversity among TB, TB-HIV and non-TB samples. Species richness was calculated based on the number of observed operational taxonomic units (OTUs) and the Chao1 estimator, which estimates the real species richness based on OTUs. Species diversity, which indicates for evenness and distribution, was estimated using the Shannon and Simpson indices. Statistical difference between different sample groups was calculated using the nonparametric Mann-Whitney *U*-test in Prism 6 software.

NanoString validation of gene expression. A set of 15 differentially abundant transcripts identified by RNA-Seq were reinvestigated and validated using a customized NanoString nCounter assay (63) (Codeset ID MtbH37Rv, NanoString Technologies). An independent set of triplicate H37Rv (exponential phase culture) was prepared as described above and total RNA extracted. Three sputum samples (SP28, SP29 and SP61) that were previously used for RNA-Seq library construction had sufficient quantity of RNA remaining and were used in the NanoString nCounter assay. Hybridisation and scanning of the NanoString assay was performed by the UCL Nanostring facility according to the manufacturer’s instructions. Briefly, customized barcoded capture/reporter probe pairs specific for each transcript were hybridized overnight at 65 °C to 2.5μg of total RNA for sputum samples and 5ng of total RNA for culture samples. Positive and negative control probe pairs were also included. Unhybridized probes were removed, and the hybridized probes were purified on an nCounter Prep Station. The barcode on each reporter probe was scanned with an nCounter Digital Analyzer to generate a quantitative measure of the hybridized RNA. Sample signal values were subtracted for background, defined as the mean number of counts for negative control probes plus 1 standard deviation. The filtered signal values were then normalized using DESeq2 and differential expression between sputum samples and exponential phase cultures were computed as described above.

References

1. WHO. 2020. Global tuberculosis report. WHO, Geneva.
2. Garcia BJ, Loxton AG, Dolganov GM, Van TT, Davis JL, de Jong BC, Voskuil MI, Leach SM, Schoolnik GK, Walzl G, Strong M, Walter ND. 2016. Sputum is a surrogate for bronchoalveolar lavage for monitoring *Mycobacterium tuberculosis* transcriptional profiles in TB patients. Tuberculosis (Edinb) 100:89-94.
3. Westermann AJ, Forstner KU, Amman F, Barquist L, Chao Y, Schulte LN, Muller L, Reinhardt R, Stadler PF, Vogel J. 2016. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. Nature 529:496-501.
4. Baddal B, Muzzi A, Censini S, Calogero RA, Torricelli G, Guidotti S, Taddei AR, Covacci A, Pizza M, Rappuoli R, Soriani M, Pezzicoli A. 2015. Dual RNA-seq of Nontypeable Haemophilus influenzae and Host Cell Transcriptomes Reveals Novel Insights into Host-Pathogen Cross Talk. MBio 6:e01765-15.
5. Nuss AM, Beckstette M, Pimenova M, Schmuhl C, Opitz W, Pisano F, Heroven AK, Dersch P. 2017. Tissue dual RNA-seq allows fast discovery of infection-specific functions and riboregulators shaping host-pathogen transcriptomes. Proc Natl Acad Sci U S A 114:E791-E800.
6. Rienksma RA, Suarez-Diez M, Mollenkopf HJ, Dolganov GM, Dorhoi A, Schoolnik GK, Martins Dos Santos VA, Kaufmann SH, Schaap PJ, Gengenbacher M. 2015. Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing. BMC Genomics 16:34.
7. Lamont EA, Xu WW, Sreevatsan S. 2013. Host-Mycobacterium avium subsp. paratuberculosis interactome reveals a novel iron assimilation mechanism linked to nitric oxide stress during early infection. BMC Genomics 14:694.
8. Pisu D, Huang L, Grenier JK, Russell DG. 2020. Dual RNA-Seq of Mtb-Infected Macrophages In Vivo Reveals Ontologically Distinct Host-Pathogen Interactions. Cell Rep 30:335-350 e4.
9. Phillips BL, Mehra S, Ahsan MH, Selman M, Khader SA, Kaushal D. 2015. LAG3 expression in active *Mycobacterium tuberculosis* infections. Am J Pathol 185:820-33.

10. Ribeiro-Rodrigues R, Resende Co T, Johnson JL, Ribeiro F, Palaci M, Sa RT, Maciel EL, Pereira Lima FE, Dettoni V, Toossi Z, Boom WH, Dietze R, Ellner JJ, Hirsch CS. 2002. Sputum cytokine levels in patients with pulmonary tuberculosis as early markers of mycobacterial clearance. Clin Diagn Lab Immunol 9:818-23.

11. Okamura H, Kashiwamura S, Tsutsui H, Yoshimoto T, Nakanishi K. 1998. Regulation of interferon-gamma production by IL-12 and IL-18. Curr Opin Immunol 10:259-64.

12. Tominaga K, Yoshimoto T, Torigoe K, Kurimoto M, Matsui K, Hada T, Okamura H, Nakanishi K. 2000. IL-12 synergizes with IL-18 or IL-1beta for IFN-gamma production from human T cells. Int Immunol 12:151-60.

13. 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. 2010. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature 466:973-7.

14. Luan HH, Medzhitov R. 2016. Food Fight: Role of Itaconate and Other Metabolites in Antimicrobial Defense. Cell Metab 24:379-87.

15. Draoui N, Feron O. 2011. Lactate shuttles at a glance: from physiological paradigms to anti-cancer treatments. Dis Model Mech 4:727-32.

16. Tan Z, Xie N, Banerjee S, Cui H, Fu M, Thannickal VJ, Liu G. 2015. The monocarboxylate transporter 4 is required for glycolytic reprogramming and inflammatory response in macrophages. J Biol Chem 290:46-55.

17. Vander Heiden MG, Cantley LC, Thompson CB. 2009. Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science 324:1029-33.

18. Cheung MK, Lam WY, Fung WY, Law PT, Au CH, Nong W, Kam KM, Kwan HS, Tsui SK. 2013. Sputum microbiota in tuberculosis as revealed by 16S rRNA pyrosequencing. PLoS One 8:e54574.

19. Beck JM, Schloss PD, Venkataraman A, Twigg Iii H, Jablonski KA, Bushman FD, Campbell TB, Charlson ES, Collman RG, Crothers K, Curtis JL, Drews KL, Flores SC, Fontenot AP, Foulkes MA, Frank I, Ghedin E, Huang L, Lynch SV,

Morris A, Palmer BE, Schmidt TM, Sodergren E, Weinstock GM, Young VB, Lung HIVMP. 2015. Multi-center Comparison of Lung and Oral Microbiomes of HIV-infected and HIV-uninfected Individuals. Am J Respir Crit Care Med 192:1335-44.

20. Huang CC, Tchetgen ET, Becerra MC, Cohen T, Hughes KC, Zhang Z, Calderon R, Yataco R, Contreras C, Galea J, Lecca L, Murray M. 2014. The effect of HIV-related immunosuppression on the risk of tuberculosis transmission to household contacts. Clin Infect Dis 58:765-74.

21. Maciag A, Dainese E, Rodriguez GM, Milano A, Provvedi R, Pasca MR, Smith I, Palu G, Riccardi G, Manganelli R. 2007. Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon. J Bacteriol 189:730-40.

22. Rose G, Cortes T, Comas I, Coscolla M, Gagneux S, Young DB. 2013. Mapping of genotype-phenotype diversity among clinical isolates of *Mycobacterium tuberculosis* by sequence-based transcriptional profiling. Genome Biol Evol 5:1849-62.

23. Reed MB, Gagneux S, Deriemer K, Small PM, Barry CE, 3rd. 2007. The W-Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. J Bacteriol 189:2583-9.

24. Domenech P, Zou J, Averback A, Syed N, Curtis D, Donato S, Reed MB. 2016. Unique regulation of the DosR regulon in the Beijing lineage of *Mycobacterium tuberculosis*. J Bacteriol 199:e00696-16.

25. Wipperman MF, Sampson NS, Thomas ST. 2014. Pathogen roid rage: cholesterol utilization by *Mycobacterium tuberculosis*. Crit Rev Biochem Mol Biol 49:269-93.

26. Upton AM, McKinney JD. 2007. Role of the methylcitrate cycle in propionate metabolism and detoxification in *Mycobacterium smegmatis*. Microbiology 153:3973-82.

27. Savvi S, Warner DF, Kana BD, McKinney JD, Mizrahi V, Dawes SS. 2008. Functional characterization of a vitamin B12-dependent methylmalonyl pathway

in *Mycobacterium tuberculosis*: implications for propionate metabolism during growth on fatty acids. J Bacteriol 190:3886-95.

28. VanderVen BC, Fahey RJ, Lee W, Liu Y, Abramovitch RB, Memmott C, Crowe AM, Eltis LD, Perola E, Deininger DD, Wang T, Locher CP, Russell DG. 2015. Novel inhibitors of cholesterol degradation in *Mycobacterium tuberculosis* reveal how the bacterium's metabolism is constrained by the intracellular environment. PLoS Pathog 11:e1004679.

29. Pandey AK, Sassetti CM. 2008. Mycobacterial persistence requires the utilization of host cholesterol. Proc Natl Acad Sci U S A 105:4376-80.

30. Solans L, Gonzalo-Asensio J, Sala C, Benjak A, Uplekar S, Rougemont J, Guilhot C, Malaga W, Martin C, Cole ST. 2014. The PhoP-dependent ncRNA Mcr7 modulates the TAT secretion system in *Mycobacterium tuberculosis*. PLoS Pathog 10:e1004183.

31. Fennelly KP, Martyny JW, Fulton KE, Orme IM, Cave DM, Heifets LB. 2004. Cough-generated aerosols of *Mycobacterium tuberculosis*: a new method to study infectiousness. Am J Respir Crit Care Med 169:604-9.

32. O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MP. 2013. The immune response in tuberculosis. Annu Rev Immunol 31:475-527.

33. O'Neill LA. 2015. A broken krebs cycle in macrophages. Immunity 42:393-4.

34. Appelberg R, Moreira D, Barreira-Silva P, Borges M, Silva L, Dinis-Oliveira RJ, Resende M, Correia-Neves M, Jordan MB, Ferreira NC, Abrunhosa AJ, Silvestre R. 2015. The Warburg effect in mycobacterial granulomas is dependent on the recruitment and activation of macrophages by interferon-gamma. Immunology 145:498-507.

35. Rogers GB, Shaw D, Marsh RL, Carroll MP, Serisier DJ, Bruce KD. 2015. Respiratory microbiota: addressing clinical questions, informing clinical practice. Thorax 70:74-81.

36. Zhou Y, Lin F, Cui Z, Zhang X, Hu C, Shen T, Chen C, Zhang X, Guo X. 2015. Correlation between Either *Cupriavidus* or *Porphyromonas* and Primary Pulmonary Tuberculosis Found by Analysing the Microbiota in Patients' Bronchoalveolar Lavage Fluid. PLoS One 10:e0124194.

37. Cui Z, Zhou Y, Li H, Zhang Y, Zhang S, Tang S, Guo X. 2012. Complex sputum microbial composition in patients with pulmonary tuberculosis. BMC Microbiol 12:276.

38. Sala C, Benjak A, Goletti D, Banu S, Mazza-Stadler J, Jaton K, Busso P, Remm S, Leleu M, Rougemont J, Palmieri F, Cuzzi G, Butera O, Vanini V, Kabir S, Rahman SMM, Nicod L, Cole ST. 2020. Multicenter analysis of sputum microbiota in tuberculosis patients. PLoS One 15:e0240250.

39. Eum SY, Kong JH, Hong MS, Lee YJ, Kim JH, Hwang SH, Cho SN, Via LE, Barry CE, 3rd. 2010. Neutrophils are the predominant infected phagocytic cells in the airways of patients with active pulmonary TB. Chest 137:122-8.

40. 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. 2016. Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[(18)F]fluoro-D-glucose positron emission and computed tomography. Nat Med 22:1090-1093.

41. Moreira-Teixeira L, Stimpson PJ, Stavropoulos E, Hadebe S, Chakravarty P, Ioannou M, Aramburu IV, Herbert E, Priestnall SL, Suarez-Bonnet A, Sousa J, Fonseca KL, Wang Q, Vashakidze S, Rodriguez-Martinez P, Vilaplana C, Saraiva M, Papayannopoulos V, O'Garra A. 2020. Type I IFN exacerbates disease in tuberculosis-susceptible mice by inducing neutrophil-mediated lung inflammation and NETosis. Nat Commun 11:5566.

42. Griffin JE, Pandey AK, Gilmore SA, Mizrahi V, McKinney JD, Bertozzi CR, Sassetti CM. 2012. Cholesterol catabolism by *Mycobacterium tuberculosis* requires transcriptional and metabolic adaptations. Chem Biol 19:218-27.

43. Botella H, Peyron P, Levillain F, Poincloux R, Poquet Y, Brandli I, Wang C, Tailleux L, Tilleul S, Charriere GM, Waddell SJ, Foti M, Lugo-Villarino G, Gao Q, Maridonneau-Parini I, Butcher PD, Castagnoli PR, Gicquel B, de Chastellier C, Neyrolles O. 2011. Mycobacterial p(1)-type ATPases mediate resistance to zinc poisoning in human macrophages. Cell Host Microbe 10:248-59.

44. Dow A, Sule P, O'Donnell TJ, Burger A, Mattila JT, Antonio B, Vergara K, Marcantonio E, Adams LG, James N, Williams PG, Cirillo JD, Prisic S. 2021. Zinc

limitation triggers anticipatory adaptations in Mycobacterium tuberculosis. PLoS Pathog 17:e1009570.

45. Perez E, Samper S, Bordas Y, Guilhot C, Gicquel B, Martin C. 2001. An essential role for phoP in *Mycobacterium tuberculosis* virulence. Mol Microbiol 41:179-87.

46. Walter ND, Dolganov GM, Garcia BJ, Worodria W, Andama A, Musisi E, Ayakaka I, Van TT, Voskuil MI, de Jong BC, Davidson RM, Fingerlin TE, Kechris K, Palmer C, Nahid P, Daley CL, Geraci M, Huang L, Cattamanchi A, Strong M, Schoolnik GK, Davis JL. 2015. Transcriptional Adaptation of Drug-tolerant *Mycobacterium tuberculosis* During Treatment of Human Tuberculosis. J Infect Dis 212:990-8.

47. Honeyborne I, McHugh TD, Kuittinen I, Cichonska A, Evangelopoulos D, Ronacher K, van Helden PD, Gillespie SH, Fernandez-Reyes D, Walzl G, Rousu J, Butcher PD, Waddell SJ. 2016. Profiling persistent tubercule bacilli from patient sputa during therapy predicts early drug efficacy. BMC Med 14:68.

48. Walter ND, de Jong BC, Garcia BJ, Dolganov GM, Worodria W, Byanyima P, Musisi E, Huang L, Chan ED, Van TT, Antonio M, Ayorinde A, Kato-Maeda M, Nahid P, Leung AM, Yen A, Fingerlin TE, Kechris K, Strong M, Voskuil MI, Davis JL, Schoolnik GK. 2016. Adaptation of *Mycobacterium tuberculosis* to Impaired Host Immunity in HIV-Infected Patients. J Infect Dis 214:1205-11.

49. Garton NJ, Waddell SJ, Sherratt AL, Lee SM, Smith RJ, Senner C, Hinds J, Rajakumar K, Adegbola RA, Besra GS, Butcher PD, Barer MR. 2008. Cytological and transcript analyses reveal fat and lazy persister-like bacilli in tuberculous sputum. PLoS Med 5:e75.

50. Agostini C, Trentin L, Zambello R, Semenzato G. 1993. HIV-1 and the lung. Infectivity, pathogenic mechanisms, and cellular immune responses taking place in the lower respiratory tract. Am Rev Respir Dis 147:1038-49.

51. Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11:485.

52. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36.

53. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589-95.

54. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-2.

55. Steiner A, Stucki D, Coscolla M, Borrell S, Gagneux S. 2014. KvarQ: targeted and direct variant calling from fastq reads of bacterial genomes. BMC Genomics 15:881.

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

57. Wei T. 2013. Corrplot: visualization of a correlation matrix. R package version 0.60.

58. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545-50.

59. Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27-30.

60. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46.

61. Ondov BD, Bergman NH, Phillippy AM. 2011. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics 12:385.

62. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335-6.

63. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, James JJ, Maysuria M, Mitton JD, Oliveri P, Osborn JL, Peng T, Ratcliffe AL, Webster PJ, Davidson EH, Hood L, Dimitrov K. 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 26:317-25.

Declarations

Ethics approval and consent to participate

The Human Research Ethics Committee of the Faculty of Health Sciences of the University of Cape Town approved the study (HREC References: 031/2012 and 568/2012) and written informed consent was obtained from all participants.

Consent for publication

No identifiable individual person's data is included in the manuscript

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files]. The RNA-Seq data have been deposited in the European Nucleotide Archive with the study number ERP012221 and accession number PRJEB10919.

Competing interests

The authors declare that they have no competing interests.

Funding

RP-JL is supported by the UK Medical Research Council (MRC/R008922/1). RJW is supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (FC00110218), the UK Medical Research Council (FC00110218), and Wellcome (FC00110218). Other funding included Wellcome (097254 for SM; 104803 and 203135 for RJW), European Union (FP-7-SysteMTb Collaborative Project 241587 and Horizon 2020 research and innovation programme (grant agreement No 637730) for TC), Marie Curie International Research Staff Exchange Scheme (FP7-PEOPLE-2011-IRSES) for NR; and the Carnegie Corporation Training Award and Discovery Foundation Academic Fellowship Award (SM). These funders had no influence on the the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. For the purpose of Open Access, the author has applied a CC BY

Authors' contributions

RPL, SM and RJW conceived and designed the experiments; SM and NR recruited, sampled and collected data from patients; RPL, MLB and AKS performed the experiments; RPL, TC, MLB, AG, SH and DBY analyzed the data; SM, NR, SH, AOG, DBY and RJW contributed materials and analysis tools; all authors contributed intellectual input; RP-JL, DBY and RJW wrote the paper.

Acknowledgements

The authors thank all the participants in this study and the health care workers and administrators at the Ubuntu Clinic. We would like to thank Meena Anissi, Leena Bhaw, and Deborah Jackson of the High Throughput Sequencing facility at the Francis Crick Institute for help with the sequencing. We thank the UCL Nanostring facility for providing the nCounter system and related services.

Figure Legends

Fig. 1 Dual host-pathogen RNA-Seq and the host transcriptome

a) Sputum samples were collected from 17 active TB and 9 non-TB respiratory symptomatic patients. Total RNA was extracted and cDNA library generated for ultra-deep RNA-sequencing. Sequence reads were first aligned to the human genome and unmapped reads were extracted for further microbiome metagenomics classification. After identifying the predominant microbiome taxa, reference-based alignment was performed to the top 10 abundant microbiome species as well as to *Mtb*. **b)** Global transcript composition profiles of TB and non-TB sputa were calculated. A reduced percentage of host reads and increased percentage of bacterial reads was recorded in non-TB samples. **c)** Heatmap showing a total of 5843 differentially expressed genes in the host transcriptomes between TB (n=17) and non-TB (n=8) sputa. Gene set enrichment analysis identified 9 pathways that were significantly enriched in TB and 2 in non-TB. The *p-value* of each enriched pathway is listed. **d)** Genes associated with IFNγ and IFNα/β signaling pathways were significantly enriched in TB samples. Red indicates upregulation in TB sputa, compared to non-TB. **e)** Evidence of T cell subset differentiation or recruitment was also observed at the transcriptional level albeit with generally low read counts. Red indicates upregulation and blue downregulation in TB *versus* non-TB sputa. **f) and g)** Metabolic reprogramming was observed in TB sputa, with decreased expression of genes in the TCA cycle and electron transport chain. The log2 fold change of TB sputa compared to non-TB is shown here and indicative of metabolic reprogramming with significant decrease in genes involved in TCA and electron transport chain. The statistical significance of each gene is listed in Supplementary Table S3. **h)** In contrast to decreased oxidative phosphorylation, there was a significant increase of genes associated with glucose uptake and lactate export in TB sputa (red) when compared to non-TB controls (blue). An increased LDHA to LDHB ratio is indicative of conversion of pyruvate to lactate. Statistical significance (p-values) are shown as asterisks: *** padj<0.001 and **** padj<0.0001. **i)** Transcript expression of genes involved in the NADPH production in the pentose phosphate pathway was also significantly higher in TB sputa. A detailed pathway map with the fold change of significant genes is shown in Supplementary Fig. S3.

Fig. 2 Global overview of sputum microbiome

a) A stacked bar chart to show the top 20 most represented microbiome genera in TB (SP12-SP61) and non-TB (SP313-SP321) sputa. SP47 had an expansion of *Haemophilus* and SP315 comprised mainly of known artefacts *Ralstonia* and *Bradyrhizobium*. These two samples were subsequently removed from all downstream analyses. **b)** Microbiome species richness and diversity were calculated. Non-TB samples (n=9) had a significantly higher number of observed operational taxonomy units (OTUs) and estimated number of true OTUs (chao1 indicator), compared to TB samples (n=17). **c)** There was no difference in species diversity as measured by the Shannon and Simpson indices, indicating species evenness and distribution did not differ between TB and non-TB groups. **d)** HIV-1 co-infection did not impact the global microbiome species richness or diversity in sputum. For panels b-d, statistical difference was calculated using Mann Whitney *U*-test and * $p<0.05$, ** $p<0.01$ and n.s. for not significant.

Fig. 3 Transcriptional profiles of sputum *Mtb*

a) Despite active TB disease, *Mtb* only accounted for 0.85±2% of all mapped bacterial reads. The percentage of *Mtb* reads was, however, significantly higher in TB-only samples, compared to TB-HIV (n=9 and n=8, respectively; $p<0.05$, Mann Whitney *U*-test). **b)** Differential gene expression between seven sputum *Mtb* samples and laboratory cultures was calculated using DESeq2. The expression data was plotted as a correlation matrix with hierarchical clustering. Exponential cultures were labeled as Rv_E1 and Rv_E2, stationary cultures were labeled as Rv_S1 and Rv_S2, and sputum samples started with the initials SP. A decrease in circle size indicates reduced correlation; red indicates a positive correlation and blue indicates negative correlation. The sputum samples showed a high degree of concordance to each other and correlated more closely to exponential cultures than to stationary cultures. **c)** Transcript abundance of ATP synthase genes in sputum *Mtb* clusters more closely to stationary phase H37Rv than to exponential phase cultures. **d)** In contrast, transcript abundance of the two major ribosomal protein operons S10 and L14 in sputum *Mtb* were found to be more similar to exponential phase H37Rv than to stationary phase cultures. Color of the heatmaps corresponds with the normalized read count of each gene. **e)** Significantly higher expression of four zinc-independent alternative ribosomal proteins was detected, along with decreased expression of the zur repressor and upregulation of *ppe3*, indicating that sputum *Mtb* was zinc-deprived. **f)** Expression of selected members of the DosR regulon is shown. Consistent with the presence of an alternative transcriptional start sites in lineage 2 isolates (22), transcript abundance of the DosR genes was significantly higher in lineage 2 than in lineage 4 sputum *Mtb*. **g)** Compared to exponential phase laboratory cultures (H37Rv), *Mtb* in sputum was found to have significantly higher expression of 34 members of the KstR and KstR2 regulons associated with cholesterol catabolism and 6 members of the downstream propionate detoxification pathways. A pathway map is shown here to illustrate the transcript expression of some of the enzymes involved in the processes. Genes that were not differentially expressed (non-significant) are colored in grey and those that were differentially expressed in sputum were colored in scale of pink and red colors according to their fold change. No downregulated genes were identified in the KstR/KstR2 regulons or either of the propionate detoxification pathways. For panels e-f, adjusted p-values (padj) were determined by DESeq2 and shown as asterisks: * padj<0.05, ** padj<0.01, *** padj <0.001 and **** padj <0.000, and n.s. for non-significant.

Description of supplementary data

Supplementary Figure S1. Principal component analysis of host transcript profiles
The host transcriptomes of sputum samples were analyzed by principal component analysis. A complete segregation of the TB (red) from the non-TB (black) samples was observed. One of the non-TB sputa (SP321) was an outlier with differential clustering pattern and was excluded from downstream analysis of the host gene expression.

Supplementary Figure S2. ltaconate biosynthesis in the host
The TCA cycle of the host in TB sputa was similar to the pattern previously described in M1 inflammatory macrophages, broken after citrate and resulted in increased production of itaconate. The ACOl enzyme that converts citrate to cis-aconitate and isocitrate was significantly down regulated, while IRGl that mediates conversion to itaconate was significantly induced in TB sputa.

Supplementary Figure S3. Pentose phosphate pathway in the host
The pentose phosphate pathway is illustrated here. The steps in the light green shade represent the oxidative branch of the pathway involved in NADPH production. Transcript abundance of enzymes that mediate the oxidative steps was significantly higher in TB sputa compared to non-TB. In contrast, there was no change or reduction of gene expression associated with the non-oxidative branch of the
pathway (shaded in dark green).

Supplementary Dataset 1 Patient characteristics
Supplementary Dataset 2 Differentially expressed genes in the human host between TB-HIV sputa and TB-only sputa
Supplementary Dataset 3 Differentially expressed genes in the human host between TB and non-TB sputa
Supplementary Dataset 4 Taxonomic classification of sputum microbiome
Supplementary Dataset 5 Differentially expressed genes in sputum Mtb compared to exponential phase H37Rv

Supplementary Dataset 6 Differentially expressed genes in sputum Mtb compared to stationary phase H37Rv

Supplementary Dataset 7 Validation by NanoString`

Figure 1

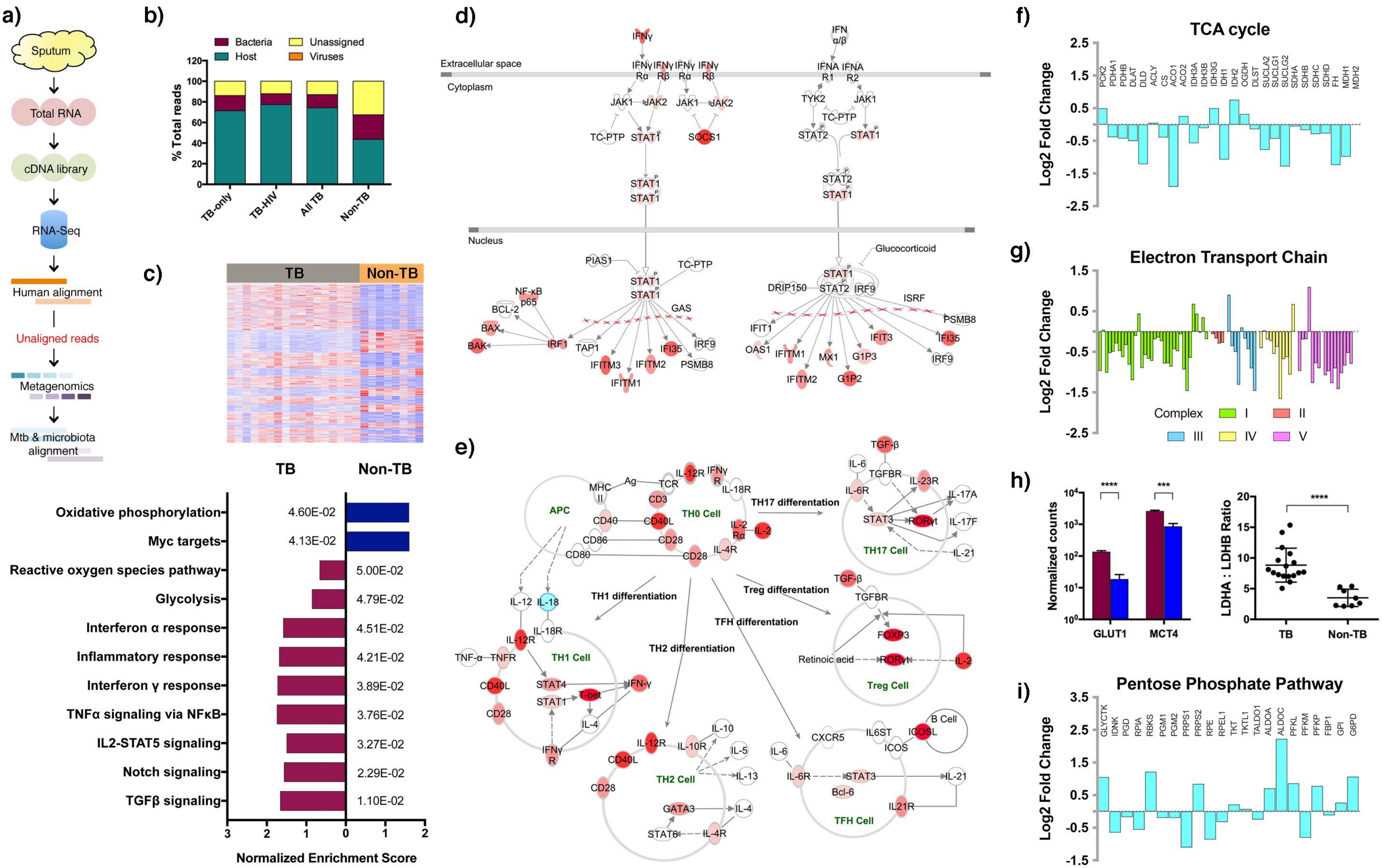

a)
Sputum
Total RNA
cDNA library
RNA-Seq
Human alignment
Unaligned reads
Metagenomics
Mtb & microbiota alignment
b)
Bacteria
Host
Unassigned
Viruses
% Total reads
0
20
40
60
80
100
120
TB-only
TB-HIV
All TB
Non-TB
c)
TB
Non-TB
TB
Non-TB
Oxidative phosphorylation 4.60E-02
Myc targets 4.13E-02
Reactive oxygen species pathway 5.00E-02
Glycolysis 4.79E-02
Interferon α response 4.51E-02
Inflammatory response 4.21E-02
Interferon γ response 3.89E-02
TNFα signaling via NFκB 3.76E-02
IL2-STAT5 signaling 3.27E-02
Notch signaling 2.29E-02
TGFβ signaling 1.10E-02
3 2 1 0 1 2
Normalized Enrichment Score
d)
Extracellular space
Cytoplasm
Nucleus
IFNγ
IFNγ Rα
IFNγ Rβ
JAK1
JAK2
TC-PTP
STAT1
SOCS1
IFN α/β
IFNA R1
IFNA R2
TYK2
STAT2
Glucocorticoid
PIAS1
NF-κB p65
BCL-2
BAX
BAK
IRF1
TAP1
IFITM3
IFITM1
IFITM2
IFI35
IRF9
PSMB8
GAS
DRIP150
ISRF
IFIT1
OAS1
MX1
G1P3
G1P2
IFIT3
e)
APC
MHC II
Ag
TCR
IL-12R
IFNγ R
IL-18R
CD3
TH0 Cell
CD40
CD40L
IL-2 Rα
IL-2
CD86
CD28
CD80
IL-4R
TH17 differentiation
TH1 differentiation
Treg differentiation
TFH differentiation
TH2 differentiation
TGF-β
IL-6
TGFBR
IL-6R
IL-23R
IL-17A
STAT3
RORγt
IL-17F
IL-21
TH17 Cell
IL-12
IL-18
TNF-α
TNFR
TH1 Cell
STAT4
T-bet
IFN-γ
STAT1
IL-4
FOXP3
Retinoic acid
Treg Cell
IL-10
IL-5
IL-13
TH2 Cell
GATA3
STAT6
IL-4R
CXCR5
IL6ST
ICOSL
B Cell
ICOS
Bcl-6
IL21R
TFH Cell
f)
TCA cycle
Log2 Fold Change
2.5
1.5
0.5
-0.5
-1.5
-2.5
PCK2 PDHA1 PDHB DLAT DLD ACLY CS ACO1 ACO2 IDH3A IDH3B IDH3G IDH1 IDH2 OGDH DLST SUCLA2 SUCLG1 SUCLG2 SDHA SDHB SDHC SDHD FH MDH1 MDH2
g)
Electron Transport Chain
Log2 Fold Change
1.5
0.5
-0.5
-1.5
-2.5
Complex I II III IV V
h)
Normalized counts
10^0 10^1 10^2 10^3 10^4
****
***
GLUT1
MCT4
LDHA : LDHB Ratio
0
5
10
15
20
****
TB
Non-TB
i)
Pentose Phosphate Pathway
Log2 Fold Change
3.5
2.5
1.5
0.5
-0.5
-1.5
GLYCTK IDNK PGD RPIA RBKS PGM1 PGM2 PRPS1 PRPS2 RPE RPEL1 TKT TKTL1 TALDO1 ALDOA ALDOC PFKL PFKM PFKP FBP1 GPI G6PD

Figure 2

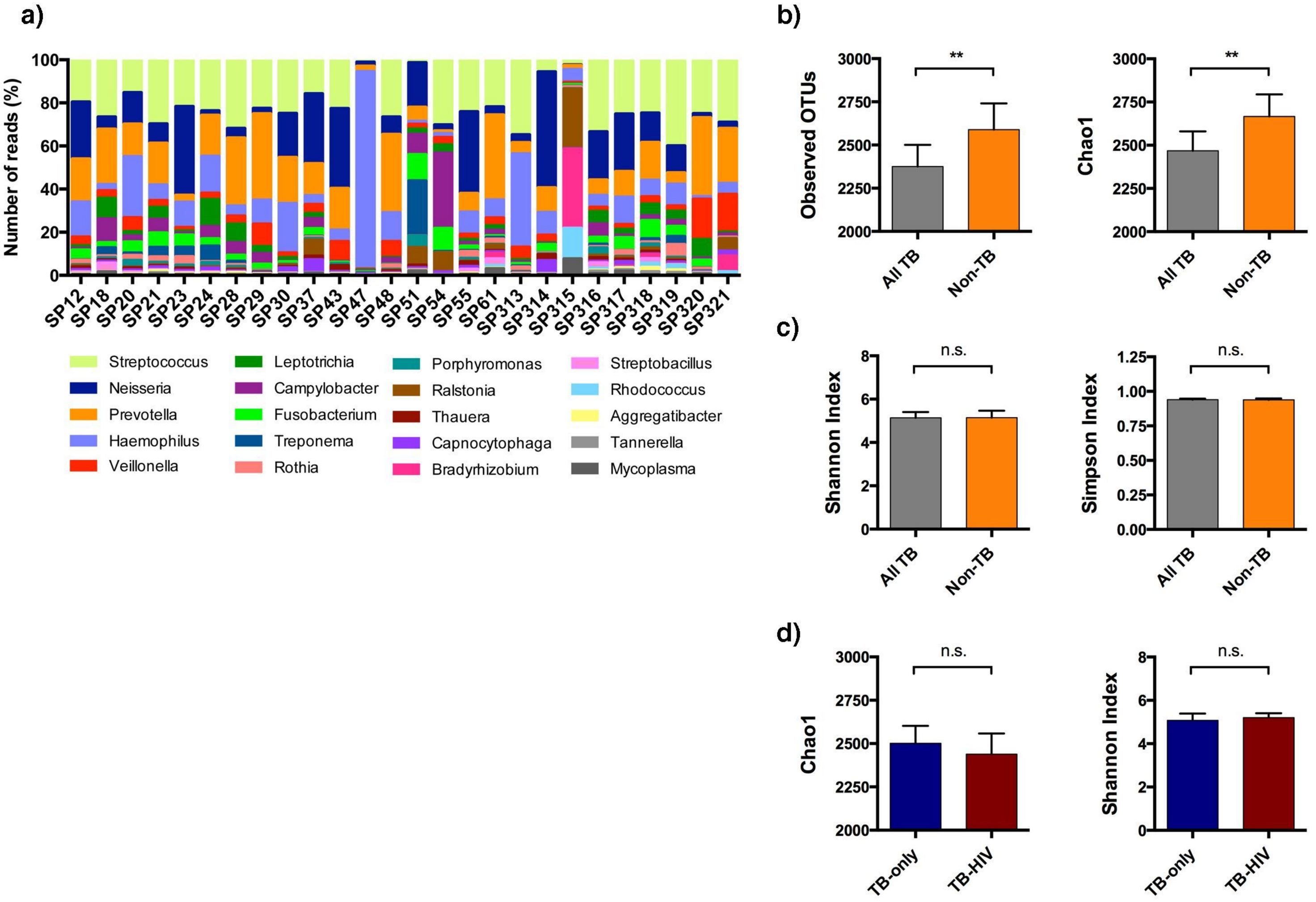
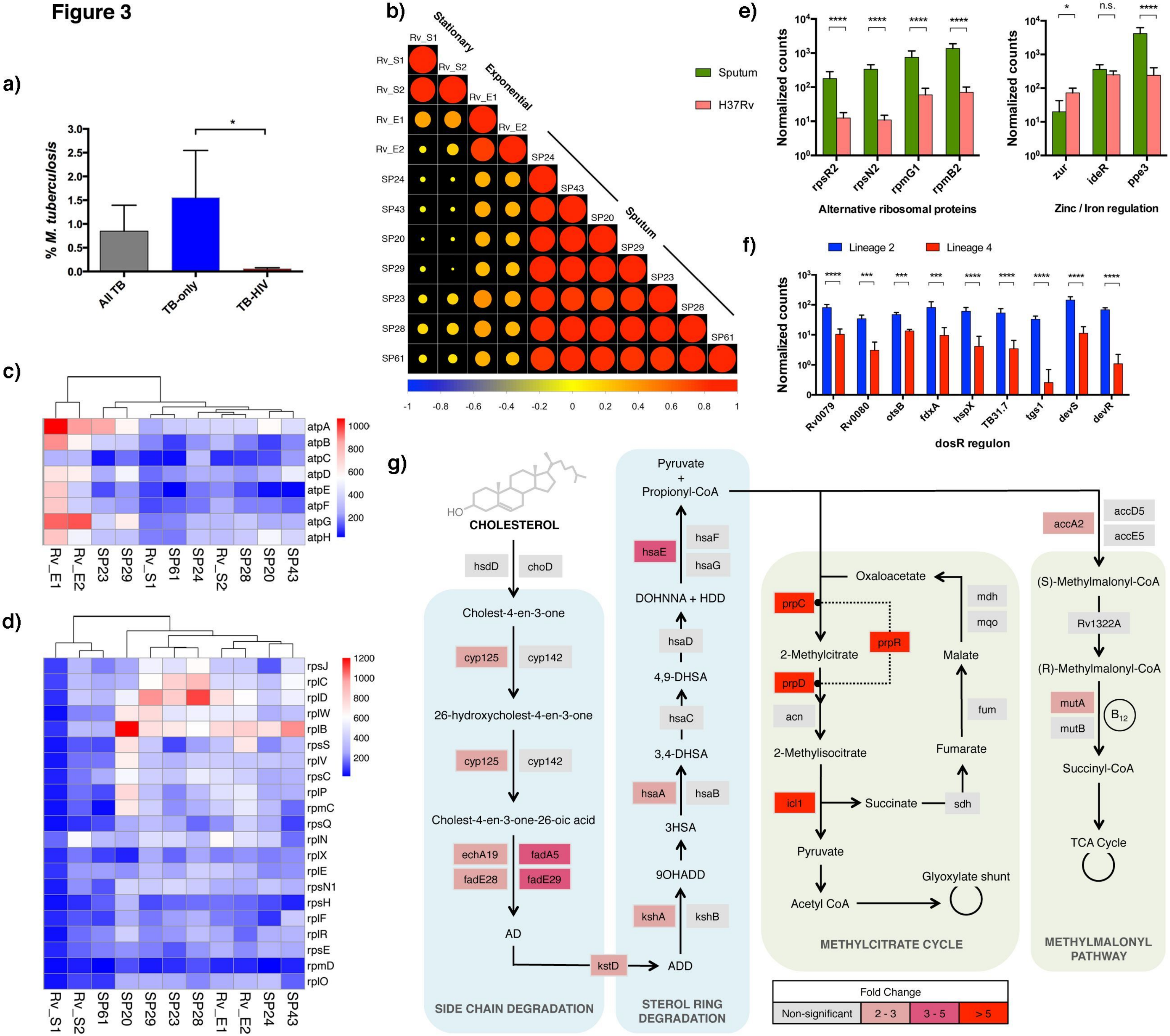

a)
Number of reads (%)
100
80
60
40
20
0
SP12
SP18
SP20
SP21
SP23
SP24
SP28
SP29
SP30
SP37
SP43
SP47
SP48
SP51
SP54
SP55
SP61
SP313
SP314
SP315
SP316
SP317
SP318
SP319
SP320
SP321
Streptococcus
Neisseria
Prevotella
Haemophilus
Veillonella
Leptotrichia
Campylobacter
Fusobacterium
Treponema
Rothia
Porphyromonas
Ralstonia
Thauera
Capnocytophaga
Bradyrhizobium
Streptobacillus
Rhodococcus
Aggregatibacter
Tannerella
Mycoplasma
b)
Observed OTUs
3000
2750
2500
2250
2000
**
All TB
Non-TB
Chao1
3000
2750
2500
2250
2000
**
All TB
Non-TB
c)
Shannon Index
8
6
4
2
0
n.s.
All TB
Non-TB
Simpson Index
1.25
1.00
0.75
0.50
0.25
0.00
n.s.
All TB
Non-TB
d)
Chao1
3000
2750
2500
2250
2000
n.s.
TB-only
TB-HIV
Shannon Index
8
6
4
2
0
n.s.
TB-only
TB-HIV

Figure 3


a)
% M. tuberculosis
3.0
2.5
2.0
1.5
1.0
0.5
0.0
*
All TB
TB-only
TB-HIV
b)
Stationary
Exponential
Sputum
Rv_S1
Rv_S2
Rv_E1
Rv_E2
SP24
SP43
SP20
SP29
SP23
SP28
SP61
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
c)
atpA
atpB
atpC
atpD
atpE
atpF
atpG
atpH
1000
800
600
400
200
Rv_E1
Rv_E2
SP23
SP29
Rv_S1
SP61
SP24
Rv_S2
SP28
SP20
SP43
d)
rpsJ
rplC
rplD
rplW
rplB
rpsS
rplV
rpsC
rplP
rpmC
rpsQ
rplN
rplX
rplE
rpsN1
rpsH
rplF
rplR
rpsE
rpmD
rplO
1200
1000
800
600
400
200
Rv_S1
Rv_S2
SP61
SP20
SP29
SP23
SP28
Rv_E1
Rv_E2
SP24
SP43
e)
Sputum
H37Rv
Normalized counts
10^4
10^3
10^2
10^1
10^0
****
****
****
****
rpsR2
rpsN2
rpmG1
rpmB2
Alternative ribosomal proteins
*
n.s.
****
zur
ideR
ppe3
Zinc / Iron regulation
f)
Lineage 2
Lineage 4
Normalized counts
10^3
10^2
10^1
10^0
0
****
***
***
***
****
****
****
****
****
Rv0079
Rv0080
otsB
fdxA
hspX
TB31.7
tgs1
devS
devR
dosR regulon
g)
HO
CHOLESTEROL
hsdD
choD
Cholest-4-en-3-one
cyp125
cyp142
26-hydroxycholest-4-en-3-one
cyp125
cyp142
Cholest-4-en-3-one-26-oic acid
echA19
fadA5
fadE28
fadE29
AD
kstD
ADD
SIDE CHAIN DEGRADATION
kshA
kshB
9OHADD
3HSA
hsaA
hsaB
3,4-DHSA
hsaC
4,9-DHSA
hsaD
DOHNNA + HDD
hsaE
hsaF
hsaG
Pyruvate
+
Propionyl-CoA
STEROL RING DEGRADATION
Oxaloacetate
prpC
prpR
2-Methylcitrate
prpD
acn
2-Methylisocitrate
icl1
Succinate
sdh
Fumarate
fum
Malate
mdh
mqo
Pyruvate
Acetyl CoA
Glyoxylate shunt
METHYLCITRATE CYCLE
accA2
accD5
accE5
(S)-Methylmalonyl-CoA
Rv1322A
(R)-Methylmalonyl-CoA
mutA
mutB
B_{12}
Succinyl-CoA
TCA Cycle
METHYLMALONYL PATHWAY
Fold Change
Non-significant
2 - 3
3 - 5
> 5