Intestinal microbiome composition influences the peripheral inflammatory state during treatment of human tuberculosis

Matthew F. Wipperman1,2,#, Shakti K. Bhattarai1,2,#, Charles Kyriakos Vorkas1,4, Venkata Suhas Maringati3, Ying Taur5, Laurent Mathurin6, Katherine McAulay7, Stalz Charles Vilbrun6, Daphie Francois6, James Bean1, Kathleen F. Walsh7, Carl Nathan8, Daniel W. Fitzgerald7, Michael S. Glickman1,4,8,*, Vanni Bucci3,9,10,*

Affiliations:
1Immunology Program, Sloan Kettering Institute
2Clinical and Translational Science Center, Weill Cornell Medicine
3Department of Microbiology and Physiological Systems, University of Massachusetts Medical School
4Division of Infectious Diseases, Weill Cornell Medicine
5Division of Infectious Diseases, Memorial Sloan Kettering Cancer Center
6Haitian Study Group for Kaposi’s Sarcoma and Opportunistic Infections (GHESKIO), Port-au-Prince, Haiti
7Center for Global Health, Weill Cornell Medicine
8Immunology and Microbial Pathogenesis Graduate Program, Weill Cornell Graduate School
9Center for Microbiome Research, University of Massachusetts Medical School
10Program in Systems Biology, University of Massachusetts Medical School
# these authors contributed equally

*Correspondence to:
Vanni Bucci, PhD
Microbiology and Physiological Systems
Program in Systems Biology
Center for Microbiome Research
University of Massachusetts Medical School
368 Plantation Street
Worcester, MA 01605
774-856-2215
vanni.bucci2@umassmed.edu

Michael S. Glickman, MD
Immunology Program, Sloan Kettering Institute
1275 York Ave
New York, NY 10065
6468882368
glickmam@mskcc.org

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.
Abstract

Although the composition of the intestinal microbiota influences systemic immune responses, the contribution of this relationship to infectious disease pathogenesis and to the resolution of infectious diseases by antibiotic therapy is poorly understood. This question is rarely examined in humans due to the difficulty in dissociating the immunologic effects of antibiotic-induced pathogen clearance and salutary microbiome alteration. To address these questions in the context of Tuberculosis, we analyzed three independent human datasets from Haiti (two longitudinal treatment and one cross-sectional control), a prototypical infection remarkable for chronic inflammation, in which we had measured sputum TB bacterial load, gut microbiota composition, and peripheral blood transcriptomics. Using data from the longitudinal datasets combined with inflammatory pathway enrichment analysis, we determined that antibiotic treatment of TB, despite significantly perturbing the gut microbiota, dampens the proinflammatory signature characteristic of active TB. Contrarily, an investigational TB treatment that failed to clear TB, but that caused similar microbiota perturbations, exacerbated peripheral inflammation. To decouple the effects of antibiotic induced changes in the microbiota from Mtb sterilization as predictors of normalization of TB associated inflammation, we applied random forest regression to the microbiome-transcriptome-sputum data from the two longitudinal datasets. We found inflammatory renormalization is positively affected by both pathogen sterilization and by the abundance of health-associated Cluster IV and Cluster Xla Clostridia. Oppositely, increases in the abundance of commonly known pathobionts such as Bacilli and Proteobacteria clusters predict inflammatory exacerbation. We independently investigated and validated these microbiota-peripheral inflammatory signature associations by applying machine learning to the peripheral gene expression and microbiota profiling in an independent human cohort of 52 healthy control individuals. Together, our findings indicate that antibiotic-induced reduction in pathogen burden and changes in the microbiome are independently associated with treatment-induced changes of the inflammatory response of active TB, and more broadly indicate that response to antibiotic therapy may be a combined effect of pathogen killing and microbiome driven immunomodulation. Our results provide support to the hypothesis that there exists clear links between microbiome composition and host peripheral gene expression in humans that can be biologically elucidated using common and well validated molecular pathway analyses.
Abstract (short)

Causal relationships between microbiome composition and host peripheral gene expression have been studied and validated extensively in mice but have not been demonstrated in humans. In this work, we investigate microbiome-periphery gene expression relationships in three unique human cohorts from Haiti, in the context of a multicenter TBRU collaboration. The objectives of this work were to quantitate the relationships in two longitudinal and one cross sectional cohorts of Tuberculosis treatment. To do this we applied machine learning methods for high-dimensional data to learn associations in humans between these commonly profiled tissues. Collectively, our results provide support to the hypothesis that there exists clear links between microbiome composition and host peripheral gene expression, that can be biologically elucidated using common and well validated molecular pathway analysis.
Introduction

There is mounting evidence that the gut microbiome has an important role in the modulation of host physiology, with a wealth of studies having associated microbiome composition and functions with differential inflammatory, neurological, and even behavioral activity. Gastrointestinal colonization by specific taxa with particular metabolic capacities has been shown to differentially modulate host biology. For example, colonization by a subset of Clostridia enhanced anti-inflammatory phenotypes in mice, and enrichment in specific members of the Bacteroides and Parabacteroides genera induced CD8+ T cell responses and anticancer activity in mice and marmosets, as well as correlating with the abundance of these immune effectors in humans. A multitude of experiments in mice have allowed for the determination of mechanisms by which intestinal mucosal-associated bacteria affect host physiology at the epithelial interface and systemically throughout their host.

Despite these observations, it is unknown whether, and to what degree microbiome changes are responsible for systemic changes in inflammatory responses in humans. This knowledge gap is due in part to the difficulty of isolating the microbiome dependent effects from other aspects of human physiology and in discerning the direction of causality in human studies. As microbial communities in the gut promote the development and maintenance of innate and adaptive immune responses, including microbiota-educated immune cells and many small molecules that circulate throughout the periphery, we would expect to observe both localized and systemic host effects upon major microbiome alterations such as treatment with antibiotics, which we can measure using technologies such as shotgun RNA sequencing (RNAseq).

Infection by Mycobacterium tuberculosis (Mtb) (the 9th leading cause of death on Earth) has been shown to have a markedly different systemic gene expression profile compared to people with latent disease, other respiratory diseases, or no known infection. Specifically, infection with Mtb leads to heightened expression of inflammatory pathways, most notably the Type I and Type II interferon pathways, with this pattern resolving with antibiotic therapy. A recent meta-analysis combining microarray and RNAseq data from studies aimed at identifying active TB transcriptional signatures, confirmed the findings about a specific set of peripheral blood transcripts that are biomarkers of active TB disease, relative to healthy individuals or those with latent TB infection (LTBI). Antibiotic treatment for active TB involves combination therapy with narrow spectrum and prodrug agents with mostly Mycobacterial-specific targets, (HRZE) is given for two months and is then followed by an HR-only administration for an additional four months, in order to achieve over 95% likelihood of Mtb clearance. The disruptive effect of HRZE therapy on the intestinal microbiome was
demonstrated in a longitudinal study in mice and cross-sectional study in humans, which indicated that the major phyla perturbed are from the class Clostridia, a group of obligate anaerobes in the gut with well described immunomodulatory effects on the host.

Given that HRZE treatment causes GI microbiota shifts that include the depletion of many Clostridia species, and given the role that these species play in modulation of host biology in mice and humans, we reasoned that there could be a connection between the microbiome alterations observed during HRZE therapy and the resolution of systemic inflammatory responses to TB. However, because HRZE therapy rapidly reduces the burden of Mtb in the early phase of treatment, it is difficult to uncouple the immunologic effects of pathogen killing from microbiota perturbation without a control group that has either pathogen killing or microbiome perturbation, but not both.

**Figure 1: Overview of cohorts, timepoints, samples, and hypothesis in this analysis.** A. This study investigates microbiome-transcriptome relationships in three separate cohorts of individuals in Haiti. Cohort 1 (longitudinal) consists of a randomized clinical trial of study volunteers, where we collected disease severity measurements (TTP), microbiome profiling, and peripheral transcriptomics at baseline, before active TB patients were randomized to either HRZE (standard of care TB treatment), or NTZ. Cohort 2 (longitudinal) consists of study volunteers who were followed throughout the course of 6 months of TB treatment, where we collected TTP, microbiome, and transcriptomics data. Finally, Cohort 3 (cross sectional) consists of healthy volunteers. These
healthy volunteers were enrolled separately through the Tri-I TBRU. Around half are healthy and TB-negative household contacts of active TB-patients, and the other half are community controls, with no know TB exposure. We performed microbiome profiling and peripheral transcriptomics on these individuals as well. B. Numbers of individuals in this study. C. Causal inference diagram demonstrating the major hypothesis this study tests.

To address this problem, we combined three independent clinical datasets for which we had gathered microbiome profiling via 16S rRNA sequencing, peripheral blood transcriptomics, and where relevant *Mtb* abundance in the sputum. The first dataset (Figure 1A,B) consists of the secondary endpoint data from a clinical trial (NCT02684240) that compared the early bactericidal effect (EBA) of standard tuberculosis (TB) therapy isoniazid (H), rifampin (R), pyrazinamide (Z), and ethambutol (E) (HRZE) to the antiparasitic drug nitazoxanide (NTZ), shown to possess antimycobacterial activity *in vitro*\(^{25,26}\). We thought this dataset to be ideal for testing our hypothesis because (1) the clinical trial demonstrated that as opposed to HRZE, NTZ did not show any antimycobacterial activity in humans\(^{26}\) (Figure 2B), and (2) preliminary microbiome analysis by our group showed that NTZ caused microbiome dysbiosis similar to that caused by HRZE, thus providing us with the desired dysbiosis in the absence of any pathogen killing or aggravation of disease. We then found that HRZE and NTZ had distinct effects on host peripheral gene expression, with HRZE causing a resolution of inflammatory gene signatures, and NTZ causing an exacerbation of them in the absence of disease progression. Because we had used secondary endpoint observations from a clinical trial dataset that was powered for its primary endpoint (reduction of *Mtb* in the sputum) and that only spanned a two-week treatment time frame, we additionally collected data from an independent 6-month longitudinal HRZE treatment cohort (Figure 1A,B). This cohort not only shows similar resolution of disease with treatment compared to the two-week HRZE arm, but also allowed us to characterize for the first time the long-term effects HRZE treatment on the intestinal microbiota and peripheral gene expression. We took advantage of these datasets to train Random Forest Regression models to assess the changes in expression of peripheral inflammatory pathways as a function of changes in microbiota species abundances and simultaneous changes in *Mtb* in the sputum. Our modeling predicted a number of bacteria to differentially affect proinflammatory responses, which were consistent with previously published experimental reports. Specifically, we found that the increase in abundance of SCFA-producing Clostridia species predicts inflammatory renormalization; oppositely, we found that an increase in abundance of oxygen-tolerant pathobionts such as *E. coli*, *Klebsiella*, *Enterococcus*, and *Streptococcus* species associate with exacerbation of proinflammatory pathways (Figure 1C). Finally, we investigated these microbiota-peripheral gene expression relationships using a cohort of healthy Haitian community controls and healthy household contacts of TB patients, previously described\(^5\) (Figure 1A,B) and confirmed the existence of these microbiome-peripheral immune gene expression signatures. Overall, we believe that these results provide support to the often stated...
hypothesis that there exists clear regulatory relationships between gut microbiome composition and peripheral gene expression, at both the immune, and gene-pathway regulatory levels in humans.
Results

Gut microbiome diversity is depleted after two weeks of HRZE or NTZ treatment

As detailed elsewhere, the GHESKIO centers in Port au Prince, Haiti conducted a prospective, randomized, early bactericidal activity (EBA) study in treatment-naive, drug-susceptible adult patients with uncomplicated pulmonary tuberculosis (TB) (ClinicalTrials.gov Identifier: NCT02684240)\textsuperscript{26}. Thirty-four participants were randomized to receive either NTZ, 1000 mg po (oral) twice daily, or standard oral therapy with isoniazid 300 mg daily, rifampin 600 mg daily, pyrazinamide 25 mg/kg daily, and ethambutol 15 mg/kg daily (referred to as HRZE) for 14 days (Figure 1A). The primary endpoint of the trial was sputum bacterial load (measured by time to culture positivity, TTP) in a BACTEC liquid culture system, a sensitive method to detect disease progression and severity in TB patients. Sputum was collected from 6pm to 9am every other day to quantify mycobactericidal activity of each treatment regimen.

HRZE resulted in a predictable increase in the TTP (corresponding to reduced bacterial load) over the first two weeks of therapy compared to baseline TTP ($p < 0.001$ for the linear mixed-effect model $TTP \sim Sex + Age + Time + Treatment + Time:Treatment + 1|ID$ where $Time:Treatment$ is the interaction term and $1|ID$ is the subject-level random effect. NTZ was used as the reference level for treatment (see Methods and Supplementary Table S1) (Figure 2B). NTZ, despite its potent in vitro activity\textsuperscript{27}, did not have any significant effect on TTP after 14 days ($p > 0.05$) (Figure 2B, Supplementary Table S1)\textsuperscript{26}. This lack of NTZ efficacy was traced to a failure of the drug to penetrate the sputum,\textsuperscript{26} and indicated that disease severity remained constant as measured by mycobacterial load. All patients were subsequently switched to HRZE standard of care treatment.

We have previously reported\textsuperscript{22} that HRZE therapy depletes members of the order Clostridiales, but the cross-sectional design of that study did not allow for conclusions about the rapidity of this effect, and most importantly, did not include pretreatment samples to allow for the assessment of baseline microbiome composition. To investigate microbiome changes induced by HRZE or NTZ, we extracted and amplified bacterial and archaeal DNA using V4 – V5 16S rDNA sequencing (see Methods). Stool samples were collected before the start of treatment (baseline) and on day 14 of therapy (Figure 1A). Using principal components analysis (PCoA) with Bray-Curtis distances, we found that the component accounting for the greatest variation in the microbiome data qualitatively represented changes in microbiome community structure that occur after two weeks of NTZ treatment (Figure 2C). Inspecting Axis 2 of Figure 2C, we found that the observed separation correlates with sequencing batch. Therefore, for any subsequent statistical modeling analysis (i.e., differential
microbiota and gene expression modeling), sequencing batch information has been controlled for by including it as fixed effect in the modeling.

To compare the effect of the two treatments to microbiome alpha diversity we calculated the Inverse Simpson Diversity Index for every microbiome sample. We then regressed the Shannon Diversity Index via linear mixed-effect modeling as $\text{Diversity} \sim \text{Sex} + \text{Age} + \text{Batch} + \text{Treatment} + \text{Time} + \text{Treatment:Time} + 1|\text{ID}$ (see Methods). We found that there were no differences in alpha diversity at baseline between the two arms, while both treatments significantly reduced microbial diversity ($p<0.01$, see Supplementary Table S1), with NTZ treatment not having a significantly different effect compared to HRZE ($p>0.05$, for the interaction term, See Supplementary Table S1) (Figure 1D).

**Figure 2:** Both HRZE and NTZ perturb the gut microbiota after two weeks of therapy. A. Set up for the clinical trial comparing bactericidal effect of HRZE and NTZ. B. Paired time to positivity (TTP) at day 0 and day 14 for the NTZ treatment cohort and HRZE treatment cohort. Data are displayed as range of three technical replicates. Linear mixed effect modeling was used to determine significance of difference in pre/post treatment in each arm as $\text{TTP} \sim 1 + \text{Sex} + \text{Age} + \text{Treatment} + \text{Time} + \text{Treatment:Time} + 1|\text{ID}$, where Treatment indicates the arm (NTZ or HRZE), Time indicates Pre or Post antibiotic administration, and : indicates the interaction term. NTZ treatment is associated with no difference in TTP between day 0 and day 14 ($p > 0.05$, see Supplementary Table S1), and thus no change in disease status, while HRZE is significant ($p < 0.05$, Supplementary Table S1), and thus a reduction in disease severity. Data for TTP where obtained from Walsh et al. 2020. C. Principal Coordinate analysis (PCoA) with Bray-Curtis distance showing differences in microbiome community structure between individuals before and after 14 days of either HRZE or NTZ treatment. The grey line connects baseline and day 14 treatment paired samples. PCoA1 clearly discriminates samples post NTZ treatment (pink triangles) from those at baseline or after HRZE treatment. D. Microbiota alpha diversity plotted using the Inverse Simpson Diversity index. Linear mixed effect modeling was used to determine the significance of difference treatment of diversity. We fitted the model $\text{Diversity} \sim 1 + \text{Sex} + \text{Age} + \text{Treatment:Time} + 1|\text{ID}$. The symbol : indicates the interaction term. HRZE was used as the reference level.
No significant difference between the two treatment at baseline was observed. Both groups display significantly reduced Inverse Simpson diversity after 14 days of treatment (p < 0.05, Supplementary Table S1).

**Taxonomic alterations in microbiome composition induced by antibiotics are more pronounced in NTZ-treated individuals.**

To identify phylotypes significantly affected by each of the two treatments, we modeled the abundance of each sequencing variant identified via dada2 (ASV) via linear mixed-effect modeling as of $ASV_i (counts) \sim Sex + Batch + Group + 1|ID$ using Limma/Voom (see Methods). This model statement enables quantifying sex and sequencing batch-dependent effects in addition to establishing effects that are due to treatment group (pre-treatment, HRZE, NTZ). We used the $1|ID$ random effect to control for baseline differences among individuals. ASVs significantly affected by the treatments were determined using a Benjamini-Hochberg false discovery rate (FDR) adjusted p-value of 0.05 (see Methods).

By running this analysis, we found that the HRZE effect on the intestinal microbiota consisted almost exclusively in the depletion of 82 ASVs, many belonging to genera from the Order Clostridiales (FDR $p < 0.05$) (Supplementary Table S2). Notably, many of these ASVs (e.g., *Blautia* spp., *Butyrivibrio* spp., *Clostridium* spp., *Eubacterium* spp., *Faecalibacterium* spp., *Gracilibacter* spp., *Oscillibacter* spp., *Roseburia* spp., *Ruminococcus* spp., *Sporobacter* spp.) are known to be involved in a number of health-associated functions such as SCFA production or bile acid transformation (Figure 3A, 3B). NTZ was instead found to have a much larger disruptive effect by depleting 387 ASVs and increasing 16 ASVs. NTZ caused a reduction of a larger swath of Clostridiales, which included all but one of the ASVs depleted by HRZE, as well as many other additional Firmicutes, (FDR $p < 0.05$). Additionally, NTZ caused an expansion of known oxygen-tolerant pathobionts including, *K. pneumoniae*, *E. coli*, *C. freundii*, *S. alactolyticus*, and *E. faecium* ($p < 0.01$) (Figure 3A, 3B).

Taken together these results demonstrate that NTZ, despite having no effect on TTP or disease severity, causes a perturbation of the intestinal microbiota which is more pronounced compared to that caused by HRZE. This includes the depletion of a large number of Clostridia, and the selection for known disease-associated, oxygen-tolerant pathobionts.
Figure 3: Overlapping and distinct microbiome perturbation induced by NTZ and HRZE. A. Volcano plots indicating the post (day 14) vs pretreatment (baseline) differences at the ASV level for HRZE and NTZ. The color of each ASV is according to the phylogenetic Order. A single linear mixed effect model for each ASV of the form $\text{ASV}_i(\text{counts}) \sim \text{Sex} + \text{Batch} + \text{Group} + 1|\text{ID}$ was fitted to determine differences due to treatment while accounting for sequencing batch and sex. The horizontal dotted lines indicate FDR < 0.05 and vertical dotted lines indicate $|\log2\text{FC}| > 1.5$. B. Within-arm unsupervised hierarchical clustering of the abundances of 404 ASVs found to be significantly affected by HRZE or NTZ treatment (FDR < 0.05). The heatmap columns are split by arm membership (including baseline randomization group), and the heatmap rows are split by ASV phylogenetic Phylum, and within the Phylum, the Order is colored as in A. The right annotations (HRZE and NTZ) indicate whether each ASV was significantly perturbed by treatment. $P$ value in $y$ axis is adjusted according to Benjamini-Hochberg.

HRZE and NTZ uniquely affect host peripheral gene expression

As the above results highlighted a differential effect on TB disease severity and intestinal microbiome composition of the two antimicrobials tested, we recognized that this presented a unique opportunity to infer possible relationship between microbiome composition, $Mtb$ bacterial load, and peripheral gene expression. As the patients were randomized at baseline before being assigned to the two treatments, we used linear mixed-
effect modeling with Limma/Voom to model the abundance of each host transcript as $Gene_i \ (Counts) \sim Sex + Batch + Group + 1|ID$ (See Methods).

We determined the functional pattern of treatment-induced changes in overall transcript abundance by performing gene set enrichment analysis (GSEA)$^{33,34}$ (See Methods) on the ranked limma/voom expression data for the baseline vs post-NTZ or post-HRZE treatment contrasts. We used GSEA to estimate enrichment for the MiSigDB Hallmark pathways$^{33,35}$, which are intended to give a broad overview of biological pathways that may be expressed. In individuals undergoing HRZE treatment, we observed a significant (FDR < 0.05) depletion at day 14 of inflammatory response, IFNα response, IFNγ response, TNFα signaling via NFκB, and IL6 JAK STAT3, all of which are consistent with the broad immunologic effects of antibiotic induced reduction in the levels of a bacterial pathogen$^{14,17,36}$ (Figure 4A). In contrast, and to our surprise, NTZ treatment showed the opposite effect. Inflammatory signaling pathways reduced by HRZE, including TNFα signaling, IFNγ signaling, and type 1 interferon signaling were all significantly enriched by NTZ treatment at day 14 (Figure 4A). Several other pathways such as hypoxia, apoptosis, and reactive oxygen species (ROS) that are considered hallmarks of immune dysregulation$^{37}$, were also enriched by NTZ treatment. As NTZ was found to perturb the microbiome while keeping Mtb bacterial load in the sputum (TTP) substantially unchanged, we hypothesized that the NTZ effect was likely not completely due to prognostic disease progression, and rather at least a partial function of microbiome alteration (Supplementary Table S6).
Figure 4: Hallmark pathway gene set enrichment analysis and gene expression comparison in HRZE and NTZ treated cohorts. A. Hallmark gene pathway changes associated with 2 weeks of HRZE (A) or NTZ (B). Positive are pathways overrepresented at 2 weeks of therapy (up), and negative are pathways underrepresented at 2 weeks (down), both compared to baseline. All pathways are significant (FDR<0.05) with the size of the arrow indicating significance. Only pathways significant in this analysis are shown from the MiSigDB Hallmark pathway set. B, C. TB-associated genes from the Berry, et al., meta-analysis14, highlighting post treatment vs baseline changes in gene expression for HRZE (B) or NTZ (C). Notably, HRZE renormalizes (i.e., towards a healthy control state) the expression of 144 validated TB inflammatory transcripts and exacerbates only 13. NTZ is found only to exacerbate four. D, E. Effect of HRZE and NTZ on blood gene expression for a set of IBD-associated genes from Palmer, et al.38. Both drugs, HRZE in D and NTZ in E cause different genes to either renormalize (HRZE 66, NTZ 34) or exacerbate (HRZE 55, NTZ 21). P value in y axis is adjusted according to Benjamini-Hochberg.

To this point our analysis has shown that while both antimicrobials were causing disruptive microbiome effects, HRZE was killing Mtb and partially renormalizing peripheral inflammatory responses, while NTZ was having the opposite effect, in the absence of any change to Mtb bacterial load. To gain a deeper understanding of gene signatures affected by each of the two drugs, we first focused our analysis on a set of validated transcriptomic markers of active TB from a meta-analysis of a large number of well-regarded TB studies (mentioned above), and examined 373 transcripts that have been associated and validated in multiple human cohorts on different sequencing platforms (microarray and RNAseq) to be differentially abundant between LTBI, active TB, and healthy control individuals19. In our study, we detected 361 of these 373 transcripts in pretreatment, active TB subjects. We defined three classes of changes to these transcripts with two weeks of HRZE or NTZ treatment: 1) renormalization (transcripts whose pre-post HRZE/NTZ fold change in expression displays the same sign (or direction) of the previously-reported fold-change between active TB and control/LTBI; 2) unchanged (transcripts with no change in expression between pre-post HRZE/NTZ administration); and 3) exacerbation (genes whose pre-post HRZE/NTZ fold-change sign is opposite to the previously-reported fold-change between active TB and control/LTBI). 157 of these active TB signature genes were significantly affected by HRZE (FDR < 0.05) (Figure 4B, Supplementary Table S3). Of these 157 affected genes, 144 (92%) were found to renormalize with the treatment (i.e., displaying the same direction of the fold change reported for active TB vs. control individuals), while 13 (8%) were found to exacerbate (i.e., opposite direction of the fold change). On the other hand, only four of these TB-related inflammatory genes were found to be affected by NTZ, and all of them (100%) displayed an exacerbation trend (Figure 4C).

Because of the disruption in microbiome composition in the absence of a change in disease severity observed in NTZ-treated people, we hypothesized that there may be other subsets of host genes that are linked to microbiome-dependent immunity that could be responsive to the observed antimicrobial perturbations. To test this, considering the already well-established link between microbiome dysbiosis and autoinflammatory
conditions such as Inflammatory Bowel Disease (IBD)\textsuperscript{39}, we selected a recently published panel of 880 genes differentially expressed in colon biopsies from IBD patients and asymptomatic controls\textsuperscript{38}. Of these 880 genes, 364 were detected in our dataset (expected given that we are profiling whole blood transcriptomics, rather than colon biopsy tissue). Despite the more limited effect on the microbiome compared to NTZ, HRZE administration was found to be responsible for a change in 117 of these genes ($FDR < 0.05$) (Figure 4D, Supplementary Table S3), while NTZ was found to be responsible for a change in 55 of them ($FDR < 0.05$) (Figure 4E, Supplementary S3). Similarly, we defined as renormalization those genes that have a post/pre fold change due to antimicrobial treatment of the same sign as control/active IBD. Interestingly, while NTZ affected a smaller subset of these IBD immune genes, it caused greater exacerbation compared to HRZE (77.8\% vs 42.0\% genes exacerbating).

**Longitudinal profiling in an early bactericidal activity HRZE treatment cohort**

To validate the observations obtained after two weeks of HRZE and to identify effects that may occur with increased treatment length we enrolled a longitudinal treatment cohort (20 individuals) to measure EBA over the course of the 6 months of HRZE treatment, with periodic sampling of sputum for TTP, stool for microbiome profiling, and whole blood for peripheral RNA sequencing. All participants received standard of care HRZE therapy (isoniazid 300 mg daily, rifampin 600 mg daily, pyrazinamide 25 mg/kg daily, and ethambutol 15 mg/kg daily). Sputum mycobacterial load (TTP) was collected at baseline, day 7, day 14, one month, two months, and six months at the completion of treatment. Stool from microbiome profiling was collected at each of these timepoints as well. Whole blood was collected at baseline, day 14, and two months. Stool and peripheral blood samples were processed for microbiome and transcriptomic profiling as for the above described clinical trial (See Methods).

Similarly, to what we observed in the HRZE arm of the clinical trial, TTP is significantly higher compared to baseline after two weeks of treatment ($p < 0.05$). TTP also significantly increases after two months compared to the two weeks timepoint confirming the presence of an additional sterilization effect of the antibiotic on $Mtb$ ($p < 0.05$, linear mixed-effect modeling, see Methods) (Figure 5A). Interestingly, microbiome diversity drops after just 7 days of treatment, increases after 1 month of treatment but remains significantly lower compared to baseline at the 6-month follow-up timepoint ($p < 0.05$, linear mixed-effect modeling, see Methods) (Figure 5B).

We performed Limma/voom differential analysis to determine effects of treatment and time on the microbiome and peripheral host transcriptomics using the same approach as above. As observed in the clinical trial and in our previous work\textsuperscript{22}, HRZE was found to depress Clostridiales after one week of treatment, with most of these
Clostridiales remaining significantly depressed compared to baseline, even at the 6-month follow-up time point (Figure 5C). At day 7 compared to baseline, 19 ASVs were depleted and 6 were increased in abundance, at day 14 compared to baseline, 74 ASVs were depleted and 2 were increased in abundance, at one month compared to baseline, 97 ASVs were depleted and none were increased in abundance. Thus, during the first month of treatment, microbiome depletion relative to individual baseline samples was evident, however, later in the course of treatment, we observed a different trajectory. Relative to baseline, at two months of HRZE, we observed 13 ASVs depleted in abundance, while 95 ASVs increased in abundance. At the 6-month mark of HRZE treatment, we observed 3 ASVs depleted in abundance while 96 ASVs increased in abundance (Figure 5C). The increased abundance of specific ASVs at the two month and 6-month mark appeared to be relatively heterogeneous between individuals, and overall, this longitudinal analysis suggests that after one month of HRZE therapy, most ASVs that would be affected by therapy are depleted (Supplementary Table S4).

With respect to the peripheral host inflammatory profiling, we observed distinct changes in gene signatures at two weeks (day 14) and two months of treatment, compared to baseline (Supplementary Table S6). We observed a similar decrease in common inflammatory pathways in the Hallmark pathway dataset described in the clinical trial (Figure 5D). Interestingly, comparing Day 56 to baseline or to Day 14, we see additional reduction in inflammatory gene signatures, potentially explained in part by the further reduction in TTP at Day 56 (Figure 5E-I). Collectively, this EBA dataset confirms many of the findings observed in the HRZE arm of the trial, comparing Day 14 to baseline in both cases, and adding additional treatment information at these later timepoints. Importantly, it can be used in addition to the clinical trial data for training the host-microbiome- Mtb model described subsequently.
Figure 5. Longitudinal profiling of HRZE treatment induced changes of microbiome composition and peripheral gene inflammatory expression. A. Time to positivity was measured at baseline, day 14, one month, and two months. B. Microbiome diversity was computed for each study volunteer at baseline, day 7, day 14, one month, two months, and 6 months. Microbiome $\alpha$ diversity was measured using the inverse Simpson index$^{40}$. C. Volcano plots showing significance of differences in microbiome composition vs fold change from baseline at Day 7, Day 14, Day 30, Day 56, and Day 180. D. Normalized enrichment scores calculated for the
Hallmark Pathway list for Day 14 vs Baseline, Day 56 vs Day 14, and Day 56 vs Baseline. E. Volcano plot showing transcripts from Berry et al.14, at Day 14 vs Baseline. F. Volcano plot showing transcripts from Berry et al.14, at Day 56 vs Day 14. G. Volcano plot showing transcripts from Berry et al.14, at Day 14 vs Baseline. H. Volcano plot showing transcripts from Palmer et al.38 of IBD cases vs controls in this study for Day 14 vs Baseline. I. Volcano plot showing transcripts from Palmer et al.38 of IBD cases vs controls in this study for Day 56 vs Day 14. J. Volcano plot showing transcripts from Palmer et al.38 of IBD cases vs controls in this study for Day 56 vs Baseline.

**Multi-omics-constrained mathematical modeling to decouple the contribution to peripheral inflammatory signature of intestinal microbiota and Mtb.** We next sought to determine the relative contribution of intestinal microbiota and Mtb dynamics in influencing peripheral inflammatory gene signatures. Specifically, our aim was to identify microbiota members whose abundances is predictive of significant changes in inflammatory pathways in our three groups of antibiotic-treated individuals. To do this, for each inflammatory hallmark pathway l identified to be significantly affected by HRZE (Clinical Trial or Long-term EBA study) or NTZ via linear mixed effects modeling, we first computed the change in enrichment score between two consecutive time points $t_{\psi}$ and $t_{\psi+1}$ for individual $s$ as $rac{y_{t_{\psi+1}s}-y_{t_{\psi}s}}{t_{\psi+1}-t_{\psi}}$. We then regressed this quantity against the corresponding fold change in abundance of for every ASV $v$ in the same interval $\frac{x_{v_{t_{\psi+1}s}}}{x_{v_{t_{\psi}s}}}$ and against the corresponding fold change in TTP, $\frac{p_{t_{\psi+1}s}}{p_{t_{\psi}s}}$. Using change from baseline values accounts for the random effect of each subject without having to incorporate this into the model statement. We solved this regression problem using Random Forest regression as in41. To train the models we used all the observations from the clinical trial and from the longitudinal EBA cohort for a total of 34 paired samples (Figure 1). We fit a model for each inflammatory using all the data from the three patients’ group (HRZE clinical trial, NTZ clinical trial and HRZE EBA) because we wanted to find patterns that are general across multiple datasets. Each model was trained using 5000 trees and a with train-validation partitioning of 80-20% of the data. We reasoned that this approach was appropriately suited for this type of “large p, small n” multi-omics dataset common in clinical research42. Importantly RFR modeling has significant advantages compared to traditional multi-linear regression techniques, because it is agnostic to model structure (e.g. non-parametric regression), it does not need to meet common assumptions underlying classical regression techniques and is able to intrinsically perform ranked feature selection. Importantly, while the interpretation of RFR is apparently less immediate compared to traditional regression (e.g. there are per-se no regression coefficients or betas), downstream analysis, which includes Permutated Importance43 and Accumulated Local Effects (ALE) calculations44 (see Methods) allows for the estimation of the significance of predictors (e.g. TTP, microbiome constituents, etc.) and of their effects on the dependent variable (e.g. host peripheral inflammatory markers).
When plotting the average slope of the ALE curves for predictors with significant Permutated Importance value ($p < 0.05$), our analysis identifies the increase in TTP (and therefore a decrease of *Mtb* in the sputum) and the increase in abundance of ASVs from Clostridia, especially members from the Cluster IV and XVIa groups\(^{45}\), which has been shown to induce anti-inflammatory responses (e.g. Treg-induction)\(^{2,3}\) and also comprehends SCFA-producing species\(^{46}\) including *E. rectale*, *F. prausnitzii*, *G. formicilis*, *E hallii*, *O ruminantium*, *D. formicerans*, *S. variable*, *B. faecis* and *B. obeum* to promote a reduction in peripheral proinflammatory response including INF\(\gamma\), INF\(\alpha\), Inflammatory Response, IL6 JAK STAT3 Signaling (Figure 6). Oppositely, an increase in the abundance of oxygen-tolerant pathobionts including *E. coli* and *E. faecium* are found to be associated with promoting inflammatory exacerbation (Figure 6) which consistent with a large body of literature demonstrating that gastrointestinal overgrowth of these species is the hallmark of gastrointestinal dysbiosis and inflammation\(^{47,48}\) and often associates with adverse clinical outcome\(^{49,50}\).

Taken together, our data and related computational analyses show that the changes in inflammatory gene expression that accompany treatment of TB is predicted by both the anti-microbial activity of the drugs that lead to pathogen clearance and by antibiotic induced changes in microbiome composition. This modeling result suggests two modules of microbiome-inflammatory effects. The first is the exacerbation of TB associated inflammation by depletion of Clostridia (especially Cluster IV and XIVa), which is evident in both the HRZE and NTZ groups. Additionally, the enhancement of pathobionts such as *E. faecium* and *E. coli*, which only occurs with NTZ, also exacerbates inflammatory pathway expression within an individual. Based on this modelling, we predict that successful disease resolution may be associated with preservation of Clostridia, whereas their depletion and consequent enhancement of dysbiosis-associated Proteobacteria and Bacilli pathobionts might slow resolution or even support inflammatory exacerbation\(^{51}\).
Figure 6. Results from Random Forest Regression Modeling to predict immune-related peripheral blood gene signatures as a function of changes in intestinal microbiota and TTP. The heatmap displays the sign of the derivative of the ALE curve (See Text). Blue/orange entries indicate features found to significantly associate with changes in a specific inflammatory pathway. Blue indicates a negative relationship, while orange a positive. Black dots are used to identify the top important predictor (according to permuted importance analysis) for each specific host pathway. This analysis shows that reduction in TB burden and positive increase in the abundance of health associated Cluster IV and XIVa Clostridia are predictor of inflammatory dampening. Oppositely increase in abundance of oxygen-tolerant pathobionts including Enterococcus, Streptococcus, and E. coli is found to predict inflammatory exacerbation.
Relationship of the microbiome and peripheral gene expression in a healthy control validation cohort

The results from our machine learning modeling on the data from both longitudinal treatment cohorts provide support to the hypothesis that specific intestinal microbiota members are associated with immune-related peripheral blood gene signatures in humans. Specifically, our modeling predicts that higher abundance of Clostridia is negatively associated with inflammation (e.g. INFα, INFγ, IL6/JAK/STAT3, Inflammatory Response gene signatures) while higher abundance of commonly known oxygen-tolerant pathobionts promotes exacerbation of these signatures. To assess the generality of these findings we hypothesized that, even in healthy individuals, different levels of colonization by these bacteria would correspond to different levels of immune-related peripheral blood gene signatures.

To test this hypothesis and to ultimately validate the finding from the modeling, we analyzed a set of human data from two healthy control cohorts. A subset of these data was previously reported in previous work from us\(^5\), and come from a cross-sectional study of TB negative healthy household contacts of active pulmonary TB patients (termed Family Contacts, FC) and healthy unexposed donors from the same community in Haiti (termed Community Controls, CC) (see Methods). For these two cohorts we have a total of 52 healthy control individuals (18 FC and 36 CC) for which we gathered both microbiome 16S rRNA sequencing data and peripheral blood transcriptomics.

We first validated that peripheral blood transcriptomic patterns for the HC and CC clustered together, which was the case. To link transcript abundance to immune pathway enrichment we utilized single-sample gene set enrichment analysis with the GSVA package in R (see Methods)\(^52\). We performed unsupervised clustering on the samples-by-pathway NES scores for all samples in this study and found that individuals from different cohorts have broad qualitative differences in distinct biological pathways (Figure 7A). While these pathways are not by any means completely representative of everything happening biologically in these individual, we do feel that they highlight clear trends across the course of the TB spectrum of disease, compared to a control population of healthy volunteers.

We performed RFR (see Methods) for each pathway against the microbiome space, and summarized the findings in Figure 7B. Surprisingly, we found the abundance of a large number of Firmicutes and particularly Clostridia to be associated with a number of the characterized Hallmark molecular pathways (Supplementary Table S7). Even more intriguingly we found that higher abundance of ASVs that are mapped to the health-associated \textit{F. prausnitzii} predicts a reduction in proinflammatory pathways including INFα, and INFγ. This validates the
findings obtained from the application of our modeling analysis on the clinical trial and EBA longitudinal cohorts (Figure 6). These results, performed in a large cohort of 52 volunteers from the Haitian community, thus reinforce the hypothesis that there exist relationships between specific gastrointestinal microbiota members and peripheral gene expression in humans, and further, that these relationships can be detected and investigated using established laboratory and clinical sequencing techniques.
Figure 7: Analysis of microbiome and blood peripheral gene expression in an independent healthy control human cohort validates association between specific microbiome members and host peripheral gene expression. A. NES scores of 50 Hallmark pathways from the MiSigDB on a per-sample basis for all cohorts in this study. NES score was calculated using the variance stabilized transformed counts from DSEeq, calculated...
with the GSVA package in R, and plotted after scaling across all samples. Columns are split based on arm or group membership and rows are split based on Hallmark pathway categorization. B. Random forest regression results associating specific microbial taxa with Hallmark pathways. Only pathways identified in the RFR model are shown. The ‘Relation’, calculated by taking the first derivative of the ALE plot for each relationship, is positive if the pathway positively associates with a particular ASV, or negative if the pathway negatively associates with a particular ASV.
Discussion

Since the advent of high-throughput microbiome characterization, it has become clear that antibiotics are one of the most common and severe perturbing influences on human microbiome composition, with both acute and longer lasting effects\textsuperscript{53,54}. It also has become evident that the specific microbiome constituents have specific effects on host immunity, including the abundance and function of immune cell subsets\textsuperscript{24}. Significant prior data have documented the effects of antibiotics on microbiome composition and function and the consequent influence of these microbiome factors on immune cell populations\textsuperscript{55}, with the majority of these findings derived using \textit{in vivo} mouse models. While there is no doubt that microbiota dynamics affect host immunity\textsuperscript{6}, it remains unknown to what degree antibiotic induced perturbation of the microbiome may modify the outcome of treatment of infection, or what relationships exist in humans between gut microbiota composition and peripheral gene expression. It is conceivable that antibiotics work to clear infection both due to direct pathogen killing and by immune modulation through the microbiome. It is also possible that the pathogen killing effect of antibiotics may be partially counteracted by detrimental immune effects induced by microbiome perturbation. Such dynamics may be particularly relevant to the treatment of chronic infections such at tuberculosis, in which antibiotic therapy is prolonged and the disease manifestations reflect a mixture of pathogen burden and the balance of inflammatory mediators that cause tissue destruction and chronic symptomatology\textsuperscript{56,57}.

Antibiotic sensitive tuberculosis is treated with six months of antibiotics with predominantly mycobacterial specific agents. In this study we report the early and late microbiome effects of HRZE therapy in subjects with active TB and demonstrate that the same changes observed in a human cross-sectional study of TB treatment\textsuperscript{22} (comparing vs cured and LTBI individuals) were present after just two weeks of treatment. As previously shown\textsuperscript{22}, we conclude that HRZE treatment has a rapid and narrow effect on the intestinal Class of Clostridia, a finding that was also demonstrated in mice\textsuperscript{21,58}. We note that given the mycobacterial-specific nature of TB drugs, and the combinatorial nature in which small molecules interact to affect the microbiome, it was difficult to predict that primarily Clostridia, in the Phylum Firmicutes, would be targeted by HRZE therapy, whereas Actinobacteria, the phylum to which \textit{Mtb} belongs, are relatively unaffected. Experiments in mice have demonstrated that this anti-Clostridia effect is primarily driven by rifampicin\textsuperscript{21}. Clostridia are immunologically active components of the microbiota through their production of metabolites such as short chain fatty acids and other compounds\textsuperscript{2,5,6,59,60}.

The clinical trial allowed us to dissect the relative contributions of pathogen killing and microbiome perturbation to disease resolution because one treatment arm, standard therapy, both reduced \textit{Mtb} bacterial burden and...
perturbed the microbiome, whereas NTZ had no effect on average *Mtb* burden, but did perturb the microbiome in a fashion that overlapped with HRZE. We were able to extend these findings in the EBA dataset with both an investigation into microbiome changes after day 7, day 14, one month, two months, and 6 months post treatment, compared to Baseline, as well as peripheral gene expression and pathway changes after day 14 and two months of HRZE therapy. The findings from these analyses may provide support to the hypothesis that antibiotic perturbation of the microbiome has systemic effects on peripheral gene expression. Further, we speculate that the large heterogeneity in the rapidity of treatment response in TB may be a partial function of heterogeneity in the effects of antibiotics on the microbiome.

To validate the inferred microbiome-host inflammatory relationship, we mined microbiome and blood transcriptomic profiling from an independent human cohort of healthy Haitian individuals. Remarkably, despite the reduced peripheral levels of inflammatory pathways compared to subjects with active TB, we again observe that members of *Clostridium* IV and XIVa predict a reduction in the expression in pro-inflammatory pathways. This validation data strongly supports our conclusion that microbiome composition sets the tone of systemic inflammation, both in disease states and in homeostatic conditions. Further, it is consistent with the prior findings in both humans and animals that Clostridia have been associated with induction of anti-inflammatory states\(^61,62\).

Finally, given the challenge of explaining the relationship between microbiome composition and peripheral gene expression with paired samples, randomized to drug combinations with vastly different effects on both body systems, we strove to use appropriate mathematical approaches for this type of analysis. For single-omic microbiome and RNAseq data, we chose to use limma/voom to model changes in these data given their ability to model many effects while using subject’s baseline values as random effects. For multi-omics integration, we used Random Forest Regression. While there are a variety of statistical and machine learning techniques able to investigate relationships between complex multiparametric datasets (“large p”: microbiome composition, RNAseq data, clinical metadata, randomization cohort, paired-sample baseline normalization, etc., and a “small n” of individuals in early phase clinical trials), Random Forests are adequate for microbiome purposes, as they have been shown to outperform Support Vector Machines in some instances, especially for continuous variable data, and need initialization of a smaller set of parameters compared to other deep-learning methods. We believe that our results highlight the utility of these models to: 1. Provide evidence for or against a particular hypothesis about clinically significant relationships between many potentially related parameters, and 2. To
provide hypothesis generating relationships between the multi-omic constituents (i.e., features) of these models, which can be further tested in mice, validation cohorts, or other model systems.

Our data indicates that within the first 14 days of treatment of tuberculosis, resolution of the active inflammatory response of TB (as measured by peripheral blood transcriptomics) may be strongly affected both by reducing Mtb burden as well as through antimicrobial-induced microbiome perturbations that may act directly on systemic immune function. Among the pathways tightly correlated with both factors are the signature activated pathways of active TB disease: IFN\(\gamma\), type I interferon, and TNF\(\alpha\)\textsuperscript{14}. There is growing evidence that the outcome of active TB reflects a mixture of pathogen burden and cytokine networks that include IL-1 and IFN\(\gamma\), with the latter acting to exacerbate disease\textsuperscript{56}. Our findings indicate that the microbiome perturbation that accompanies TB treatment is a predictor of the normalization of these same pathways during early treatment, suggesting that microbiome perturbation could modify or predict the rapidity of disease resolution.

In the first two weeks of treatment, pathogen killing is the dominant factor, but microbiome dependent modulation of inflammatory responses during treatment may assume an important role during the later phases of treatment when pathogen killing slows. The validation of the relationships between microbiome composition and peripheral gene expression in a healthy control cohort, especially for the collective expression of these same pro- and anti-inflammatory pathways, suggests that these relationships may extend into other populations. Whether these relationships are causal, or biomarkers of another state will remain at the forefront of future study design. Future work will be directed to applying the analytical tools and study design presented here to later time points in the TB treatment course to examine whether microbiome perturbation during treatment associates with clinically relevant treatment outcomes, and whether the abundance of Clostridia correlates with rapidity of Mtb sterilization or the resolution of the inflammatory response that accompanies active TB. Such data might help support trials to test microbiome monitoring as a predictor of TB treatment outcome or help understand interindividual heterogeneity in treatment outcomes.
Materials and Methods

Ethical statement and study approval

All volunteers provided written informed consent to participate in this study. All human studies were reviewed and approved by the IRBs of both Weill Cornell Medicine and Groupe Haitien d’etude du Sarcome de Kaposi et des Infections Opportunistes (GHESKIO) Centers (Port-au-Prince, Haiti). Participants provided informed consent prior to peripheral blood draw for whole blood collection and stool collection for 16S rDNA sequencing. All methods and procedures were performed in accordance with the relevant institutional guidelines and regulations.

Donor recruitment and protection of human subjects

Longitudinal treatment cohort: Donors were enrolled through the Clinical Trials Unit at GHESKIO. Pulmonary TB was diagnosed by clinical symptoms, chest radiograph consistent with pulmonary TB, and positive molecular testing. All participant samples were deidentified on site using a barcode system before they were shipped to Weill Cornell Medicine (WCM)/Memorial Sloan Kettering Cancer Center (MSKCC) for analysis. All clinical metadata was collected on site and managed through the REDCap data management system.

Human healthy control arm: We recruited families of active pulmonary TB patients where at least two siblings within the family were diagnosed with active TB. These criteria were designed to select for households with high risk of transmission of Mtb. Household contacts were then recruited if they had been sleeping in the same house with a TB case for at least one month during the six months prior to the TB case diagnosis. Contacts underwent clinical screening for active TB symptoms and IGRA testing. Healthy donors without history of TB contacts or disease were recruited from the same community as a control group for exposure and also underwent clinical screening for active TB symptoms and IGRA testing. All donors provided informed consent prior to peripheral blood donation for whole blood collection for RNAseq and stool submission for DNA extraction and 16S rDNA sequencing.

Microbial DNA extraction from stool

DNA extraction from stool was performed as described. Stool specimens were collected and stored for less than 24 hours at 4°C, aliquoted (~2 ml each), frozen at −80°C, and shipped to WCM/MSKCC. About 200 – 500 mg of stool from frozen samples was suspended in 500 μl of extraction buffer (200 mM Tris-HCl [Thermo Fisher Scientific], pH 8.0; 200 mM NaCl [Thermo Fisher Scientific]; 20 mM EDTA [MilliporeSigma]), 210 μl of 20% SDS, 500 μl of phenol/chloroform/isoamyl alcohol (25:24:1; MilliporeSigma), and 500 μl of 0.1-mm–diameter
zirconia/silica beads (Biospec Products). Samples were lysed via mechanical disruption with a bead beater (Biospec Products for 2 minutes, followed by 2 extractions with phenol/chloroform/isoamyl alcohol [25:24:1]). DNA was precipitated with ethanol and sodium acetate at −80°C for at least 1 hour, resuspended in 200 μl of nuclease-free water, and further purified with QIAamp DNA Mini Kit (Qiagen) according to the manufacturer’s protocols. DNA was eluted in 200 μl of nuclease-free water and stored at −20°C.

16S rDNA sequencing and bioinformatic analysis
Primers used to amplify rDNA were: 563F (59-nnnnnnnn-NNNNNNNNNNN-AYTGGGYDTAAAGN G-39) and 926R (59-nnnnnnnn-NNNNNNNNNNN-CCGTCAATTYHTTTR AGT-39). Each reaction contained 50 ng of purified DNA, 0.2 mM dNTPs, 1.5 μM MgCl2, 1.25 U Platinum TaqDNA polymerase, 2.5 μl of 10× PCR buffer, and 0.2 μM of each primer. A unique 12-base Golay barcode (Ns) preceded the primers for sample identification after pooling amplicons. One to 8 additional nucleotides were added before the barcode to offset the sequencing of the primers. Cycling conditions were the following: 94°C for 3 minutes, followed by 27 cycles of 94°C for 50 seconds, 51°C for 30 seconds, and 72°C for 1 minute, where the final elongation step was performed at 72°C for 5 minutes. Replicate PCRs were combined and were subsequently purified using the Qiaquick PCR Purification Kit (Qiagen) and Qiagen MinElute PCR Purification Kit. PCR products were quantified and pooled at equimolar amounts before Illumina barcodes and adaptors were ligated on using the Illumina TruSeq Sample Preparation procedure. The completed library was sequenced on an Illumina Miseq platform per the Illumina recommended protocol.

Forward and reverse 16S MiSeq-generated amplicon sequencing reads were dereplicated and sequences were inferred using dada2. Potentially chimeric sequences were removed using consensus-based methods. Taxonomic assignments were made using BLASTN against the NCBI refseq rna database. These files were imported into R and merged with a metadata file into a single Phyloseq object.

Peripheral blood transcriptomics
Collection of peripheral blood and evaluation of host gene expression follows our previous published work. Briefly, peripheral blood was collected into Tempus™ Blood RNA tubes (Applied Biosystems) for the HRZE/NTZ trial cohort, as well as the control cohort. RNA was extracted using the Tempus™ Spin RNA Isolation Kit (Ambion), with addition of on-column DNase treatment (AbsoluteRNA, Ambion). For the EBA longitudinal cohort, PAXgene tubes were used to collect blood according to the manufacturers protocol. RNA was ribo-depleted by polyA selection and libraries were generated using TruSeq (Illumina, San Diego, CA). Paired-end RNA sequences
(50×50PE) were generated with HiSeq 2500 (Illumina) with at least 50 million reads per sample. Sequence integrity was verified using FastQC (Babraham Bioinformatics). Sequences were aligned to the human genome (version hg38) using STAR aligner and transcript counts were estimated using featurecounts. Quality of aligned and counted reads was assessed using Quality of RNA-Seq ToolSet (QoRTs).

**Statistical and Computational Analysis**

**Linear mixed effect models:**

For the clinical trial, to identify the significance of the influence of sex, age, treatment groups (HRZE and NTZ) and time of treatment on time to positivity (TTP), we implemented linear mixed effect model as $TTP \sim Sex + Age + Time * Treatment + 1|ID$. Similarly, to associate the significance of the effect of sex, age, treatment groups (HRZE and NTZ) and sequencing batches and time of treatment on microbiota diversity (Inv Simpson), Inv Simpson measure was modeled as $Diversity \sim Sex + Age + Batch + Time * Treatment + 1|ID$, where:

- $1|ID$ is used as a random effect to account for individual differences
- $Sex$ represents if an individual is male or female
- $Time$ indicates Day 0 and Day 14
- $Treatment$ indicates HRZE or NTZ group
- $Batch$ represents the sequencing batch information

Alpha diversity indices were computed using phyloseq package in R and the implementation of linear mixed effect models were carried out using nlme package in R.

For the EBA longitudinal cohort, to identify the significance of the influence of sex, age, and treatment time on time to positivity (TTP), we implemented linear mixed effect model as $TTP \sim Sex + Age + Time + 1|ID$. Similarly, to identify the significance of the influence of sex, age, and treatment time on microbiota diversity on microbiota diversity (Inv Simpson), we performed linear mixed effect modeling $Diversity \sim Sex + Age + Time + 1|ID$.

**Differential analysis for microbial ASVs and host genes:** Both raw 16Sr rRNA microbiota ASVs and peripheral blood RNAseq gene-expression counts were modeled using the limma/voom pipeline. This allowed us to use linear mixed-effect modeling of gene/ASV counts as of $Count \sim sex + batch + group + 1|ID$. This model statement enables quantifying sex and sequencing batch-dependent effects in addition to establishing effects that are due to treatment group (pre-treatment, HRZE, NTZ). For EBA cohort, we used similar differential analysis approach as clinical trial by modeling gene/ASV counts as $Count \sim Sex + Batch + Time + 1|ID$, where $Time$ represents the Day 0, Day 14 and Day 56 time points. The advantage of limma is that we could use
subject as a random effect to control for baseline differences among individuals, important in this clinical setting. Significance of ASVs affected by the treatment were determined using a Benjamini-Hochberg false discovery rate (FDR) adjusted p-value of 0.05 from the modeling-produced contrast lists (e.g. HRZE vs. Pre, NTZ vs Pre, Day 0 – EBA vs. Day 14 – EBA). To determine how the anti-TB treatment affects both microbiome and peripheral gene expression profiles we performed differential analysis on the counts data obtained by microbiome DNA and peripheral blood RNA sequencing. As the primary endpoint of the clinical trial was powered to determine differences in Mtb load (TTP), we determined the statistical power available to identify significant differences in the abundance of both microbiota ASVs, and in the expression of peripheral genes. We ran power calculations to determine that with 16 pre and 16 post treatment microbiome samples and 8 pre and 8 post treatment RNAseq samples for the HRZE cohort, with 80% power at a significance level (α) of 0.05, we could detect a fold change of 1.4 for microbiome difference and a fold change of 1.8 for mRNA transcripts. In the NTZ cohort, with 18 pre and 18 post treatment microbiome samples and 14 pre and 14 post treatment RNAseq samples, with 80% power at α<0.05, we can detect a fold change of 1.4 for microbiome differences and a fold change of 1.6 for mRNA transcripts. In the EBA cohort with 20 baseline, 10 day 7, 18 day 14, 13 one month, 13 two month, and 11 six month follow up microbiome samples, with 80% power at a significance level (α) of 0.05, we can detect a fold change of 1.4 (day 7), 1.36 (day 14), 1.4 (one, two, and 6 months). In the EBA cohort with 19 baseline, 19 day 14, and 13 two-month RNAseq samples, with 80% power at a significance level (α) of 0.05, we can detect a fold change of 1.4 for mRNA transcripts at day 14, and 1.5 at two months. Power calculations were performed with the RNAseqPower package in R. For microbiome data we calculated a biological coefficient of variation of 0.3, and for RNAseq, we used a coefficient of variation of 0.4. We estimated the expected minimum fold change that we could observe for each group based on the sample size, sequencing depth, and an α of 0.05. To visualize trends of transcript fold changes, we used scatter plots and calculated post-vs-pre fold changes for all transcripts throughout.

Host-Microbiome-Mtb modeling:

We assessed the relative contribution of the intestinal microbiota and Mtb dynamics towards peripheral gene expressions using Random forest regression (RFR). Instead of modeling each gene/transcript profile as a function of microbiome and TTP, we mapped our gene expression data to a set of 50 Hallmark Pathways via GSVA. To avoid having correlated samples from same individual in a model, we instead modeled the changes in normalized enrichment score of hallmark pathways at two time points (Day 0 and Day 14) as a function of change
in microbiome and TTP at corresponding timepoints. For each hallmark pathway $l$ identified to be significantly affected by HRZE (Clinical Trial or Long-term EBA study) or NTZ via linear mixed effects modeling, we first computed the change in normalized enrichment score $\triangle NES_l^i$ between two consecutive time points $Day 0$ and $Day 14$ as $\triangle NES_l^i = NES_l^{14} - NES_l^0$ for each individual. We then regressed this quantity against the corresponding Log2 fold change of normalized counts value (NEV) of every ASV $v$ as $\log_2(\text{NEV}_{14} / \text{NEV}_0)$ in the same interval and against the log fold change in TTP i.e. $\log(TTP_{14} / TTP_0)$. Normalized expression value (NEV) is the CPM (counts per million) obtained by normalizing the raw counts by the library sizes and multiplying by one million.

To train the models we used observations from the clinical trial and from the longitudinal EBA cohort for a total of 34 paired samples. We fit a model for each significant pathway using all the data from the three patients’ group (HRZE clinical trial, NTZ clinical trial and HRZE EBA) because we wanted to find patterns that are general across multiple datasets. Each model was trained using 5000 generated trees and a train-validation partitioning of 80-20% of the data.

Permutated importance\textsuperscript{43} measure is used to assess the importance and significance of predictors (e.g. TTP, microbiome constituents, etc.) towards the dependent variable (e.g. pathways). Higher the permutation importance of a predictor, higher is the association towards the outcome variable. Accumulated Local Effects (ALE) plots\textsuperscript{44} are used to estimate the relationship between the predictors towards the outcome variable. To simplify the effect of predictors on pathways into a monotonic relationship, we compute the slope of a fitted straight line of ALE plots and summarize the direction of the slope into a positive/negative effect of predictors towards outcome variable.

Random Forest Regression Analysis of control cohort: To investigate microbiome-pathway relationships in FC and CC cohort, we modeled the normalized enrichment score (NES) of each hallmark pathway as a function of normalized expression value (NEV) of ASVs for corresponding samples. Each RFR model was trained using 5000 generated trees and a train-validation partitioning of 80-20% of the data. Permutated importance\textsuperscript{43} and (ALE) plots\textsuperscript{44} were used to assess importance and relationship of ASVs towards pathways.

Within sample GSEA analysis: The ssGSEA (single sample gene set enrichment analysis) method\textsuperscript{69} was used to profile within-sample differences between pathways from the MiSigDB Hallmark pathways list\textsuperscript{35}, or other MiSigDB lists (e.g., KEGG), with the GSVA package in R\textsuperscript{52}. The MiSigDB Hallmark pathway list is a well validated set of general curated biological pathways that give insight into specific biological and cellular processes. Additionally, we obtained a list of well validated active TB signatures from the TBSignatureProfilier R package\textsuperscript{70}.
Variance stabilized transformed (vst) counts derived from DESeq2 were used as input into the GSVA function in the GSVA R package with default parameters (kcdf="Gaussian") and scaled Normalized Enrichment Scores (NES) were plotted as heatmaps. Importantly, unlike classical GSEA, this analysis is agnostic to sample phenotype.

Identification of differential pathways post antibiotic treatment in both the clinical trial and EBA cohort was performed using linear mixed effect model where we modeled the normalized enrichment score (NES) of each pathways as \( NES \sim \text{sex} + \text{batch} + \text{group} + 1|ID \) and \( NES \sim \text{sex} + \text{batch} + \text{Time} + 1|ID \), respectively. Significance of pathways affected by the treatment were determined using a Benjamini-Hochberg false discovery rate (FDR) adjusted p-value of <0.05. Significant pathways in both trials were used as outcome/dependent variables for the RFR models.

**Data availability statement**

Data on Time to Positivity where obtained from Walsh et al. 2020 and are available on Github at [https://wipperman.github.io/TBRU/](https://wipperman.github.io/TBRU/). 16S rDNA sequencing data is deposited with the SRA under accession no. PRJNA445968. Peripheral blood transcriptomic data are deposited with the SRA under accession no. PRJNA445968. Code to analyze the data presented in this manuscript and to reproduce all of the figures and results is available on Github at [https://wipperman.github.io/TBRU/](https://wipperman.github.io/TBRU/).

**Author Contributions**

Patient recruitment, enrollment, and sample collection were contributed by LM, KM, KFW, JB, and SCV; Laboratory experiments were performed by MFW and CKV; Data analysis was performed by MFW, SB, VB; Wrote manuscript: MFW, SB, MSG, VB; Edited manuscript: all authors.

MFW and SB are co-first authors, and MSG and VB are co-last authors. Co-authorship and author order were determined by recognition that the integration of the nuances clinical trial data and mathematical modeling are different skillsets found in different laboratory environments. Each were important components to the validity and message of this manuscript.

**Conflicts of Interest**

MFW is currently an employee and shareholder of Regeneron Pharmaceuticals, Inc. MSG reports consulting fees and equity in Vedanta Biosciences, Inc., and consulting fees from Takeda Pharmaceutical Co., Ltd. VB is supported by a Sponsored Research Agreement from Vedanta Biosciences, Inc.
Funding
MFW, CKV, YT, KW, CN, DWF, MSG acknowledge funding from the Tri-I TBRU (grant: U19AI111143). CV acknowledges support from K08AI132739. MFW acknowledges support from the National Center for Advancing Translational Sciences (grant: TL1TR002386-02). Support for NCT02684240 came primarily from the Abby and Howard P. Milstein Program in Chemical Biology and Translational Medicine. VB acknowledges support from the National Science Foundation (grant: 1458347). This work was supported by P30 CA008748.
References

1. Martin, C.R., Osadchiy, V., Kalani, A. & Mayer, E.A. The Brain-Gut-Microbiome Axis. Cell Mol Gastroenterol Hepatol 6, 133-148 (2018).

2. Atarashi, K., et al. Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. Nature 500, 232-236 (2013).

3. Atarashi, K., et al. Induction of colonic regulatory T cells by indigenous Clostridium species. Science 331, 337-341 (2011).

4. Tanoue, T., et al. A defined commensal consortium elicits CD8 T cells and anti-cancer immunity. Nature 565, 600-605 (2019).

5. Vorkas, C.K., et al. Mucosal-associated invariant and gammacdelta T cell subsets respond to initial Mycobacterium tuberculosis infection. JCI Insight 3(2018).

6. Campbell, C., et al. Extrathymically Generated Regulatory T Cells Establish a Niche for Intestinal Border-Dwelling Bacteria and Affect Physiologic Metabolite Balance. Immunity 48, 1245-1257 e1249 (2018).

7. Levan, S.R., et al. Elevated faecal 12,13-diHOME concentration in neonates at high risk for asthma is produced by gut bacteria and impedes immune tolerance. Nat Microbiol (2019).

8. Belkaid, Y. & Hand, T.W. Role of the microbiota in immunity and inflammation. Cell 157, 121-141 (2014).

9. Ryan, F.J., et al. Changes in the Composition of the Gut Microbiota and the Blood Transcriptome in Preterm Infants at Less than 29 Weeks Gestation Diagnosed with Bronchopulmonary Dysplasia. mSystems 4, e00484-00419 (2019).

10. Grigg, J.B. & Sonnenberg, G.F. Host-Microbiota Interactions Shape Local and Systemic Inflammatory Diseases. J Immunol 198, 564-571 (2017).

11. Wang, Z., Arat, S., Magid-Slav, M. & Brown, J.R. Meta-analysis of human gene expression in response to Mycobacterium tuberculosis infection reveals potential therapeutic targets. BMC Syst Biol 12, 3 (2018).

12. Kaforou, M., et al. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. PLoS Med 10, e1001538 (2013).

13. Zak, D.E., et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. Lancet 387, 2312-2322 (2016).

14. Berry, M.P., et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature 466, 973-977 (2010).

15. Lesho, E., et al. Transcriptional responses of host peripheral blood cells to tuberculosis infection. Tuberculosis (Edinb) 91, 390-399 (2011).

16. Ottenhoff, T.H., et al. Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis. PLoS One 7, e45839 (2012).

17. Bloom, C.I., et al. Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. PLoS One 8, e70630 (2013).

18. Cliff, J.M., et al. Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. J Infect Dis 207, 18-29 (2013).

19. Singhania, A., et al. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. Nat Commun 9, 2308 (2018).

20. Aljajyoussi, G., et al. Pharmacokinetic-Pharmacodynamic modelling of intracellular Mycobacterium tuberculosis growth and kill rates is predictive of clinical treatment duration. Sci Rep 7, 502-502 (2017).

21. Namasiyayam, S., et al. Longitudinal profiling reveals a persistent intestinal dysbiosis triggered by conventional anti-tuberculosis therapy. Microbiome 5, 71 (2017).

22. Wipperman, M.F., et al. Antibiotic treatment for Tuberculosis induces a profound dysbiosis of the microbiome that persists long after therapy is completed. Sci Rep 7, 10767 (2017).
23. Faith, J.J., Ahern, P.P., Ridaura, V.K., Cheng, J. & Gordon, J.I. Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. Scien Transl Med, 220ra211 (2014).

24. Geva-Zatorsky, N., et al. Mining the Human Gut Microbiota for Immunomodulatory Organisms. Cell 168, 928-943 e911 (2017).

25. Shigyo, K., et al. Efficacy of nitazoxanide against clinical isolates of Mycobacterium tuberculosis. Antimicrob Agents Chemother 57, 2834-2837 (2013).

26. K. F. Walsh, K.M., M. H. Lee, S. C. Vilbrun, L Mathurin, D Jean Francois, M Zimmerman, F Kaya, N Zhang, K Saito, O Ocheretina, R Savic, V Dartois, W. D. Johnson, J. W. Pape, C Nathan, D. W. Fitzgerald. Early bactericidal activity trial of nitazoxanide for pulmonary tuberculosis. Antimicrob Agents and Chemotherapy (2020).

27. Harausz, E.P., et al. Activity of nitazoxanide and tizoxanide against Mycobacterium tuberculosis in vitro and in whole blood culture. Tuberculosis (Edinb) 98, 92-96 (2016).

28. Wagner, B.D., et al. On the Use of Diversity Measures in Longitudinal Sequencing Studies of Microbial Communities. Frontiers in Microbiology 9(2018).

29. Ritchie, M.E., et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47-e47 (2015).

30. Parada Venegas, D., et al. Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. Frontiers in Immunology 10(2019).

31. Heikken, A., et al. Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. Microbiome 7, 75 (2019).

32. Kim, S., Covington, A. & Pamer, E.G. The intestinal microbiota: Antibiotics, colonization resistance, and enteric pathogens. Immunol Rev 279, 90-105 (2017).

33. Subramanian, A., et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545-15550 (2005).

34. Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. bioRxiv, 060012 (2019).

35. Liberzon, A., et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 1, 417-425 (2015).

36. Glickman, M.S., Cox, J.S. & Jacobs, W.R., Jr. A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of Mycobacterium tuberculosis. Mol Cell 5, 717-727 (2000).

37. Krzywinska, E. & Stockmann, C. Hypoxia, Metabolism and Immune Cell Function. Biomedicines 6(2018).

38. Palmer, N.P., et al. Concordance between gene expression in peripheral whole blood and colonic tissue in children with inflammatory bowel disease. PLoS One 14, e0222952-e0222952 (2019).

39. Knox, N.C., Forbes, J.D., Van Domselaar, G. & Bernstein, C.N. The Gut Microbiome as a Target for IBD Treatment: Are We There Yet? Curr Treat Options Gastroenterol 17, 115-126 (2019).

40. Magurran, A.E. Measuring biological diversity, (John Wiley & Sons, 2013).

41. Haran, J.P., et al. Alzheimer’s Disease Microbiome Is Associated with Dysregulation of the Anti-Inflammatory P-Glycoprotein Pathway. MBio 10(2019).

42. Johnstone, I.M. & Titterington, D.M. Statistical challenges of high-dimensional data. Philos Trans A Math Phys Eng Sci 367, 4237-4253 (2009).

43. Altmann, A., Tolosi, L., Sander, O. & Lengauer, T. Permutation importance: a corrected feature importance measure. Bioinformatics 26, 1340-1347 (2010).

44. Apley, D.W. & Zhu, J. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. in arXiv e-prints (2016).

45. Collins, M.D., et al. The phylogeny of the genus Clostridium: proposal of five new genera and eleven new species combinations. Int J Syst Bacteriol 44, 812-826 (1994).

46. Lopetuso, L.R., Scaldaferr, F., Petito, V. & Gasbarrini, A. Commensal Clostridia: leading players in the maintenance of gut homeostasis. Gut Pathog 5, 23-23 (2013).
47. Garrett, W.S., et al. Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host Microbe* **8**, 292-300 (2010).

48. Seishima, J., et al. Gut-derived Enterococcus faecium from ulcerative colitis patients promotes colitis in a genetically susceptible mouse host. *Genome Biology* **20**, 252 (2019).

49. Peled, J.U., et al. Microbiota as Predictor of Mortality in Allogeneic Hematopoietic-Cell Transplantation. *New England Journal of Medicine* **382**, 822-834 (2020).

50. Freedberg, D.E., et al. Pathogen colonization of the gastrointestinal microbiome at intensive care unit admission and risk for subsequent death or infection. *Intensive Care Medicine* **44**, 1203-1211 (2018).

51. Zeng, M.Y., Inohara, N. & Nuñez, G. Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunology* **10**, 18-26 (2017).

52. Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).

53. Dethlefsen, L. & Relman, D.A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc Natl Acad Sci U S A* **108** Suppl 1, 4554-4561 (2011).

54. Lavelle, A., et al. Baseline microbiota composition modulates antibiotic-mediated effects on the gut microbiota and host. *Microbiome* **7**, 111 (2019).

55. Morgan, A., et al. Uncovering effects of antibiotics on the host and microbiota using transkingdom gene networks. *Gut* **64**, 1732-1743 (2015).

56. Mayer-Barber, K.D., et al. Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk. *Nature* **511**, 99-103 (2014).

57. Pirofski, L.-A. & Casadavall, A. Antimicrobial therapy in the context of the Damage-response framework: The prospect of optimizing therapy by reducing host damage. *Antimicrobial agents and chemotherapy*, AAC.01800-01819 (2019).

58. Namasivayam, S., Sher, A., Glickman, M.S. & Wipperman, M.F. The Microbiome and Tuberculosis: Early Evidence for Cross Talk. *MBio* **9**(2018).

59. Arpaia, N., et al. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504**, 451-455 (2013).

60. Gold, M.C., et al. Human mucosal associated invariant T cells detect bacterially infected cells. *PLoS Biol* **8**, e1000407 (2010).

61. Skelly, A.N., Sato, Y., Kearney, S. & Honda, K. Mining the microbiota for microbial and metabolite-based immunotherapies. *Nature Reviews Immunology* **19**, 305-323 (2019).

62. Schirmer, M., Garner, A., Vlamakis, H. & Xavier, R.J. Microbial genes and pathways in inflammatory bowel disease. *Nature Reviews Microbiology* **17**, 497-511 (2019).

63. Harris, P.A., et al. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* **42**, 377-381 (2009).

64. Callahan, B.J., et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-583 (2016).

65. Dupnik, K.M., et al. Blood transcriptomic markers of Mycobacterium tuberculosis load in sputum. *Int J Tuberc Lung Dis* **22**, 950-958 (2018).

66. Dobin, A., et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2012).

67. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)* **30**, 923-930 (2014).

68. Hartley, S.W. & Mullikin, J.C. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC bioinformatics* **16**, 224-224 (2015).

69. Barbie, D.A., et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108-112 (2009).
Jenkins D, Z.Y., Johnson W, Odom A, Love C. TBSignatureProfiler: Profile RA-Seq Data Using TB Pathway Signatures. R package version 1.0.0. in https://github.com/compbiomed/TBSignatureProfiler. (2020).