

RESEARCH

Open Access



Profiling the airway in the macaque model of tuberculosis reveals variable microbial dysbiosis and alteration of community structure

Anthony M Cadena^{1,2†}, Yixuan Ma^{3†}, Tao Ding^{3†}, MacKenzie Bryant^{1,4}, Pauline Maiello¹, Adam Geber³, Philana Ling Lin⁵, JoAnne L Flynn^{1*} and Elodie Ghedin^{3,6*} 

Abstract

Background: The specific interactions of *Mycobacterium tuberculosis* (Mtb), the causative agent of tuberculosis (TB), and the lung microbiota in infection are entirely unexplored. Studies in cancer and other infectious diseases suggest that there are important exchanges occurring between host and microbiota that influence the immunological landscape. This can result in alterations in immune regulation and inflammation both locally and systemically. To assess whether Mtb infection modifies the lung microbiome, and identify changes in microbial abundance and diversity as a function of pulmonary inflammation, we compared infected and uninfected lung lobe washes collected serially from 26 macaques by bronchoalveolar lavage over the course of infection.

Results: We found that Mtb induced an initial increase in lung microbial diversity at 1 month post infection that normalized by 5 months of infection across all macaques. Several core genera showed global shifts from baseline and throughout infection. Moreover, we identified several specific taxa normally associated with the oral microbiome that increased in relative abundance in the lung following Mtb infection, including *SR1*, *Aggregatibacter*, *Leptotrichia*, *Prevotella*, and *Campylobacter*. On an individual macaque level, we found significant heterogeneity in both the magnitude and duration of change within the lung microbial community that was unrelated to lung inflammation and lobe involvement as seen by positron emission tomography/computed tomography (PET/CT) imaging. By comparing microbial interaction networks pre- and post-infection using the predictive algorithm SPIEC-EASI, we observe that extra connections are gained by *Actinomycetales*, the order containing Mtb, in spite of an overall reduction in the number of interactions of the whole community post-infection, implicating Mtb-driven ecological reorganization within the lung.

(Continued on next page)

* Correspondence: joanne@pitt.edu; elodie.ghedin@nyu.edu

[†]Anthony M Cadena, Yixuan Ma and Tao Ding contributed equally to this work.

¹Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

³Center for Genomics and Systems Biology, Department of Biology, New York University, New York, USA

Full list of author information is available at the end of the article



(Continued from previous page)

Conclusions: This study is the first to probe the dynamic interplay between Mtb and host microbiota longitudinally and in the macaque lung. Our findings suggest that Mtb can alter the microbial landscape of infected lung lobes and that these interactions induce dysbiosis that can disrupt oral-airway boundaries, shift overall lung diversity, and modulate specific microbial relationships. We also provide evidence that this effect is heterogeneous across different macaques. Overall, however, the changes to the airway microbiota after Mtb infection were surprisingly modest, despite a range of Mtb-induced pulmonary inflammation in this cohort of macaques.

Keywords: Mycobacterium tuberculosis, Lung and airway microbiota, Microbiome, 16S rRNA gene, Cynomolgus macaque

Background

Microbial dysbiosis is increasingly recognized as a significant factor influencing human disease and host immunity in a variety of contexts spanning allergy, autoimmunity, cancer, and infectious disease [1–6]. In the context of infectious disease, studies have shown that the microbiome has important influences on the immunological landscape by altering immune regulation, inflammation, and pathogen colonization [2, 7–9]. The gut has been the primary focus of a number of studies linking the immune system and an organ's microbiome [3, 10, 11], but it is increasingly clear that other human immune surfaces and interfaces, including the lung, are actively shaped by host-microbe interactions [12–15]. How specific pathogens modify and influence host microbial landscapes and site-specific immunity, particularly in the lung, is poorly understood. In particular, very little is known about the interaction of the lung microbiome and *Mycobacterium tuberculosis* (Mtb), the bacillus that causes tuberculosis (TB). Human Mtb infection presents along a spectrum of host outcomes with only ~10% presenting with active TB, a clinically defined state of infection that includes chronic cough, hemoptysis, weight loss, night sweats, and fever, and may manifest for years until clinically diagnosed [16–18]. By contrast, the majority of Mtb infections (90%) are successfully contained in an asymptomatic state termed latent TB infection (LTBI) that can persist for their lifetime [16, 18, 19]. Current estimates place 2 billion humans within this clinically latent reservoir [19, 20] who are at risk of reactivation and progression to active TB. The precise mechanisms driving the variability in human Mtb infection outcome are poorly understood but may include the earliest interactions between pathogen and the local lung microbiota [21].

Previous work in human TB and the lung microbiome suggested that TB patients have altered diversity in their airway microbiota compared to uninfected humans [22–24] but these studies have been limited by their cross-sectional nature and some are confined to the sputum. The distribution of some genera (*Stenotrophomonas*, *Phenylobacterium*, and *Cupriavidus*) was unique to

TB patients [23, 24]. Certain patterns of microbiota (e.g., presence of *Pseudomonas spp.*) were observed in patients with recurrent TB and treatment failure, suggesting that the lung microbiota interacts with both host and pathogen [25]. Specifically, alterations in *Treponema* and *Atopobium* were associated with recurrent TB suggesting that an altered “normal” microbiome correlated with ongoing susceptibility to TB [25]. Studies examining the microbiota of the oropharynx between healthy controls and TB patients showed differences in diversity and abundance of particular organisms although no differences in major phyla were observed [26]. Other studies in experimental murine models of infection have noted distinct changes in the gut microbiome following Mtb challenge [27], although the precise effects and relationship of these changes on course of infection were not explored. A more recent study profiled alterations in the gut microbiota after tuberculosis antibiotic treatment in mice and observed a distinct change in bacterial taxa that persisted months after cessation of treatment [28]. Collectively, this body of work hints at a related immune–lung microbiome axis that extends from the lung to the oropharynx, as well as the gut, that exhibits significant cross-talk following Mtb infection.

Here, using a macaque model of TB, we explored for the first time whether Mtb infection alters the lung microbiota in a significant and durable manner, and whether the microbial shift is associated with pulmonary inflammation. By serially sampling the airway of 26 infected macaques via bronchoalveolar lavage (BAL), we have provided the first survey of the microbial lung landscape over time in the context of this chronic lung pathogen.

Results

Lung microbial richness and composition shift during Mtb infection

To determine whether infection with Mtb had an effect on the microbial composition of the macaque lung, we sampled 26 macaques at baseline (2 weeks pre-infection) and at 3 time points post-infection (months 1, 4, and 5 post-infection) (Table 1). For each time point, a

Table 1 Number of samples for each sample type, including controls

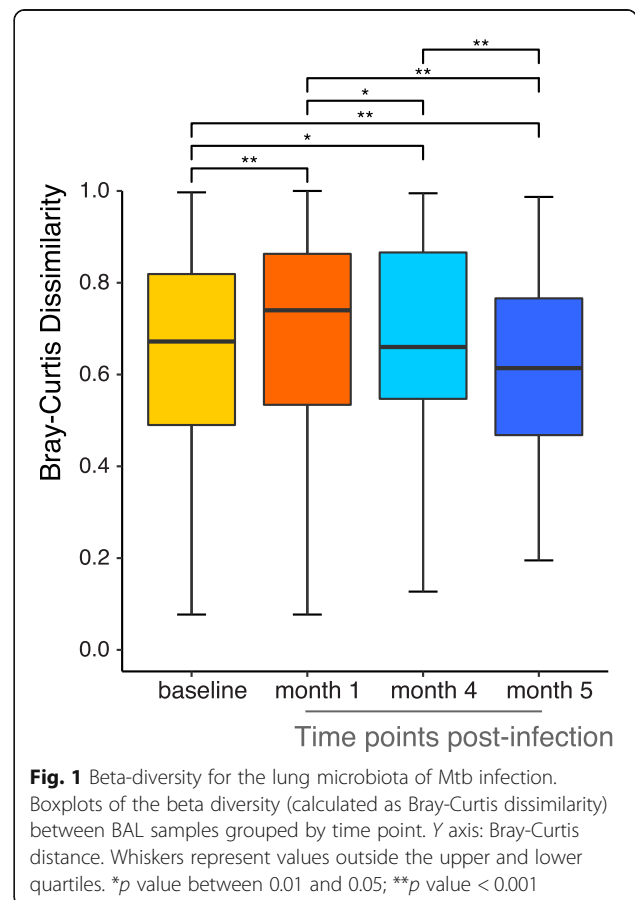
Cohort	Macaques	Oral washes	Bronchoscope controls	BAL of lobes	Reagent control	Total number
1	10	36	36	70	20	162
2	16	55	55	108	32	250
Total	26	91	91	178	52	412

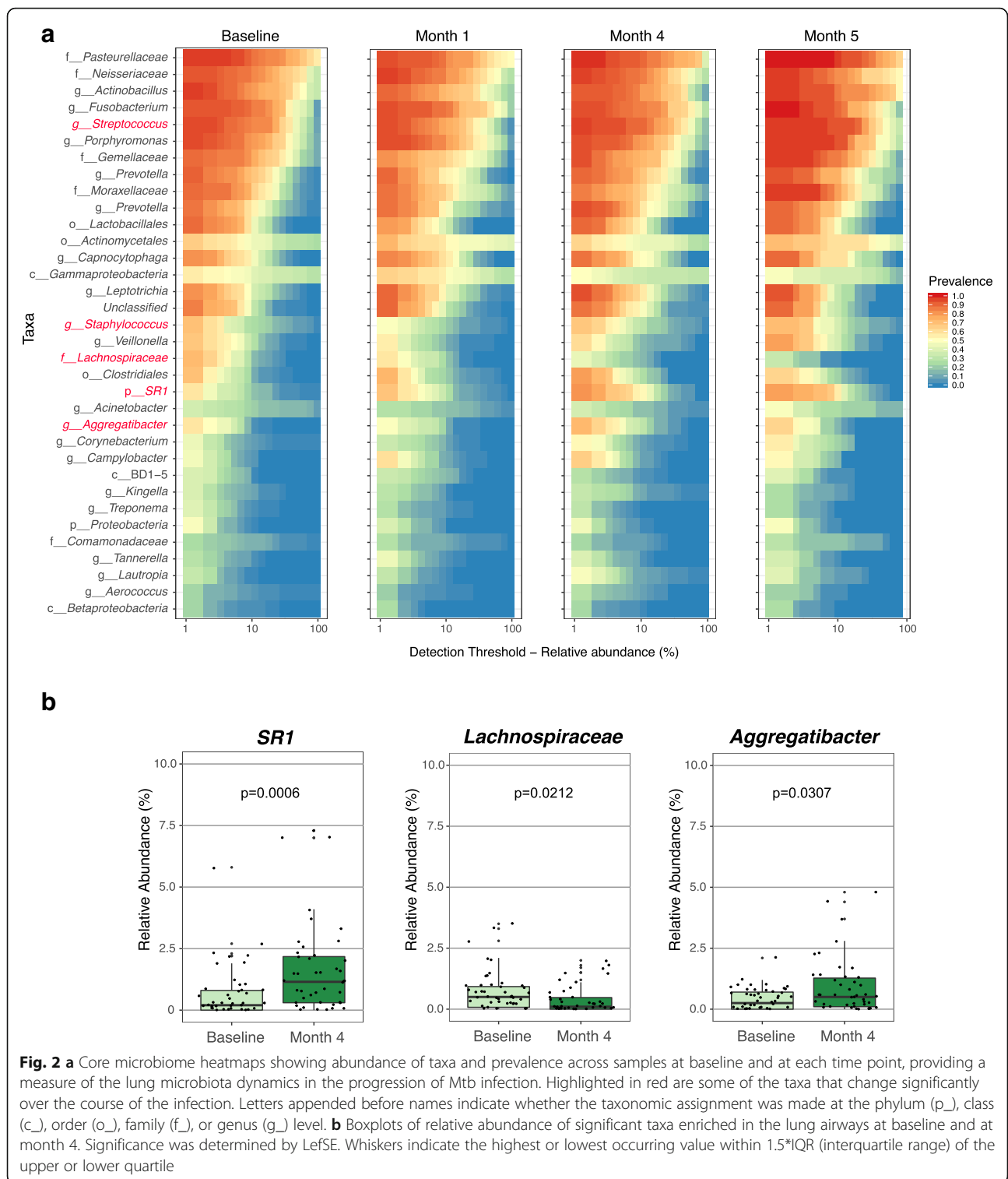
bronchoalveolar lavage (BAL) sample was collected from the right and left lobes, separately. As a negative control, a saline wash sample of the bronchoscope was also obtained. To confirm that the BAL sampling would specifically access the lungs and not be biased by the microbial community of the oral cavity, we also collected oral wash from each animal. We determined by statistical testing that the microbial communities of these two compartments were significantly different ($p < 0.001$), as visualized by t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis (Additional file 1: Figure S1).

We compared the microbial communities between samples (Beta-diversity) at baseline and post-infection using Bray-Curtis distance as a measure of similarity (the more similar the communities, the shorter the distance). Visualized by t-SNE, there appears to be no clustering of the samples based on their infection status (baseline versus post-infection) or by specific time point (Additional file 2: Figure S2). However, when we performed a Wilcoxon test, we observed a small but significant change post-infection compared to the baseline, with an initial overall increase in Beta-diversity at 1 month followed by a decline in Beta-diversity at months 4 and 5 (Fig. 1), indicating that microbial communities of the samples are becoming more similar at these later time points.

When performing analyses of the microbial communities at the level of single macaques, we see unique compositions of bacterial taxa that shifted over the course of the infection (Additional file 3: Figure S3). Although the macaques were given the same food, they were single-housed, thus the variability observed in the respiratory microbiota is not surprising given the genetic variability of macaques and the different exposures of each monkey over time, including antibiotic treatment (prior to Mtb infection) as well as the host variability in TB [18]. To identify global shifts in microbial composition across all macaques, we focused on the core microbiota, which is comprised of taxa with a minimum relative abundance of 0.1% that are shared by 95% of the subjects. We thus identified four core microbiota, one for each time point, and we visualized the common taxa across these time points (Fig. 2a). We compared baseline samples with post-infection samples at each time point by Linear discriminant analysis Effect Size (LEfSe). We identified three taxa that shifted significantly between baseline and month 4, a time point at which host outcome is well established [21, 29]. *Lachnospiraceae* (p value = 0.0212) was

present at a higher relative abundance in the pre-infection compared with post-infection samples at month 4, whereas *SRI* (p value = 0.0006) and *Aggregatibacter* (p value = 0.0307) were enriched post-infection at month 4 (Fig. 2b). The family *Lachnospiraceae* and the unculturable Candidate division SR1 bacteria have previously been found in the mammalian gut [30] and mouth [31, 32], respectively, and an oral strain of *Aggregatibacter* was sequenced from the Rhesus macaque [33], indicating that these are likely commensal bacteria that have shifted in relative abundance with Mtb infection. Comparisons of core microbiota across other time points also led to the identification of taxa such as *Staphylococcus* and *Streptococcus* from month 1 to month 4 ($p = 0.049$ and $p = 0.040$, respectively).





Community abundance and structure are not associated with pulmonary inflammation or lobe involvement in TB
 To investigate whether the dynamics of the microbiota are associated with progression of disease and its accompanying inflammation in the macaques, we performed a

multivariate statistical analysis that determines associations between clinical metadata and microbial community abundance (MaAsLin). We did not identify any specific taxa that were specifically correlated with pulmonary inflammation throughout the course of infection

in these animals (data not shown). We further explored correlations with inflammation and disease involvement by looking at the community structure as a whole. To do that, we partitioned the data into community types using Dirichlet multinomial mixture models [34]. We identified four major community types in the lung microbiota. Community types B and C had the lowest alpha-diversities (Fig. 3a) and were dominated by one taxon each: *Gammaproteobacteria* and *Actinomycetales*, respectively (Fig. 3b). Community types A and D were each represented by a mix of taxa at different relative abundances. *Capnocytophaga* was over-represented in community type D but at a relative abundance between 2 and 3% (not shown). When labeling the samples by their community types and visualizing their beta-diversity, we see that communities B and C stand out as the tightest clusters (Fig. 3c). To determine the stability of the community types over the course of the infection, we mapped the communities at each time point across the lung samples and in each monkey (Fig. 3d). Monkeys were rank ordered by their inflammation status (least to greatest), as measured by their total lung [¹⁸F]fluoro-D-glucose (FDG) uptake by positron emission tomography and computed tomography (PET/CT) scans at 4–5 months post infection (Table 2). We previously published that total lung FDG activity correlates with thoracic bacterial burden in macaques [35]. Lung side involvement was determined by visual confirmation of disease pathology by PET/CT scan at the respective time point post-infection. We observed several community shifts during infection in a number of monkeys, but the microbial community appears stable in others. Community types B and C were over-represented in samples that come largely from two monkeys (IDs 20615 and 9815). Community type D was the most prominent community found throughout infection, and between left and right lung lobes. There was no discernible relationship between the community types or shift in community type over time and inflammation status or lung side involvement (Fig. 3d).

Infection with Mtb leads to modified microbial interactions

To determine how specific microbes potentially interact with each other in the community and how their interactions are disrupted by Mtb infection, we compared co-occurrence of taxa over the course of the Mtb infection using the SPIEC-EASI algorithm (Fig. 4), which infers a graphical model using conditional independence between OTUs [36, 37]. In general, we see that the total number of microbial interactions, as indicated by the number of edges between the nodes, decreased post-infection (pre: $n = 122$, post: $n = 104$) (Fig. 4 and Table 3). Of the taxa profiled, *Porphyromonas*,

Capnocytophaga, and *Moraxellaceae* had the greatest loss in predicted interactions with the latter losing all of its interactions indicating it has become conditionally independent of other taxa present. By contrast, *Actinomycetales*—the order that includes mycobacteria—which only had 3 connections (all positive) in pre-infection samples, had 9 connections in post-infection samples, including 2 positive and 7 negative correlations. This implies that Mtb affects the overall network and competes with specific taxa. We also observed that the diverse class *Gammaproteobacteria*, which contains a number of important human pathogens including *Pseudomonas*, *Yersinia*, *Salmonella*, *Vibrio*, and *Escherichia coli* [38] established six new taxonomic interactions in the airway and became a central node following Mtb infection.

Discussion

In this study, we coupled our cynomolgus macaque model of TB [35, 39] with the ability to serially sample the airways of animals throughout the course of lung infection for microbiome analysis. We exploited the [¹⁸F]-FDG PET/CT imaging analyses that we have developed for use in this model to assess infection trajectory and severity combined with microbial profiling [29, 40–45] to determine whether Mtb infection changed the lung microbial community and the duration of that effect. As a result, our study is the first targeted exploration of the lung microbiome in an animal model that faithfully replicates human Mtb infection outcomes and pathology. Leveraging these imaging and sequencing technologies with our TB animal model, we found that Mtb induces an initial shift in the overall microbial richness that peaks at 1 month post infection and returns to baseline by 5 months across our infected macaques (Fig. 1); there was no discernible clustering that occurred between pre-infection and any time points post-infection of our monkeys in aggregate (Additional file 2: Figure S2). Importantly, we observed a high degree of variability in airway microbial change across the different macaques: several monkeys displayed robust microbial shifts while other monkeys harbored more stable lung communities (Additional file 3: Figure S3) likely reflecting the outbred nature of the animals, variable pre-Mtb exposures and treatments, and the heterogeneity in Mtb infections in macaques [18]. We postulate that this observed microbial heterogeneity may in turn contribute to the spectrum of host outcome seen in TB by modifying initial airway interactions between host cells and Mtb or specific granuloma microenvironments.

When we profiled core taxa for specific changes in relative abundance and prevalence following Mtb infection, we found several shifts of residents associated with the mammalian oral microbiota into the airway. These included members of the *Aggregatibacter*, *Streptococcus*,

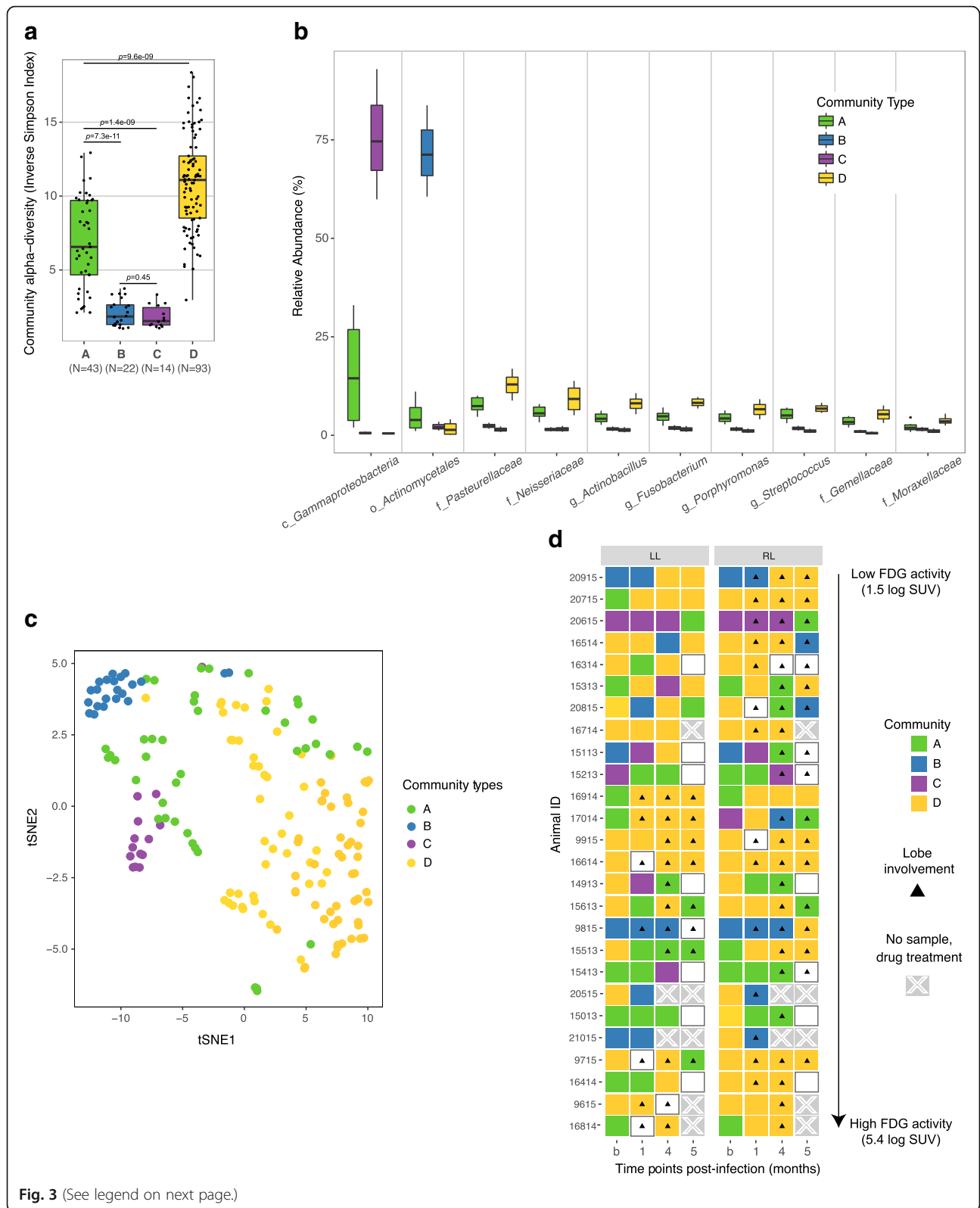


Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Community types found in the lung microbiota. **a** Alpha diversity. p values from Wilcoxon rank sum tests are included. The alpha diversities of community types B and C were overall significantly lower (p value = $2.2e-16$) than those of community types A and D. **b** Relative abundance of taxa over-represented in each community type. Letters appended before names indicate whether the taxonomic assignment was made at the class (c_), order (o_), family (f_), or genus (g_) level. **c** tSNE, colored by community types. **d** Community type that predominates in each monkey at each time point. Triangles represent radiographic evidence of lung pathology (i.e., a granuloma or other grossly visible disease at a resolution of 1 mm by CT) that coincide with microbiome sampling at the indicated time point. From top to bottom (arrow), monkeys were ordered, low to high, by the log transformed total [^{18}F]-FDG activity at 4–5 months. Xs indicate samples not collected; white (blank) boxes indicate that the sequence depth was not sufficient for analysis (fewer than 1000 sequence reads) and so the sample was not included

and *Staphylococcus* (Fig. 2) genera. Their enrichment over the course of Mtb infection hints at increased transit of oral microbes into the lower airway. While LEfSe analysis identified both *SRI* and *Aggregatibacter* as significantly enriched at 4 months post challenge, *Lachnospiraceae*, a family associated with the intestinal microbiome [46] and recently linked to adiposity following broad spectrum antibiotic challenge [30], was significantly decreased.

Table 2 [^{18}F]-FDG uptake values for each monkey at 4–5 months post infection by PET/CT

Macaques	PET HOT (4–5 months (SUV))	Log ₁₀ (PET HOT 4–5 months)
20,915	36.27	1.56
20,715	655.30	2.82
20,615	903.38	2.96
16,514	911.11	2.96
16,314	1102.36	3.04
15,314	1320.82	3.12
20,815	1426.94	3.15
16,714	3498.34	3.54
15,113	3701.98	3.57
15,213	5264.63	3.72
16,914	5724.37	3.76
17,014	6492.61	3.81
9915	7166.13	3.86
16,614	7832.35	3.89
14,913	8517.50	3.93
15,613	10,563.85	4.02
9815	14,896.56	4.17
15,513	18,738.09	4.27
15,413	27,612.06	4.44
20,515	30,032.82	4.48
15,013	33,494.44	4.52
21,015	39,013.51	4.59
9715	53,815.37	4.73
16,414	121,015.76	5.08
16,814	277,171.63	5.44

To further address functional microbe relationships within the airways, we profiled microbe interaction networks at both pre- and post-infection using SPIEC-EASI [37]. We saw that the majority of taxa have a net loss of their total predicted interactions (both positive and negative) post Mtb challenge; by contrast, *Actinomycetales*, the order that contains Mtb, had an overall increase in negative correlations. Other notable network shifts included the transition of *Gammaproteobacteria* from a peripheral node to a central node with the predicted increase of six new edges following Mtb challenge. Collectively, these microbial correlation networks imply that Mtb may alter the functional interactions of microbes residing within the airways during infection, possibly to better establish a niche for infection. It posits unique microbe-microbe relationships secondary to infection that may initiate local dysbiosis and potentially influence barrier integrity, lung immunity, and host inflammation [13, 14, 47].

While relating specific host disease status and TB outcome with alteration of lung microbiota was beyond the scope of this pilot study (profiled monkeys were pooled from multiple studies with differing infection lengths and experimental endpoints), we observed remarkable stability in the overall lung microbiota post Mtb challenge. This relative stability was evident in both involved and uninvolved lung lobes (lobes with and without visible disease) and in lung lobes with both high and low levels of pulmonary inflammation (Fig. 3). Although this stability was not entirely universal, as was highlighted above with some macaques exhibiting greater shifts in both relative abundance and community structure, it is surprising that a robust lung bacterial infection did not cause a more significant dysbiosis of the airways. Interestingly, Morris and colleagues [48] also reported an absence of systemic or persistent alteration of the macaque lung microbiota following long-term SHIV infection despite notable host-specific variation and concomitant immunosuppression. Similar non-discriminatory results were seen in a recent study of the oral and lung microbiota of HIV-infected individuals (treatment naïve median CD4 count = 668 cells/mm³ and HAART median CD4 count = 618 cells/mm³) compared against HIV-negative controls [49]. A separate lung study of

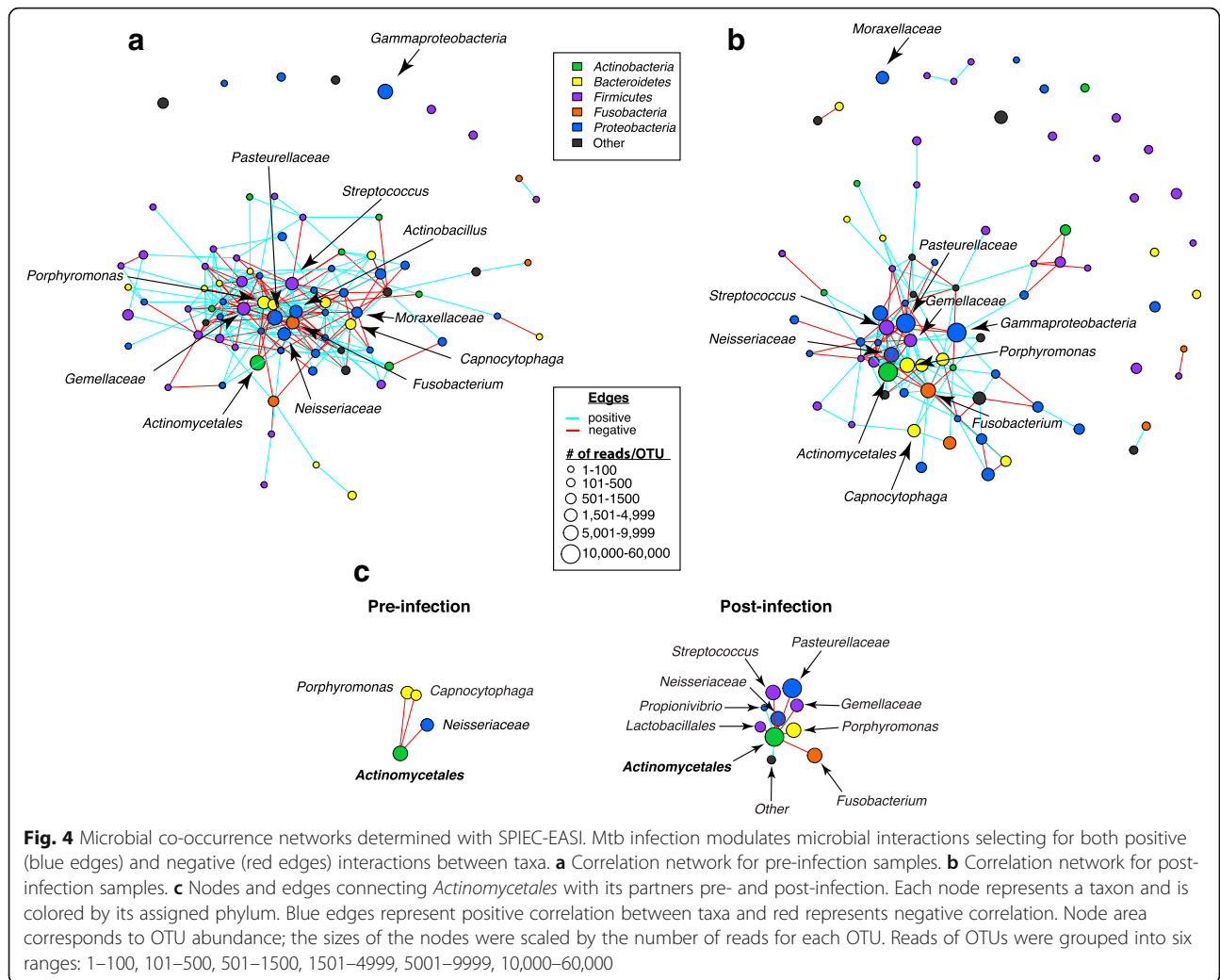


Fig. 4 Microbial co-occurrence networks determined with SPIEC-EASI. Mtb infection modulates microbial interactions selecting for both positive (blue edges) and negative (red edges) interactions between taxa. **a** Correlation network for pre-infection samples. **b** Correlation network for post-infection samples. **c** Nodes and edges connecting *Actinomycetales* with its partners pre- and post-infection. Each node represents a taxon and is colored by its assigned phylum. Blue edges represent positive correlation between taxa and red represents negative correlation. Node area corresponds to OTU abundance; the sizes of the nodes were scaled by the number of reads for each OTU. Reads of OTUs were grouped into six ranges: 1–100, 101–500, 501–1500, 1501–4999, 5001–9999, 10,000–60,000

Table 3 Summary of edges (positive and negative) for each taxon

	Pre-infection			Post-infection		
	Total	Positive	Negative	Total	Positive	Negative
<i>Pasteurellaceae</i>	18	9	9	17	7	10
<i>Neisseriaceae</i>	12	9	3	9	6	3
<i>Actinobacillus</i>	13	9	4	8	6	2
<i>Actinomycetales</i> *	3	0	3	9	2	7
<i>Gammaproteobacteria</i>	0	–	–	6	2	4
<i>Fusobacterium</i>	18	7	11	15	8	7
<i>Streptococcus</i>	14	6	8	15	5	10
<i>Porphyromonas</i>	18	10	8	11	8	3
<i>Moraxellaceae</i>	8	2	6	0	–	–
<i>Gemellaceae</i>	17	9	8	10	5	5
<i>Capnocytophaga</i>	11	6	5	4	4	0

*Order to which Mtb belongs

HIV+ individuals with more advanced disease (baseline mean CD4 count = 262 cells/mm³) did, however, find significant alterations in both bacterial diversity and composition compared to an uninfected population [50]. The observations from the human studies, the macaque SHIV model, and the findings in our Mtb-infected cynomolgus macaques suggest that the bacterial landscape within the airway is a highly variable mucosal surface with indications of both resistance and plasticity that is influenced by the type of pathogen, the initial status of the lung, the severity of disease, host inflammation, and the incidence of exacerbation [13, 14, 47]. However, there are key features of tuberculosis that may contribute to the relatively modest overall changes in airway microbiota we observed following Mtb infection. First, Mtb is a relatively slowly evolving infection, since the bacteria replicate quite slowly in vitro and in vivo; our previous studies suggested a doubling time in macaque lung of ~40 h in the first month of infection [51]. Acute lung infections could have a stronger impact on airway microbiome composition [14, 52]. Second, TB is primarily a parenchymal infection rather than a strict airway infection, so there may be stronger effects on the actual lung vs. the airway microbiota. It is possible that much more severe TB disease, including cavitary disease, would have a larger impact on the airway microbiome. Third, Mtb infection results in the formation of granulomas, which are designed to compartmentalize the infection, protecting the lung from further infection spread, and thus effects on microbiome may be more local. Finally, unlike the gut, the lungs and airways are sparsely populated by microbes, and there may be adequate space for new microbial agents to enter and replicate without major changes to the microbial residents. A better understanding of the interaction of the lung microbiota, TB disease pathogenesis, and host response will rely on a more targeted exploration of the local, granuloma microenvironment [18, 53, 54], as well as sequencing approaches including transcriptomics and metagenomics that complement and enrich 16S rRNA gene sequence analyses.

Conclusions

By serially sampling the BAL of both infected and uninfected lung lobes in macaques over the course of Mtb infection, we have provided an initial survey into how Mtb modifies the microbial environment of the lung. We observed alteration of the lung microbiota post challenge in early infection as well as significant intra-host variation that reflects the spectrum of outcomes observed in TB. We identified several changes among the microbial interactome that posit an Mtb-mediated reshuffling of intra-microbial relationships. However, despite the presence of a wide range of pulmonary

inflammation in the macaques studied here, there was a relatively modest effect overall on the airway microbiome in response to Mtb infection, and the effects observed were quite variable among subjects.

Methods

Macaque infections and imaging, microbiome experimental design, and sample collection

In this study, we analyzed oral washes (OW), bronchoalveolar lavages (BALs), and bronchoscope control samples of 26 cynomolgus macaques (*Macacca fascicularis*) obtained from Valley Biosystems (Sacramento, CA). All animals underwent baseline blood and chemical profiles and were housed in biosafety level 3 facilities in accordance with the standards of the Animal Welfare Act and the *Guide for the Care and Use of Laboratory Animals*. All macaque procedures and protocols were approved by the University of Pittsburgh's Institutional Animal Care and Use Committee.

The macaques were infected with <25 CFU Mtb strain Erdman via bronchoscopic instillation into the right lower lung, as previously described [39]. A 5 mL oral wash was first obtained from the cheek pouch. A 5 mL saline wash of the sterilized bronchoscope was also obtained to detect any residual DNA within the scope. The mouth was swabbed with an antiseptic agent (chlorohexane) immediately prior to endotracheal insertion of the bronchoscope. Saline lavage (7 mL each lobe) of the right and left lower lung lobes was performed with scope sterilization between entering each lobe. On average, 4–5 mL were obtained from both sides of the lung (left and right lobes) for each macaque at 2 weeks before infection (considered the baseline), and 1, 4, and 5 months post-infection. The samples were stored at –80 °C until they were processed (Table 1). Macaques were serially imaged throughout infection with 2-deoxy-2-[¹⁸F]fluoro-D-glucose positron emission tomography and computed tomography ([¹⁸F]-FDG PET/CT), as previously published [29, 54, 55], to track inflammation and disease progression. Total lung inflammation (total FDG activity) is a cumulative summation of [¹⁸F]-FDG uptake (as measured by standard uptake value [SUV] with positive PET above 2.3) [35, 56] (Table 2).

DNA isolation, amplification, and sequencing

DNA was isolated from 5 mL of the oral wash, 5 mL of the bronchoscope wash, and 4–5 mL lavage from each lobe using the PowerSoil® DNA Isolation Kit (MO BIO Laboratories Inc.). DNA isolations were performed at the University of Pittsburgh in a dedicated biosafety level 3 cabinet that was first exposed to a UV light source for 15 min. Reagent and sterile PBS were also processed through the DNA isolation procedure as negative controls. The DNA was eluted in 100 µL elution buffer. Each sample was quantified with the Qubit 2.0

Fluorometer (ThermoFisher Scientific Inc.). The samples were stored at $-20\text{ }^{\circ}\text{C}$ until shipment to New York University for analysis.

Extracted DNA was used in a PCR reaction with sequencing primers targeting the V4 hypervariable region of the bacterial 16S ribosomal RNA (rRNA) gene. 6 μL of DNA were amplified in a final volume of 25 μL with 0.35 μL Q5 Hot Start High-Fidelity DNA Polymerase (New England BioLabs Inc.), 5 μL 5X Q5 Buffer, 0.5 μL dNTP and 5 μM primers (515F/806R) [57], 5 μM barcoded reverse primer and water. Cycling conditions were 94 $^{\circ}\text{C}$ for 2 min, then 33 cycles at 94 $^{\circ}\text{C}$ for 30 s, 55 $^{\circ}\text{C}$ for 30 s, and 72 $^{\circ}\text{C}$ for 30 s, followed by 72 $^{\circ}\text{C}$ for 10 min.

Reactions were cleaned up using 0.65 \times volume of AMPure XP Beads (Agencourt) and eluted into 20 μL low TE, pH 8.0 on the Bravo Automated Liquid Handling Platform (Agilent Technologies). Eluted PCR products were quantified with a Quant-iT double-stranded DNA (ds-DNA) High-Sensitivity Assay Kit (Invitrogen) according to the manufacturer's instructions and combined into a pool with equal amounts of each amplicon. Samples were then pooled and each pool was re-purified with 0.65 \times volume of AMPure XP Beads (Agencourt). All pools were re-run on a HSD1000 ScreenTape with the TapeStation instrument (Agilent Technologies) and quantified with the Qubit 2.0 Fluorometer. Three pools were combined into a final pool with equal amounts of each amplicon and re-purified with 0.65 \times volume of AMPure XP Beads to ensure library quality. The pool was quantified by qPCR with the KAPA Library Quantification Kit (KAPA Biosystems) for the Roche 480 Light-Cycler system. The final library was 2 \times 250 bp paired-end sequenced on the MiSeq Illumina Sequencer at the Genomics Core Facility of the Center for Genomics and Systems Biology (CGSB), New York University.

Sequence data processing and analysis

The 16S rRNA gene sequences were processed using the Quantitative Insights into Microbial Ecology (QIIME) pipeline for analysis of microbiome data [58]. The reads were end-paired with the 1.1.2 *ea-utils* [59], filtered, and de-multiplexed. Chimeric sequences were identified and filtered out with the ChimeraSlayer sequence detection tool [60]. Sequences were clustered into Operational Taxonomic Units (OTUs) at a 97% similarity threshold with the open-reference approach, which begins by running a close-reference step followed by a de novo step that clustered the sequences that failed closed-reference assignment. Taxonomic assignments of the sequences were made using the RDP (Ribosomal Database Project) classifier [61], which classifies sequences to the species level, aligning the reference sequences picked for each OTU against the 16S rRNA gene database Greengenes 13_5

(version of May 2013) [62]. Assignments of taxonomy were refined through phylogenetic methods using PyNAST [63] to generate an alignment for each cluster from which a phylogenetic tree was generated using the FastTree approximately maximal likelihood method [64]. We identified *Burkholderia* as being one of the dominant genera in the bronchoscope controls, and so any OTU matching that genus was removed from the analyses. After removal of *Burkholderia*, we identified 16,218 OTUs (from an original total of 16,243) with a mean of 26,600 reads per sample and 511 genus-level OTUs. Singletons and OTUs present in less than 10% of the samples were filtered out of the dataset to reduce statistical noise, leaving the 98 most abundant OTUs for downstream analyses. Samples with fewer than 1000 reads were removed, and all others were subsampled/rarified to 1000 reads.

The QIIME output data were imported in RStudio (Version 1.0.136) with the Bioconductor package phyloseq [65], including subsetting, normalizing, and plotting of the input data. The taxonomic and OTU-based profiles were used in a series of ordination, clustering, and community diversity analyses designed to identify significant shifts in 16S rRNA gene profiles between samples across time points and for further statistical analyses. Alpha diversities of different groups, calculated as the inverse Simpson Index, were compared using the Wilcoxon signed-rank test. The beta-diversity was calculated as Bray-Curtis dissimilarity [66] and analyzed/compared using t-Distributed Stochastic Neighbor Embedding (t-SNE) [67] which allows dimensionality reduction.

For OTU differential relative abundance analysis, the Linear Discriminant Analysis Effect Size (LEfSe) [68] method was applied. This method consists of a Kruskal-Wallis test followed by subsequent Wilcoxon rank-sum tests on subgroups. This analysis was performed using the Huttenhower lab online Galaxy web application [69]. For the association analysis between microbial abundance and lung inflammation, Multivariate Association with Linear Models (MaAsLin) analysis was applied [70]. This analysis was performed using the MaAsLin R package (https://bitbucket.org/biobakery/maaslin/downloads/Maaslin_0.0.4.tar.gz).

Ecological correlation networks were constructed to identify interactions of the microbial community associated with Mtb infection. Taxa counts were normalized using total sum scaling (also known as relative abundance) followed by centered log ratio scaling [71]. Each network was built using SParse Inverse Covariance estimation for Ecological Association Inference (SPIE-C-EASI) package version 0.1 in R (<https://github.com/zdk123/SpiecEasi>) [37, 36]. The sparse graphical lasso (glasso) setting was used and the optimal sparsity parameter was selected, based on the Stability Approach to Regularization Selection (StARS) [72]. The StARS

variability threshold was set to 0.1 for all networks. Networks were analyzed using functions of the R package *igraph* version 1.0.1 [73].

Community types [74] of the lung microbiome were identified using the Dirichlet Multinomial Mixture (DMM) model [34], implemented in *mothur* [75]. The Laplace approximation was used for selecting the best number of community types, and the optimal number of community types was found when the minimum Laplace value was observed.

Additional files

Additional file 1: Figure S1. tSNE showing clustering of oral and BAL samples. Yellow: oral wash (OW); Red: bronchoalveolar lavage samples (BAL); triangles: pre-infection samples; circles: post-infection samples. (PDF 132 kb)

Additional file 2: Figure S2. tSNE of the lung microbiota, measured by Bray-Curtis distances. The different colors correspond to the different time points. Yellow: baseline, $n = 52$ samples; orange: month 1, $n = 47$; light blue: month 4, $n = 46$; dark blue: month 5, $n = 27$. (PDF 702 kb)

Additional file 3: Figure S3. Relative abundance of main taxa in the lung microbiota. For most monkeys, the microbial composition is relatively similar between right and left lobes within the same monkey, while the microbiota across monkeys is divergent. b: baseline; 1, 4, 5: months post-infection. LL: left lower lobe; RL: right lower lobe. (PDF 287 kb)

Abbreviations

[¹⁸F]-FDG: 2-deoxy-2-[¹⁸F]fluoro-D-glucose; 16S rRNA gene: 16S subunit of the ribosomal RNA gene; BAL: Bronchoalveolar lavage; CT: Computed tomography; DMM: Dirichlet multinomial mixture; LEM: Linear discriminant analysis effect size; MaAsLin: Multivariate Association with Linear Models; Mtb: *Mycobacterium tuberculosis*; OTU: Operational taxonomic unit; PET: Positron emission tomography; QIIME: Quantitative Insights into Microbial Ecology; SPIEC-EASI: Sparse Inverse Covariance estimation for Ecological Association Inference (pronounced as "speak easy"); STARS: Stability Approach to Regularization Selection; TB: Tuberculosis; t-SNE: t-Distributed Stochastic Neighbor Embedding

Acknowledgements

We thank the members, laboratory technicians, and veterinary staff of the Flynn, Lin, and Ghedin laboratories for their technical expertise and assistance with animal care, sample processing, bioinformatics analysis, and study design.

Funding

This work was made possible by support from the National Institutes of Health, R03 AI122067 (JLF, PLL, and EG), R01 AI111871 (PLL), T32 AI089443 (AMC), and the Bill and Melinda Gates Foundation (JLF).

Availability of data and materials

The sequencing data in this study are available in the Sequence Read Archive (SRA) under accession number SRP134056 (BioProject accession PRJNA436553).

Authors' contributions

AMC developed the concept, analyzed and interpreted data, performed animal experiments, collected samples, and wrote the manuscript. YM performed the sequencing experiments, analyzed and interpreted data, and wrote the manuscript. TD analyzed the sequence data, performed the statistical analyses, and interpreted data. MB performed sample collection and data analysis. PM performed PET/CT analysis. AG did sample processing, sequence library preparation, and data analysis. PLL, JLF, and EG developed the concept, supervised all analyses and interpretation. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. ²Present address: Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. ³Center for Genomics and Systems Biology, Department of Biology, New York University, New York, USA. ⁴Present address: Department of Pediatrics, University of California, San Diego School of Medicine, La Jolla, California, USA. ⁵Department of Pediatrics, Children's Hospital of Pittsburgh of University of Pittsburgh Medical Center, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. ⁶College of Global Public Health, New York University, New York, NY, USA.

Received: 23 March 2018 Accepted: 20 September 2018

Published online: 09 October 2018

References

1. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. *Nat Rev Genet.* 2012;13:260–70.
2. Honda K, Littman DR. The microbiome in infectious disease and inflammation. *Annu Rev Immunol.* 2012;30:759–95.
3. Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol.* 2009;9:313–23.
4. Roy S, Trinchieri G. Microbiota: a key orchestrator of cancer therapy. *Nat Rev Cancer.* 2017;17:271–85.
5. Thaiss CA, Zmora N, Levy M, Elinav E. The microbiome and innate immunity. *Nature.* 2016;535:65–74.
6. Brennan CA, Garrett WS. Gut microbiota, inflammation, and colorectal cancer. *Annu Rev Microbiol.* 2016;70:395–411.
7. Relman DA. Microbial genomics and infectious diseases. *N Engl J Med.* 2011;365:347–57.
8. Kane M, et al. Successful transmission of a retrovirus depends on the commensal microbiota. *Science.* 2011;334:245–9.
9. Kuss SK, et al. Intestinal microbiota promote enteric virus replication and systemic pathogenesis. *Science.* 2011;334:249–52.
10. Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. *Nat Rev Immunol.* 2016;16:341–52.
11. Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JL. Human nutrition, the gut microbiome and the immune system. *Nature.* 2011;474:327–36.
12. Marsland BJ, Gollwitzer ES. Host-microorganism interactions in lung diseases. *Nat Rev Immunol.* 2014;14:827–35.
13. Dickson RP, Martinez FJ, Huffnagle GB. The role of the microbiome in exacerbations of chronic lung diseases. *Lancet.* 2014;384:691–702.
14. Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB. The microbiome and the respiratory tract. *Annu Rev Physiol.* 2016;78:481–504.
15. Iliev ID, Leonardi I. Fungal dysbiosis: immunity and interactions at mucosal barriers. *Nat Rev Immunol.* 2017;10:635–646.
16. Zumla A, Raviglione M, Hafner R, von Reyn CF. Tuberculosis. *N Engl J Med.* 2013;368:745–55.
17. Barry CE 3rd, et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Microbiol.* 2009;7:845–55.
18. Cadena AM, Fortune SM, Flynn JL. Heterogeneity in tuberculosis. *Nat Rev Immunol.* 2017;17:691–702.
19. Getahun H, Chaisson RE, Raviglione M. Latent *Mycobacterium tuberculosis* infection. *N Engl J Med.* 2015;373:1179–80.
20. World Health Organization. Global tuberculosis report 2015. Geneva: World Health Organization; 2015.

21. Cadena AM, Flynn JL, Fortune SM. The Importance of First Impressions: Early Events in Mycobacterium tuberculosis Infection Influence Outcome. *mBio*. 2016;7:e00342–16.
22. Cheung MK, et al. Sputum microbiota in tuberculosis as revealed by 16S rRNA pyrosequencing. *PLoS One*. 2013;8:e54574.
23. Cui Z, et al. Complex sputum microbial composition in patients with pulmonary tuberculosis. *BMC Microbiol*. 2012;12:276.
24. Zhou Y, et al. Correlation between either Cupriavidus or Porphyromonas and primary pulmonary tuberculosis found by analysing the microbiota in patients' bronchoalveolar lavage fluid. *PLoS One*. 2015;10:e0124194.
25. Wu J, et al. Sputum microbiota associated with new, recurrent and treatment failure tuberculosis. *PLoS One*. 2013;8:e83445.
26. Botero LE, et al. Respiratory tract clinical sample selection for microbiota analysis in patients with pulmonary tuberculosis. *Microbiome*. 2014;2:29.
27. Winglee K, et al. Aerosol Mycobacterium tuberculosis infection causes rapid loss of diversity in gut microbiota. *PLoS One*. 2014;9:e97048.
28. Namasivayam S, et al. Longitudinal profiling reveals a persistent intestinal dysbiosis triggered by conventional anti-tuberculosis therapy. *Microbiome*. 2017;5:71.
29. Coleman MT, et al. Early changes by (18)Fluorodeoxyglucose positron emission tomography coregistered with computed tomography predict outcome after Mycobacterium tuberculosis infection in cynomolgus macaques. *Infect Immun*. 2014;82:2400–4.
30. Cho I, et al. Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature*. 2012;488:621–6.
31. Takeshita T, et al. Discrimination of the oral microbiota associated with high hydrogen sulfide and methyl mercaptan production. *Sci Rep*. 2012;2:215.
32. Campbell JH, et al. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc Natl Acad Sci U S A*. 2013;110:5540–5.
33. Karched M, Furgang D, Planet PJ, DeSalle R, Fine DH. Genome sequence of *Aggregatibacter actinomycetemcomitans* RHAA1, isolated from a rhesus macaque, an Old World primate. *J Bacteriol*. 2012;194:1275–6.
34. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One*. 2012;7:e30126.
35. Maiello P, et al. Rhesus macaques are more susceptible to progressive tuberculosis than cynomolgus macaques: a quantitative comparison. *Infect Immun*. 2018;86(2).
36. Kurtz ZD, et al. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol*. 2015;11:e1004226.
37. Tipton L, et al. Fungi stabilize connectivity in the lung and skin microbial ecosystems. *Microbiome*. 2018;6:12.
38. Williams KP, et al. Phylogeny of gammaproteobacteria. *J Bacteriol*. 2010;192:2305–14.
39. Capuano SV, et al. Experimental Mycobacterium tuberculosis infection of Cynomolgus macaques closely resembles the various manifestations of human M. tuberculosis infection. *Infect Immun*. 2003;71:5831–44.
40. Lin PL, et al. Early events in Mycobacterium tuberculosis infection in cynomolgus macaques. *Infect Immun*. 2006;74:3790–803.
41. Lin PL, et al. Quantitative comparison of active and latent tuberculosis in the cynomolgus macaque model. *Infect Immun*. 2009;77:4631–42.
42. Cui L, et al. Topographic diversity of the respiratory tract mycobiome and alteration in HIV and lung disease. *Am J Respir Crit Care Med*. 2015;191(8):932–42.
43. Cui L, et al. The microbiome and the lung. *Ann Am Thorac Soc*. 2014;4(11 Suppl):S227–32.
44. Morris A, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med*. 2013;187:1067–75.
45. Twigg HL 3rd, et al. Use of bronchoalveolar lavage to assess the respiratory microbiome: signal in the noise. *Lancet Respir Med*. 2013;1:354–6.
46. Goodman AL, et al. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc Natl Acad Sci U S A*. 2011;108:6252–7.
47. Huffnagle GB, Dickson RP, Lukacs NW. The respiratory tract microbiome and lung inflammation: a two-way street. *Mucosal Immunol*. 2016;10(2):299–306.
48. Morris A, et al. Longitudinal analysis of the lung microbiota of cynomolgus macaques during long-term SHIV infection. *Microbiome*. 2016;4:38.
49. Beck JM, et al. Multicenter comparison of lung and Oral microbiomes of HIV-infected and HIV-uninfected individuals. *Am J Respir Crit Care Med*. 2015;192:1335–44.
50. Twigg HL 3rd, et al. Effect of advanced HIV infection on the respiratory microbiome. *Am J Respir Crit Care Med*. 2016;194:226–35.
51. Ford CB, et al. Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. *Nat Genet*. 2011;43:482–6.
52. Dickson RP, Erb-Downward JR, Huffnagle GB. Towards an ecology of the lung: new conceptual models of pulmonary microbiology and pneumonia pathogenesis. *Lancet Respir Med*. 2014;2:238–46.
53. Lin PL, et al. Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. *Nat Med*. 2014;20:75–9.
54. Martin CJ, et al. Digitally barcoding Mycobacterium tuberculosis reveals in vivo infection dynamics in the macaque model of tuberculosis. *mBio*. 2017;8(3).
55. Coleman MT, et al. PET/CT imaging reveals a therapeutic response to oxazolidinones in macaques and humans with tuberculosis. *Sci Transl Med*. 2014;6:265ra167.
56. White AG, et al. Analysis of 18FDG PET/CT imaging as a tool for studying Mycobacterium tuberculosis infection and treatment in non-human primates. *J Vis Exp*. 2017;(127):e56375.
57. Caporaso JG, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A*. 2011;108(Suppl 1):4516–22.
58. Caporaso JG, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Meth*. 2010;7:335–6.
59. Aronesty E. ea-utils: Command-line tools for processing biological sequencing data. *Durham: Expression Analysis*; 2011.
60. Haas BJ, et al. Chimeric 16S rRNA sequence formation and detection in sanger and 454-pyrosequenced PCR amplicons. *Genome Res*. 2011;21:494–504.
61. Cole JR, et al. The ribosomal database project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res*. 2008;37:D141–5.
62. De Santis TZ, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol*. 2006;72:5069–72.
63. Caporaso JG, et al. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*. 2009;26:266–7.
64. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*. 2010;5:e9490.
65. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8:e61217.
66. Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr*. 1957;27:325–49.
67. van der Maaten LJP, Hinton GE. Visualizing High-Dimensional Data using t-SNE. *J Mach Learn Res*. 2008;9:2579–605.
68. Segata N, et al. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12:R60.
69. Afgan E, et al. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res*. 2018;46:W537–44.
70. Morgan XC, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol*. 2012;13:R79.
71. Aitchison J. A new approach to null correlations of proportions. *Math Geol*. 1981;13:175–89.
72. Liu H, Roeder K, Wasserman L. Stability approach to regularization selection (stars) for high dimensional graphical models. In: *Advances in neural information processing systems*; 2010. p. 1432–40.
73. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJ Complex Syst*. 2006;1695:1–9.
74. Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. *Nature*. 2014;509:357–60.
75. Schloss PD, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75:7537–41.