OPEN ACCESS

Original research

# More than *Mycobacterium tuberculosis:* site-of-disease microbial communities, and their functional and clinical profiles in tuberculous lymphadenitis

Georgina R Nyawo,[1,2] Charissa C Naidoo,[1,2] Benjamin Wu,[3] Imran Sulaiman,[3] Jose C Clemente,[4] Yonghua Li,[3] Stephanie Minnies,[1] Byron W P Reeve,[1] Suventha Moodley,[1,2] Cornelia Rautenbach,[5,6] Colleen Wright,[7] Shivani Singh,[3] Andrew Whitelaw,[5,6] Pawel Schubert,[5,7] Robin Warren,[1] Leopoldo Segal,[3] Grant Theron [1,2]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/thorax-2022-219103).

For numbered affiliations see end of article.

**Correspondence to**
Professor Grant Theron, DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Stellenbosch University Faculty of Medicine and Health Sciences, Cape Town, Western Cape, South Africa; gtheron@sun.ac.za

TBScience 2022, hosted at the Union World Conference on Lung Health 2022.

Received 20 April 2022
Accepted 31 October 2022



**To cite:** Nyawo GR, Naidoo CC, Wu B, *et al*. *Thorax* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/thoraxjnl-2022-219103

## ABSTRACT

**Background** Lymphadenitis is the most common extrapulmonary tuberculosis (EPTB) manifestation. The microbiome is important to human health but uninvestigated in EPTB. We profiled the site-of-disease lymph node microbiome in tuberculosis lymphadenitis (TBL).

**Methods** Fine-needle aspiration biopsies were collected from 158 pretreatment presumptive TBL patients in Cape Town, South Africa. 16S Illumina MiSeq rRNA gene sequencing was done.

**Results** We analysed 89 definite TBLs (dTBLs) and 61 non-TBLs (nTBLs), which had similar α- but different β-diversities (p=*0.001*). Clustering identified five lymphotypes prior to TB status stratification: *Mycobacterium*-dominant, *Prevotella*-dominant and *Streptococcus*-dominant lymphotypes were more frequent in dTBLs whereas a *Corynebacterium*-dominant lymphotype and a fifth lymphotype (no dominant taxon) were more frequent in nTBLs. When restricted to dTBLs, clustering identified a *Mycobacterium*-dominant lymphotype with low α-diversity and non-*Mycobacterium*-dominated lymphotypes (termed *Prevotella-Corynebacterium*, *Prevotella-Streptococcus*). The *Mycobacterium* dTBL lymphotype was associated with HIV-positivity and features characteristic of severe lymphadenitis (eg, larger nodes). dTBL microbial communities were enriched with potentially proinflammatory microbial short-chain fatty acid metabolic pathways (propanoate, butanoate) vs nTBLs. 11% (7/61) of nTBLs had *Mycobacterium* reads BLAST-confirmed as *Mycobacterium tuberculosis* complex.

**Conclusions** TBL at the site-of-disease is not microbially homogeneous. Distinct microbial community clusters exist that, in our setting, are associated with different clinical characteristics, and immunomodulatory potentials. Non-*Mycobacterium*-dominated dTBL lymphotypes, which contain taxa potentially targeted by TB treatment, were associated with milder, potentially earlier stage disease. These investigations lay foundations for studying the microbiome's role in lymphatic TB. The long-term clinical significance of these lymphotypes requires prospective validation.

### WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Lymphadenitis is the most frequent extrapulmonary tuberculosis manifestation. The microbiome is critical for human health, however, the microbiome at the site-of-disease in patients with tuberculosis lymphadenitis is completely uncharacterised, including whether distinct microbial clusters (which we term 'lymphotypes') are associated with clinically important patient characteristics.

### WHAT THIS STUDY ADDS

⇒ Surprisingly, patients with confirmed tuberculosis lymphadenitis often had bacterial taxa other than Mycobacterium dominant at the site-of-disease (Prevotella, Streptococcus, Corynebacterium). Such patients had milder forms of disease (eg, less swelling, less HIV) whereas patients with the Mycobacterium-dominated lymphotype had increased microbial functional capacity for proinflammatory short-chain fatty acids and more severe disease.

### HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our findings have relevance for clinical staging and treatment of tuberculosis lymphadenitis, which we show to not be microbially homogeneous, and suggest that the site-of-disease in tuberculosis lymphadenitis is, prior to shifting to becoming Mycobacterium-dominated, first characterised by Prevotella, Streptococcus and/or Corynebacterium dominance and milder disease. Lastly, given that Streptococcus and Corynebacterium are themselves capable of causing lymphadenitis and susceptible to first-line TB treatment, such treatment may alleviate pathology in tuberculosis lymphadenitis by, in part, killing taxa other than Mycobacterium.

## INTRODUCTION

Tuberculosis (TB), which kills 1.5 million people globally each year (including 214 000 people

with HIV), causes extrapulmonary TB (EPTB).[1] EPTB accounts for ~16% of all TB, up to half of all TB in people living with HIV (PLHIV)[2] and has high mortality.

TB lymphadenitis (TBL) is the most common EPTB manifestation, accounting for 70% of EPTB and most frequently affects peripheral and cervical lymph nodes.[3 4] TBL occurs after *Mycobacterium tuberculosis* (*Mtb*) enters the airways, is taken up by phagocytic cells, and transported to lymph nodes where granulomas may form. These steps are also necessary for priming T-cells to generate adaptive immune responses for microbial killing mediated by cytokines and other effector mechanisms.[5]

Lymph nodes have an important role in TB pathogenesis: enlargement has been documented following exposure, even if only a fraction of patients with enlarged nodes develop active disease.[6] Animal studies show lymph nodes can be sites of TB reactivation (*Mtb* DNA found in new lung granulomas share unique DNA barcodes with *Mtb* previously only found in lymph nodes).[7] Furthermore, pathologically normal lymph nodes obtained at autopsy from humans without active TB can, when used to inoculate animals, cause active disease,[8] suggesting these lymph nodes contained live *Mtb* (and hence *Mtb* DNA). Lymph nodes are therefore hypothesised to serve as a *Mtb* growth and persistence niche[6] that can spread to bodily sites[9] (in animals lymph node infection almost always accompanies infection in the lungs[7]; suggesting that TB may primarily be a lymphatic rather than pulmonary disease.[10] For example, the lymph nodes of participants with subclinical TB pathology demonstrate enhanced metabolic activity on positron emission tomography (PET)-CT scans.[11] Together these studies show that lymph nodes have an important role in TB pathogenesis, however, the determinants of why *Mtb* sometimes successfully establishes itself in the lymph nodes and subsequently proliferates, including the potential role of other microbes, is understudied. Key to understanding this is characterising the local site-of-disease.

The microbiota modulates immune responses via microbially-derived metabolites known as short-chain fatty acids (SCFAs).[12] Enriched pulmonary SCFAs predict TB risk in HIV-infected individuals stable on ART, and ex vivo addition of butyrate inhibits *Mtb*-induced proinflammatory responses.[13] Two studies assessed lymph node microbial content,[14 15] both in mesenteric lymph nodes in Crohn's disease where reduced diversity was observed. The site-of-disease microbiome in TB is underexamined[16]: in bronchoalveolar lavage fluid (BALF), active pulmonary TB was associated with *Mycobacterium* enrichment and *Streptococcus* depletion.[17 18]

The site-of-disease microbiome in TBL (including in HIV-endemic settings where TB is common) remains uncharacterised. Therefore, given the apparent role of the lymph nodes in TB pathogenesis, and the importance of the microbiome as a modulator of immunity, we characterised the site-of-disease lymph microbiome in presumptive TBL patients from a high HIV burden setting[19] before the potentially confounding effects of antibiotic-based TB treatment.

## METHODS

### Patient recruitment and follow-up

Presumptive TBL participants (≥18 years) were recruited from Tygerberg Academic Hospital in Cape Town, SA (25 January 2017–11 December 2018). Participants were programmatically referred for a routine fine needle aspiration biopsy (FNAB) via the skin for the investigation of lymphadenopathy as described.[19] Eligible participants were not on TB treatment within 6 months. Clinical and demographic data were collected by interview and medical record review. Patients programmatically diagnosed with TBLwere initiated on treatment, and study staff assessed treatment response by telephonic follow-up ≥12 weeks. The study had no role in patient management.

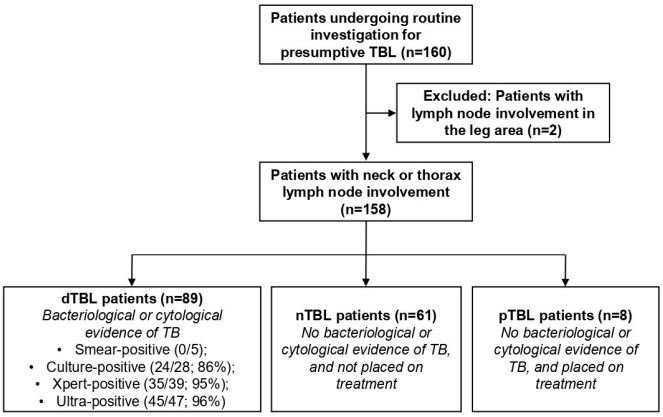



**Figure 1** Study flow chart. Fine-needle aspirates, skin and saline controls were collected from presumptive TBL patients. dTBLs, definite-TBL; MGIT960 culture, mycobacteria growth indicator tube 960 liquid culture; nTBLs, non-TBLs; pTBLs, probable TBLs; Smear: Smear microscopy; Ultra: Xpert MTB/RIF Ultra; Xpert: Xpert MTB/RIF.

### Specimen collection and processing

For each patient, two background DNA sampling controls were collected in microcentrifuge tubes prior to lymph node aspiration: a skin swab (collected into saline; Ysterplaat Medical Supplies, Cape Town, South Africa) of the site to be punctured, followed by a saline flush of the syringe to be used for aspiration. Aspiration and microbiological procedures are in online supplemental methods. Aspirated material from the third pass was collected into 500 µL sterile saline and stored at −80°C until batched DNA extraction.

### Routine specimen testing

Patients were categorised based on lymphatic or non-lymphatic mycobacteriological evidence, provided by the government programmatic laboratory (National Health Laboratory Service), and/or clinical decision to start treatment by the responsible clinician thereafter.

### Case definitions

Briefly, definite-TBLs (dTBLs) had at least one *Mtb* complex (MTBC)-positive extrapulmonary or pulmonary specimen by Xpert or culture (figure 1). Alternatively, they had site-of-disease cytology compatible with active TB. Probable-TBLs (pTBLs) did not meet dTBL criteria but commenced treatment empirically. Non-TBLs (nTBLs) had no microbiological or cytological evidence of TB. Further detail is in online supplemental table S1.

### Microbial DNA extraction and sequencing

DNA was extracted from specimens and controls using the PureLink Microbiome DNA Purification Kit (Invitrogen, Carlsbad, USA). The 16S rRNA gene V4 hypervariable region (150 bp read length) was amplified and sequenced

(paired-ends) on the Illumina MiSeq platform. Lymph, skin swab and one in five saline flushes were extracted and sequenced.

### Microbiome data analysis

16S rRNA gene sequences were processed, denoised and analysed in Quantitative Insights Into Microbial Ecology (QIIME 2, v2020.8)[20] and DADA2[21] using closed-reference picking by assigning taxonomy at a 97% similarity against representative sequences in Greengenes (V.13.8).[22] QIIME2 outputs (phylogenetic tree, feature table, taxonomy) and metadata were imported into R (V.3.5.2) and analyses done using *phyloseq*.[23] Shannon's index was calculated with *vegan*[24] as measure α-diversity (within-sample diversity). Bray-Curtis distances were calculated as a measure of β-diversity (between-sample diversity) and were visualised as principal coordinate analysis plots. Dirichlet-Multinomial Mixtures (DMM) modelling was done to estimate the optimal number of clusters based on microbial compositional similarity.[25] These clusters are herewith referred to as 'lymphotypes'.

### Inferred metagenome

Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) V.2.1.3-b[26] was used to predict gene family abundance with PICRUSt2 default options (picrust2_pipeline.py). The resulting gene table was mapped against the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database, and pathway abundances were inferred from predicted KEGG ORTHOLOGY (KO) abundances.

### Differential abundance analysis

Differentially abundant taxa and pathways were identified using *DESeq2* (V.1.22.2), which internally corrects and normalises data.[27] Feature tables were pruned to have a mean relative abundance ≥5% in 0.5% of samples.[28] *DESeq2* was run on PICRUSt outputs to identify common pathways in oL4 versus each lymphotype (overall patients), and dL3 versus each other lymphotype (dTBLs). *DESeq2* outputs with abundances and significance values for each discriminatory taxon and pathway were obtained (see online supplemental material: *DESeq2* Tables). A false discovery rate (FDR)-adjusted p≤0.2 and ≤0.05 was considered significant for taxa and pathways, respectively.

### Statistical analyses

Statistical analysis was done in GraphPad Prism V.7 (GraphPad Software, USA), STATA V.16 (StataCorp) and R V.4.2 (R Core Team, 2022). The proportions test was done to determine whether a specific variable was more frequent in different groups (eg, patients of different TB status).[29] For analysis of microbiome data, non-parametric tests were used as microbiome data are not normally distributed.[30] The Mann-Whitney or Wilcoxon signed rank test was used for unpaired and paired comparisons between two groups respectively (eg, α-diversity). Kruskal-Wallis with Dunn's test was used for comparison involving more than two groups (eg, relative abundance comparisons). Spearman's rank correlation was used to measure the association between mycobacterial relative abundance and continuous variables (eg, lymph node size). Permutational multivariate analysis of variance (PERMANOVA) was computed with 999 permutations for β-diversity differences, and $R^2$ used to measure the proportion variation explained by a variable.[20] The Benjamini-Hochberg procedure was used to correct for multiple comparisons by controlling for FDR.[28] For analysis of continuous variables in different groups (eg, lymph node size different TB status), the D'Agostino-Pearson omnibus normality test was done to evaluate normality, and the relevant parametric or nonparametric test was chosen based on the normality test. A p≤0.05 was considered significant for all comparisons, unless otherwise specified.

**Table 1** Demographic and clinical characteristics of patients with presumptive TBL

| | Patients with presumptive TB (n=158) | | | |
|---|---|---|---|---|
| | Total (n=158)* | dTBL (n=89) | nTBL (n=61) | P value |
| Age, years | 36 (21–44) | 35 (29–40) | 38 (30–49) | 0.053 |
| Female | 85/159 (53) | 48/89 (54) | 35/61 (57) | 0.677 |
| HIV† | 77/156 (49) | 49/89 (55) | 23/59 (39) | 0.055 |
| CD4+cells/μL | 166 (90–308) | 155 (76–251) | 250 (139–458) | **0.027** |
| CD4+<200 cells/μL | 47/77 (61) | 32/49 (65) | 11/23 (48) | 0.159 |
| On ART† | 38/76 (50) | 21/49 (43) | 14/22 (64) | 0.105 |
| Previous TB† | 36/156 (23) | 24/88 (27) | 9/60 (15) | 0.078 |
| Pulmonary TB | 27/36 (75) | 17/24 (71) | 8/9 (89) | 0.281 |
| Extrapulmonary TB | 9/36 (25) | 7/24 (29) | 1/9 (11) | 0.281 |
| Tobacco smoking† | 44/157 (28) | 21/89 (24) | 22/60 (37) | 0.084 |
| Antibiotic use within 1 year of recruitment† | 41/155 (26) | 22/87 (25) | 16/60 (27) | 0.851 |
| At recruitment | 24/41 (59) | 10/22 (45) | 11/16 (69) | 0.154 |
| Lymph node characteristics: sites | | | | |
| Neck | 138/158 (87) | 78/89 (88) | 55/61 (90) | 0.632 |
| Deep anterior cervical | 65/138 (47) | 36/78 (46) | 24/55 (47) | 0.774 |
| Deep lateral cervical | 25/138 (18) | 15/78 (19) | 10/55 (18) | 0.879 |
| Superficial | 15/138 (11) | 6/78 (8) | 9/55 (16) | 0.120 |
| Supraclavicular | 19/138 (14) | 16/78 (21) | 3/55 (5) | **0.015** |
| Head | 13/138 (10) | 4/78 (5) | 9/55 (16) | **0.032** |
| Thorax | 20/158 (13) | 11/89 (12) | 6/61 (10) | 0.632 |
| Axillary (vs breast) | 16/21 (81) | 9/11 (82) | 3/5 (60) | 0.350 |
| Lymph node characteristics: size, $cm^2$ | 4 (2–9) | 4 (2–9) | 4 (4–9) | 0.150 |
| Specimen appearance | | | | |
| Bloody (vs chylous) | 130/158 (82) | 66/89 (72) | 57/61 (93) | **0.003** |

Bolded items indicate that p values are significant at p<0.05.
Pulmonary or extrapulmonary previous TB refers to the most recent prior TB episode.
Data are n/N (%) or median (IQR).
*Probable TBLs (pTBLs) excluded from table.
†Missing data: HIV (n=2); On ART (n=1); previous TB (n=2); smoking (n=1); antibiotic use within 1 year of recruitment (n=3).
ART, antiretroviral therapy; dTBLs, definite tuberculous lymphadenitis; nTBLs, non tuberculous lymphadenitis; pTBLs, probable-TBLs.

## RESULTS

### Cohort characteristics

We had 89 dTBLs, 61 nTBLs (figure 1) and 8 pTBLs (henceforth excluded due to small n), the characteristics of which are in table 1.

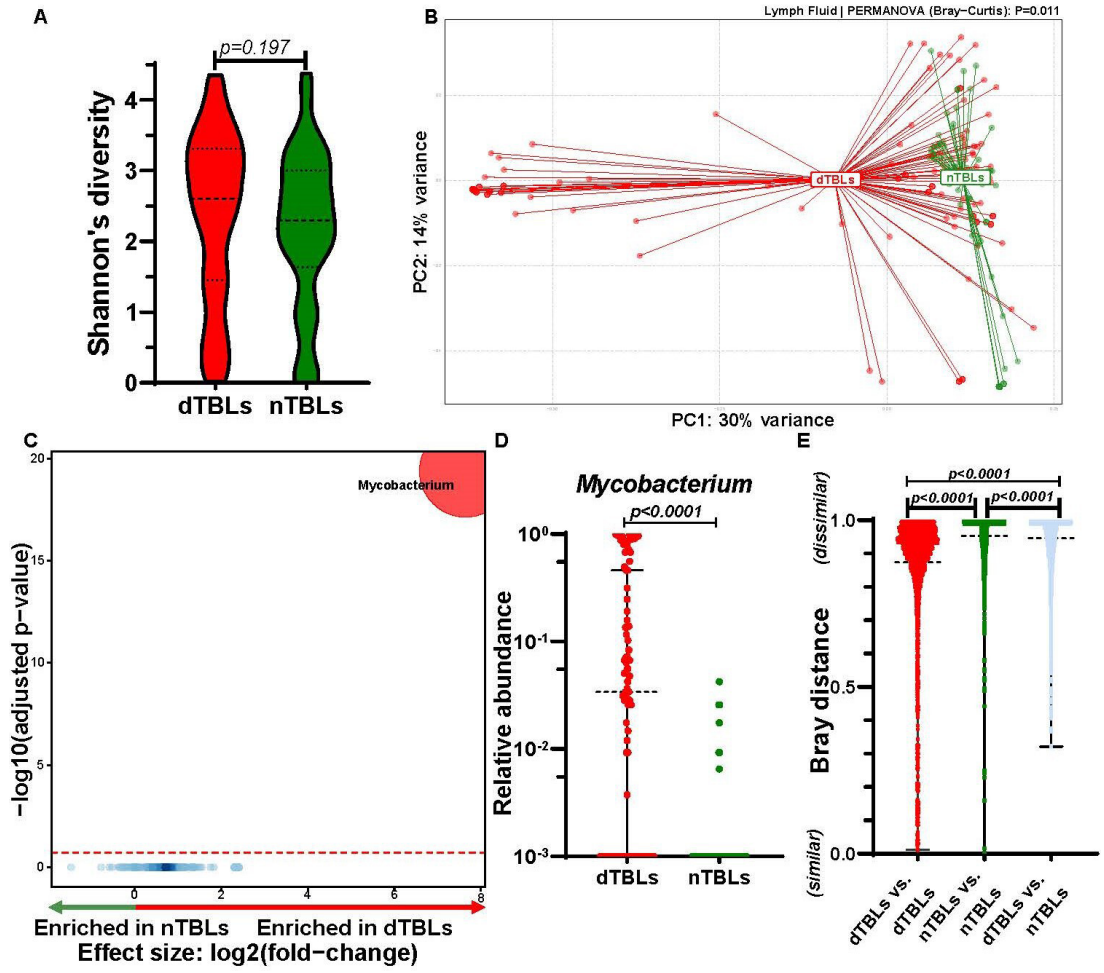



**Figure 2** dTBLs have a distinct microbiome to nTBLs with *Mycobacterium* enrichment. (A) Although α-diversity was similar, (B) β-diversity differed. *Mycobacterium* was enriched in dTBLs compared with nTBLs based on (C) differential abundance testing and (D) relative abundance. Discriminatory taxa appear above the threshold (red dotted line, FDR=0.2). (E) dTBLs were more compositionally similar to each other than nTBLs. dTBLs, definite-TBLs; FDR, false discovery rate; nTBLs: non-TBL; PERMANOVA, permutational multivariate analysis of variance; TBL, tuberculous lymphadenitis.

### Lymph microbiome is distinct from background sampling controls

To assess the degree of potential carry-over from skin commensals and background DNA on equipment used for biopsies, two background DNA sampling controls were collected and subjected to the same procedures as the actual samples. We compared the microbiome in skin and saline flushes with that of the lymph fluid. Lymph fluid had similar α-diversity to background controls but different β-diversity resulting from an enrichment of *Mycobacterium* (online supplemental figure S1A–D), thus background contamination is unlikely.

### *Mycobacterium* enrichment in dTBLs drives differences with nTBLs

We evaluated the overall difference between the microbial communities of TB groups by comparing their microbial diversity and composition. α-Diversity was similar in dTBLs and nTBLs (figure 2A), but β-diversity differed and *Mycobacterium* was the most discriminatory taxon (figure 2B,C; online supplemental figure S2 has similar comparisons with pTBLs included) appearing at several fold higher frequencies than in nTBLs (figure 2D). When patients with antibiotic use within the last year were excluded, α-diversity differences by TB status were detected (lower in dTBLs, online supplemental results). Bray distances within nTBLs were greater than within dTBLs (figure 2E), thus dTBLs were more like each other than nTBLs to each other (likely reflecting the mixture of different disease pathologies in the nTBLs and relative homogeneity of dTBLs). These results show that lymph microbial communities differ in dTBLs and nTBLs, and the microbiome of TBL is characterised by a significant enrichment of *Mycobacterium*.

### MTBC DNA found in tuberculous and nontuberculous lymph nodes

*Mycobacterium* reads were present in 64% (57/89) of dTBLs and 11% (7/61; $p<0.0001$) of nTBLs (online supplemental figure S3) and, when sequences underwent BLAST, all reads matched with *Mtb*, suggesting that none of these patients had environmental mycobacteria. There was a higher relative abundance of *Mycobacterium* reads in dTBLs (0.034 (IQR 0.001–0.460) vs 0.001 (0.001–0.001), $p<0.0001$; figure 2D), and the 16S rRNA gene sequencing positively correlated with TB diagnostic tests, but not with lymph node size (online supplemental figure S4A,B). These results suggest that MTBC DNA is found in most dTBL lymph nodes and occasionally occurs in nTBL lymph nodes.

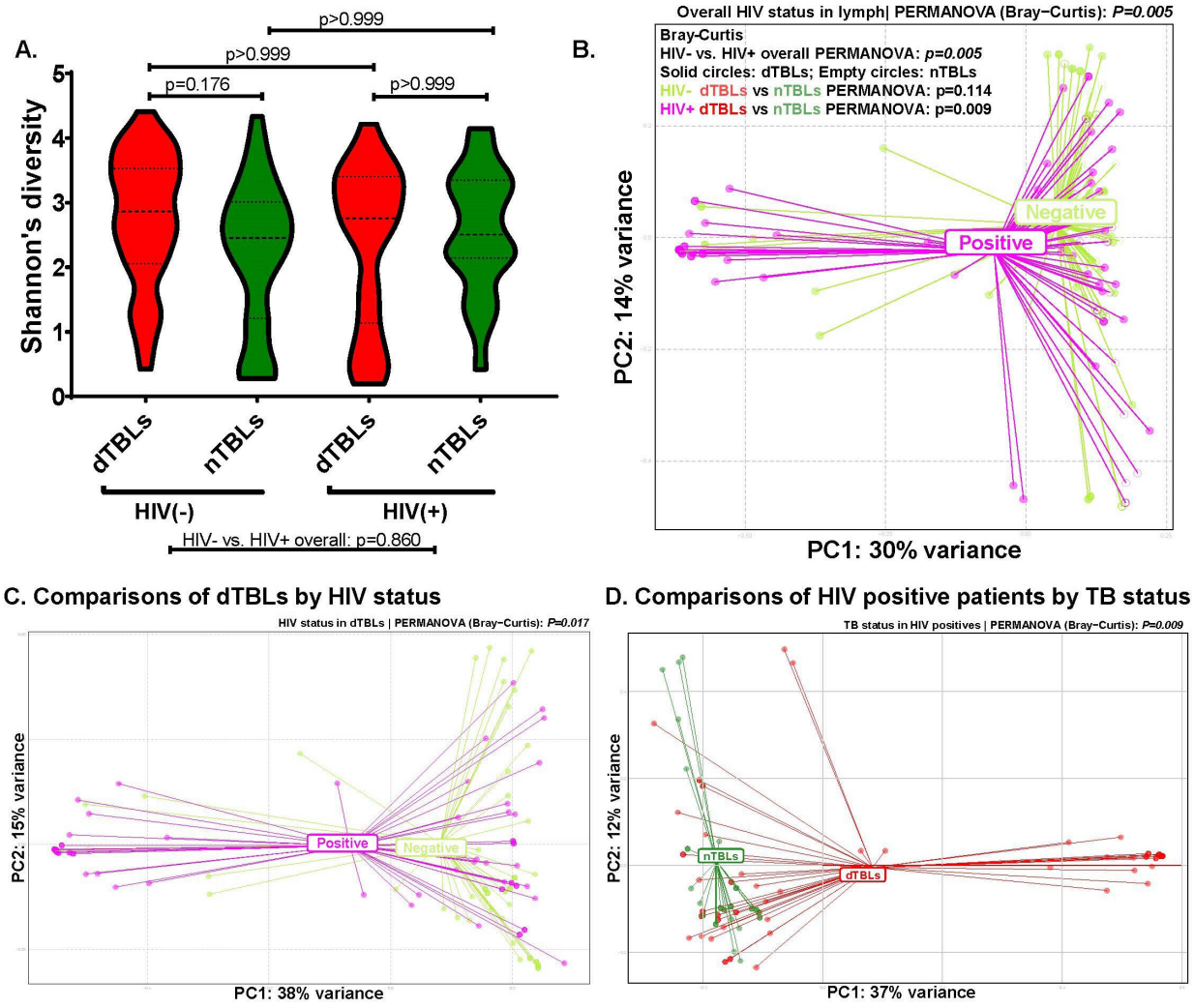



**Figure 3** Microbiome differences in HIV-positive dTBLs versus nTBLs but not in HIV-negative dTBLs vs nTBLs. (A) α-Diversity did not differ by HIV or TBL statuses, (B) however, β-diversity differed between HIV-positives and -negatives overall (shaded circles are dTBLs, empty circles are nTBLs). β-diversity differed (C) by HIV status in dTBLs only and (D) by TBL status in HIV-positives only. d-TBLs, definite TBLs; nTBLs, non-TBLs; PERMANOVA, permutational multivariate analysis of variance; TBL, tuberculous lymphadenitis.

### Differences by HIV status

HIV is a known risk factor for TB. We assessed its association with the lymph microbiome first in all patients irrespective of TB status and next within dTBLs or nTBLs. Overall, α-diversity did not differ by HIV status (figure 3A), but β-diversity did (figure 3B). β-diversity differences by HIV status persisted within dTBLs (p=0.017, figure 3C) but not nTBLs. In people with the same HIV status, β-diversity differed between dTBLs vs nTBLs only in HIV-positives (p=*0.009*, figure 3D) where dTBLs were *Mycobacterium*-enriched (online supplemental figure S5B). In PERMANOVA analyses, HIV status was only significantly associated with β-diversity in dTBLs and not nTBLs (online supplemental table S2).

### Lymphotype identification and their associations with clinical characteristics

We further explored this data using DMM to identify potential clusters in the TBL microbiome. These clusters were termed 'lymphotypes', and we evaluated associations between each lymphotype(s) and patients' clinical characteristics.

Overall: We examined whether all patients could be grouped into distinct lymphotypes; these were termed overall lymphotypes (oLs). Five oLs with differing α-diversities and β-diversities were identified (figure 4A–C, online supplemental table S3), with the *Mycobacterium*-dominated (figure 4D) oL4 showing the least α-diversity. While no taxa were differentially abundant in oL1 versus other oLs (online supplemental figure S6A–C), oL2, oL3 and oL5 were enriched relative to oL4 in *Corynebacterium*, *Prevotella* and *Streptococcus*, respectively (figure 4E–G). The patients in all oLs were associated with distinct clinical characteristics. The majority of nTBLs occurred in highly diverse oLs with a heterogeneous mixtures of taxa; likely reflecting the spectrum of pathologies in people with TBL ruled out. oL1 was associated with characteristics indicative of less severe lymphadenitis (less TB and HIV involvement). In contrast, oL4 was associated with characteristics resembling more severe lymphadenitis (bigger lymph nodes, chylous FNABs, previous TB, HIV (with a smaller proportion of PLHIV on ART, likely to have lower CD4 counts) and TB involvement. Therefore, in summary, oL1 appears to be associated with less severe forms of lymphadenitis, whereas oL4 was associated more severe forms (online supplemental table S4).

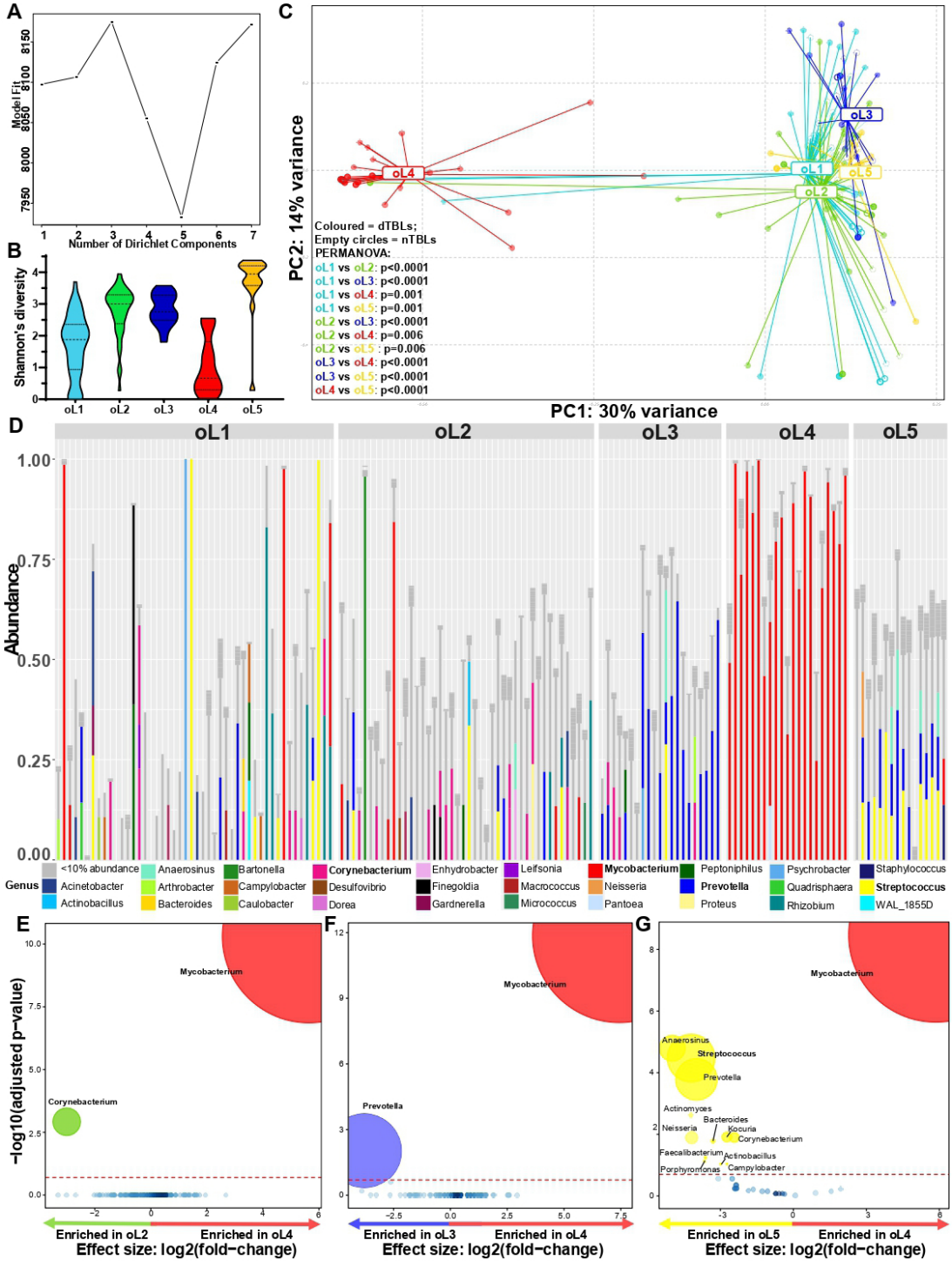



**Figure 4** Five overall lymphotypes observed in presumptive TBL. (A) Laplace approximation identified five clusters. (B) OL5 had the highest α-diversity. (C) β-diversity differed between each lymphotype (shaded circles dTBLs, empty circles nTBLs). (D) Stacked bar plots showing OL1 with a heterogeneous mixture of genera, OL2 dominated by *Corynebacterium*, OL3 dominated by *Prevotella*, OL4 dominated by *Mycobacterium*, and OL4 dominated by *Streptococcus*. Bolded taxa represent dominating taxa. (E) *Corynebacterium* was enriched in OL2; (F) *Prevotella* enriched in oL3, (G) *Mycobacterium* enriched in oL4, and *Streptococcus* enriched in OL5. Significantly more discriminatory taxa (bolded) appear closer to the left or right and higher above the threshold (red dotted line, FDR=0.2) as significance increases. Relative taxa abundance is indicated by circle size. dTBLs, definite-TBL; FDR, false discovery rate; nTBLs, non-TBL; oL, overall lymphotype; PERMANOVA, permutational multivariate analysis of variance; TBL, tuberculous lymphadenitis.

Within patients of the same TB status: We then examined whether patients within each TB group could be grouped into distinct lymphotypes. Within dTBLs, three lymphotypes (termed dTBL lymphotypes; dL) with differing β-diversities were identified (figure 5A,B), and dominated by; dL1: *Prevotella* and *Corynebacterium;* dL2: *Prevotella* and *Streptococcus*; and dL3: *Mycobacterium* (figure 5C–F). These dLs were termed *Prevotella-Corynebacterium*,

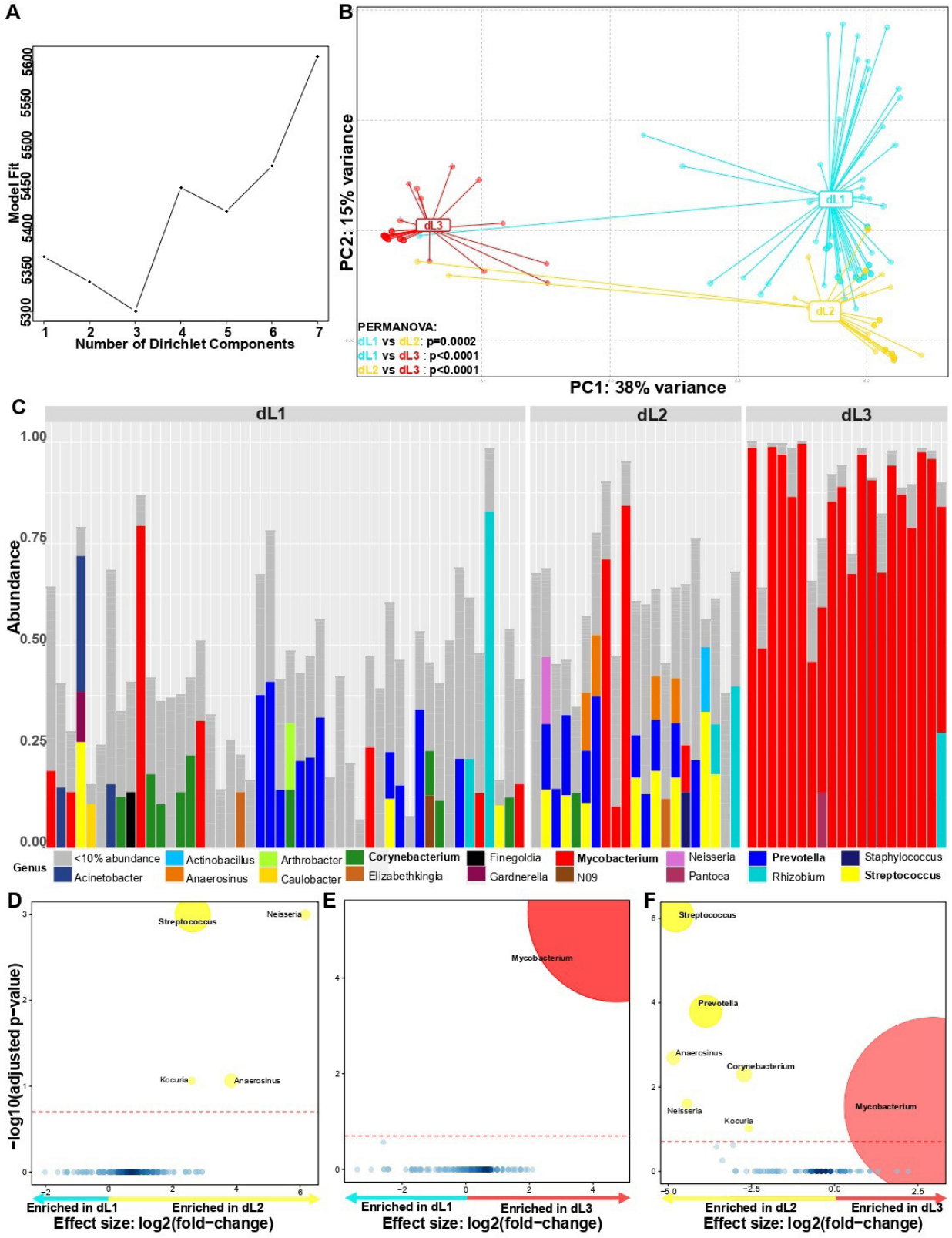



**Figure 5** Three dTBL lymphotypes identified in dTBLs. (A) Best model fit based on Laplace approximation identified three clusters within dTBLs. (B) β-diversity differed between lymphotypes. (C) Stacked bar plots showing dL1 comprised of *Mycobacterium* and accompanying heterogenous taxa, dL2 dominated by *Prevotella* and *Streptococus*, and dL3 dominated by *Mycobacterium*. Bolded taxa represent dominating taxa. (D) NO taxa were enriched in dL1, (E) L2 was enriched in *Streptococcus,* (F) and *Mycobaterium* was enriched in dL*3*. Significantly more discriminatory taxa (bolded) appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0.2) as significance increases. Relative taxa abundance is indicated by circle size. dL, dTBL lymphotype; dTBL, definite-TBL; FDR, false discovery rate; TBL, tuberculous lymphadenitis.

*Prevotella-Streptococcus* and *Mycobacterium*, respectively. dL3s were more likely to be HIV-positive compared with dL1s, with larger lymph nodes, compared with dL1s and ddL2s. Lastly, dL2s are more likely to be female than dL1s (online supplemental table S5). Together, these differences suggest dL3 is associated with more severe TBL than other dLs. Within nTBLs, no lymphotypes were identified (online supplemental figure S7).

### Predictive metagenome profiling shows increased SCFA metabolism

We further predicted the bacterial metagenome content and made functional inferences of the microbiome using the PICRUSt algorithm. Differences among pathways between groups were evaluated and visualised by *DESeq2* analysis. In dTBLs, 'fatty acid metabolism', 'benzoate degradation', 'propanoate metabolism' and 'butanoate metabolism' were enriched, suggesting increased SCFA production (figure 6). These SCFA-related pathways were enriched in PLHIV overall and, within dTBLs (figure 7A,B).

In addition, when comparing inferred pathways in the 5 oLs, a similar core of pathways was enriched in oL4. In contrast, versus oL4, oL1 was enriched in 'epithelial cell signalling in *Helicobacter pylori* infection', oL2 and oL5 were enriched in 'carbohydrate digestion and absorption', and oL3 was enriched in 'dioxin degradation' (online supplemental figure S9A–H). When comparing the three dLs, *Mycobacterium*-dominated oL3 was, compared with each other dLs, enriched in the similar core

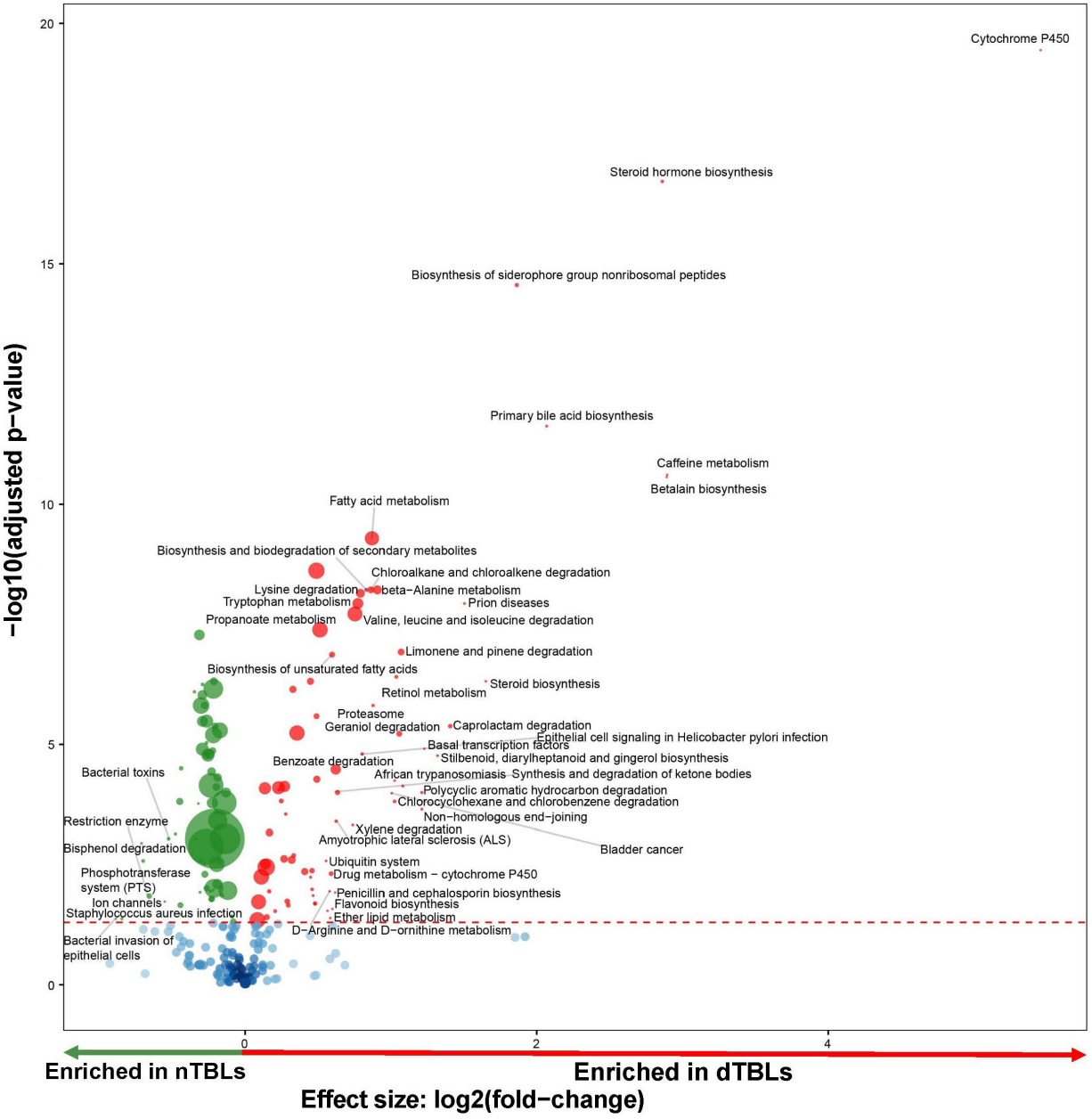



**Figure 6** Enriched microbial capacity for SCFA pathways in dTBLs vs nTBLs. Volcano plot depicting differentially abundant microbial pathways in dTBLs vs nTBLs inferred by PICRUSt2. key pathways of interest are bolded including aminobenzoate degradation, benzoate degradation and propanoate degradation. Significantly more discriminatory pathways appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0.05) as significance increases. Relative gene abundance is indicated by circle size. dTBLs, definite-TBLs; FDR, false discovery rate; nTBLs, non-TBL; SCFA, short-chain fatty acids.

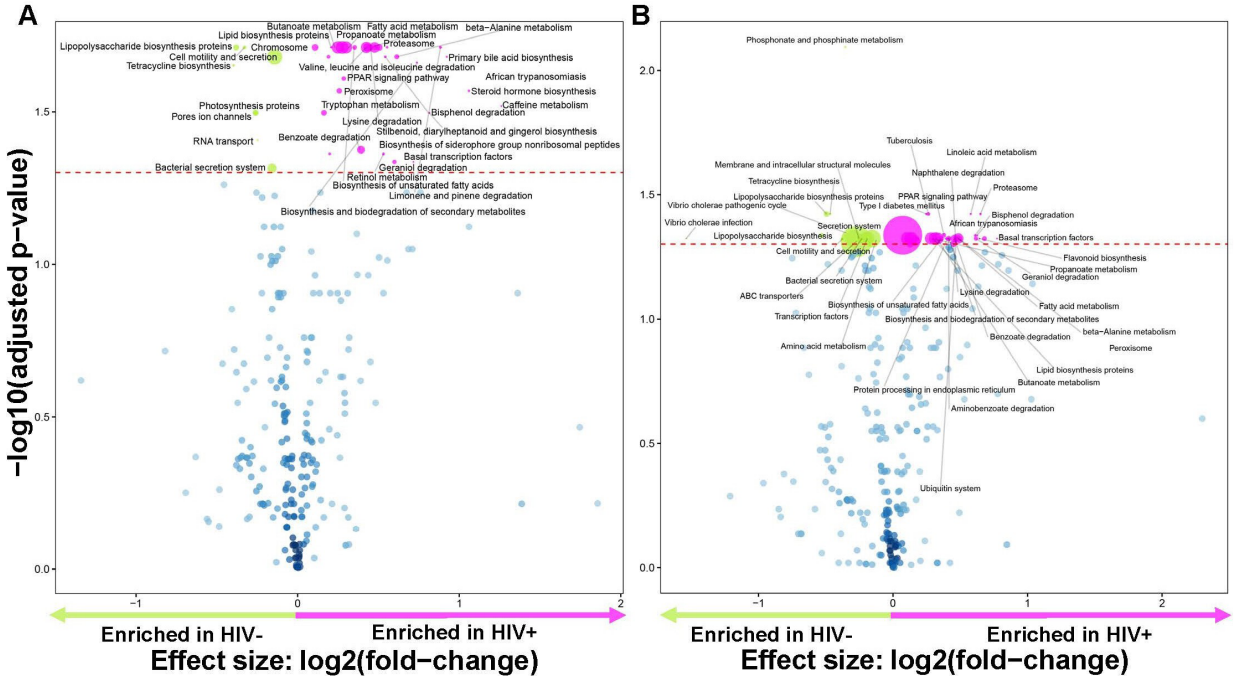



**Figure 7** Predicted metagenome function reveals increased capacity for SCFA production in HIV-positive versus HIV-negative patients overall, and in dTBLs. Volcano plot depicting functional pathways differing between (A) HIV-positive and HIV-negative patients with presumptive TBL and (B) in dTBLs. Key pathways of interest include butanoate metabolism, propanoate metabolism and benzoate degradation. Significantly more discriminatory pathways appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0.05). Relative pathway abundance is indicated by circle size. dTBLs, definite-TBL; FDR, false discovery rate; SCFA, short-chain fatty acids.

pathways as the *Mycobacterium*-dominated oL4 in all patients (figure 8C; online supplemental figure S10). These results show that pathways involved in fatty acid-related, amino acid-related and SCFA-related inferred microbial pathways were significantly enriched in dTBLs and *Mycobacterium* lymphotypes (oL4 and dL3).

## DISCUSSION

We characterised the local microbial environment in patients with lymphadenitis undergoing investigation for TB in an HIV-endemic setting. Our key findings are: (1) lymphatic microbial communities in dTBLs clustered into three distinct 'lymphotypes' we termed '*Prevotella-Corynebacterium*', '*Prevotella-Streptococcus*' and '*Mycobacterium*', (2) the *Mycobacterium* dTBL lymphotype was associated with HIV-positivity and other clinical features characteristic of severe lymphadenitis and (3) dTBLs relative to nTBLs were functionally enriched in fatty acid-related, amino acid-related and SCFA-related microbial metabolic pathways with known immunomodulatory effects (the *Mycobacterium* lymphotype was most enriched in these pathways than other dTBL lymphotypes). Finally, (4) dTBLs without *Mycobacterium* reads and nTBLs with *Mycobacterium* reads were identified. These data show TBL at the site-of-disease is not microbially homogenous and that distinct clusters of microbial communities exist associated with different clinical characteristics. The long-term significance and importance of these lymphotypes requires prospective evaluation.

We identified three lymphotypes within dTBLs termed '*Prevotella-Corynebacterium*', '*Prevotella-Streptococcus*' and '*Mycobacterium*', distinguished by different relative abundances of these taxa (*Prevotella* co-occurred in the first two lymphotypes). These individual taxa are enriched in respiratory secretions from pulmonary TB cases.[31 32] Furthermore, within dTBLs, *Streptococcus* is associated with low BMI and extent of lung damage.[32] *Prevotella* in BALF also positively correlates with SCFA concentrations and independently predicts incident TB in people without co-prevalent TB.[13] Compared with the other dTBL lymphotypes, '*Mycobacterium*' was associated with severe disease and most frequently occurred in PLHIV, agreeing with diagnostics studies that show stronger baseline mycobacterial PCR test readouts predict long term clinical outcomes in pulmonary[33] and extrapulmonary TB.[34] Together, these data show distinct lymphotypes are associated with different clinical characteristics and suggests that patients with the most severe *Mycobacterium*-dominated lymphotype may initially progress through different site-of-disease microbial states characterised by *Corynebacterium*, *Streptococcus* and/or *Prevotella* domination. Studies with longitudinal follow-up and repeat sampling are required to examine whether these lymphotypes have potential for clinical staging.

Importantly, *Corynebacterium* and *Streptococcus* often dominated in dTBL patients. Members of both taxa are causative agents of lymphadenitis and, even though these patients have TBL confirmed via conventional diagnostics, *Corynebacterium* and *Streptococcus* may therefore cocontribute to pathology and symptoms.[35–37] Coincidently, these taxa fall within the antimicrobial spectrum of first-line TB treatment,[16] meaning that this regimen may, in part, cure lymphadenitis by killing *Corynebacterium* and *Streptococcus* in addition to *Mycobacterium*.

Microbial pathways predicted to be most enriched in dTBLs involved fatty acid, amino acid and SCFAs (benzoate, propanoate) metabolism; all of which are associated with pulmonary TB disease compared with sick patients without TB.[38 39] SCFAs in particular suppress immune pathways involved in IFN-γ and IL-17A production and, ex vivo, limit macrophage-mediated kill of *Mtb*. SCFA concentrations hence predict incident TB in

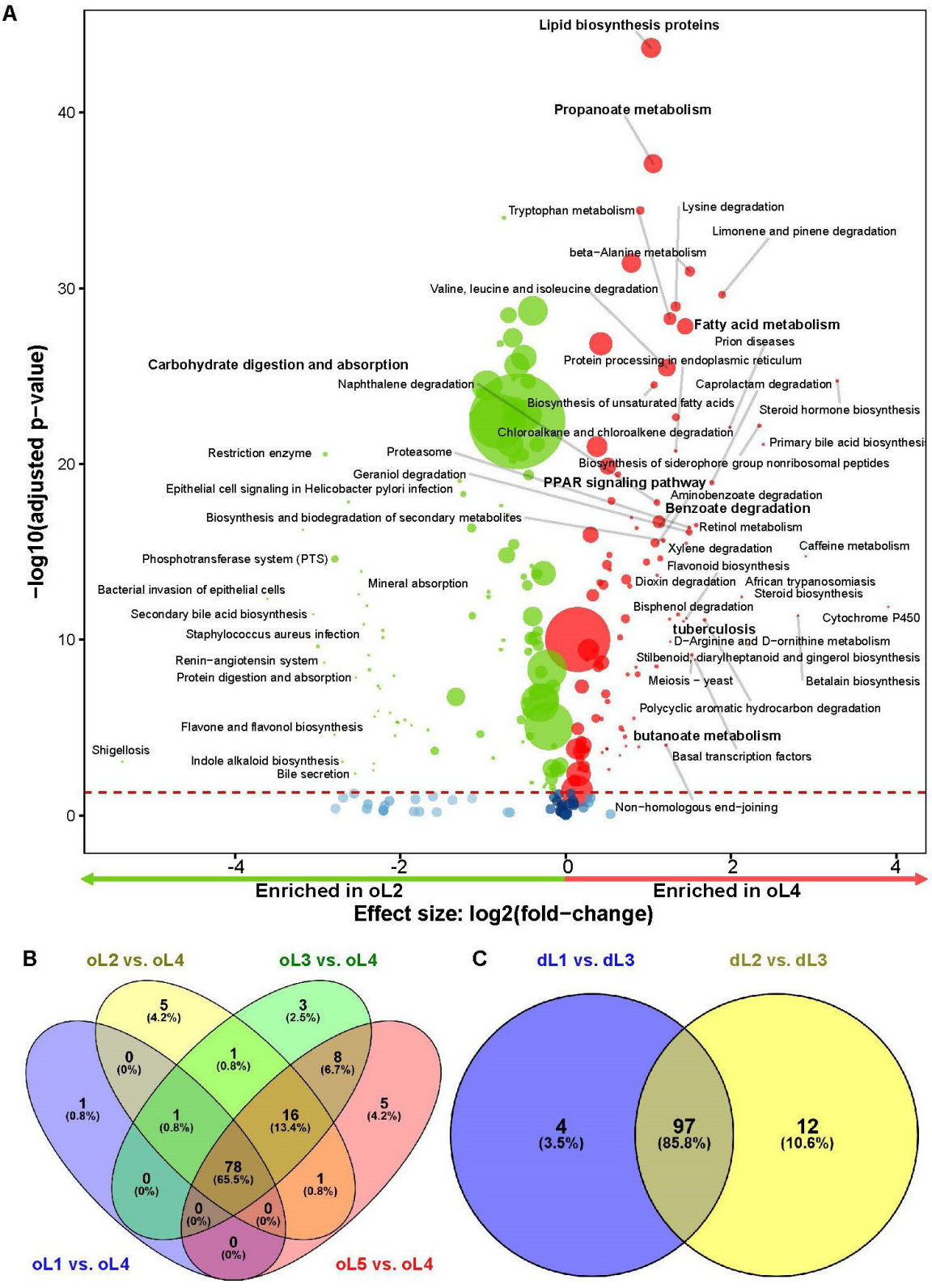



**Figure 8** Differential microbial pathways between lymphotypes showing similar core pathways enriched in the *Mycobacterium*-dominated lymphotype. (A) Volcano plot showing differentially abundant microbial pathways inferred by PICRUSt2 in oL2 vs oL4 representing pathways enriched in oL4 compared with every other oL in all patients (overall including dTBLs and nTBLs). Significantly more discriminatory pathways appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0.05) as significance increases. Relative pathway abundance is indicated by circle size. (B) 65.5% of all inferred pathways enriched in oL4 compared with each other oLS were common, while (C) 85.8% were common in dL3 compared with each other dTBL lymphotypes.Differentially enriched pathways common in all comparisons with the *Mycobacterium* dominant lymphotype included pathways involving lipid biosynthesis, fatty acids and SCFA metabolism, that is, lipid biosynthesis proteins, propanoate metabolism, benzoate degradation, and valine, leucine and isoleucine degradation. dL, dTBL lymphotype; dTBLs: definite-TBLs; FDR, false discovery rate; nTBLs: non-TBL; oL: overall lymphotype; SCFA: short-chain fatty acid.

patients.[13] Our research therefore suggests that the inflammation associated with lymphadenopathy is in part caused by the presence of microbes including but not limited to *Mycobacterium* that are able to produce SCFAs that interfere with these immunological pathways; revealing potentially new therapeutic targets to reduce lymphadenopathy.

We detected *Mtb* DNA in nTBLs. These reads could be from subclinical infection, previous TB exposure or disease, where the DNA was transported to the lymph node. *Mtb* DNA has been found in the lymph nodes of healthy individuals and primates exposed to TB, where the sites are hypothesised to serve as a *Mtb* growth and persistence niche.[6] dTBLs without *Mycobacterium* reads were also documented, however, 16S rRNA sequencing has known suboptimal sensitivity for *Mycobacterium*, in part due to low 16s RNA gene copy number.[40]

Our study has strengths and limitations. Patients were sampled once, as close as possible to treatment initiation; animal models might permit repeat invasive sampling especially if treatment is withheld. The programmatic context enabled large numbers of patients to be recruited, however, detailed long-term follow-up, which could include imaging of lymph nodes and more detailed measurements of differential responses to treatment, was not possible. We did not perform any viability tests, and since 16S gene sequencing is DNA based, the DNA may have originated from live, dead or nonculturable bacteria. Future studies could use meta-transcriptomics or culturomics to investigate this. We also used an FDR-adjusted p value threshold of 0.2 to identify differentially abundant taxa because this study was designed to be hypothesis generating and lower thresholds did not generate such taxa. Furthermore, the use of PICRUSt to infer potential function from 16S rRNA gene sequencing is a limitation. Follow-up studies using shotgun metagenomics, are necessary for inferring biological function and can more comprehensively describe the microbiota beyond bacteria. Our study was designed to describe the site-of-disease microbiome in TBL in a setting with a high burden of TB and HIV. Further research in different settings and populations is needed to validate our findings, especially those findings pertaining to microbial community clustering and the relationship between individual clusters and clinical characteristics.

In conclusion, we show dTBL patients have a distinct microbiome at the site of disease, characterised by three lymphotypes (*Mycobacterium*, *Prevotella-Corynebacterium*, *Prevotella-Streptococcus*). This dysbiosis of the lymphatic microbiome likely contributes to pathophysiology, including inflammatory state and clinical severity, which itself may reflect the chronicity of TB disease. TBL does therefore not appear to be a microbially homogenesis disease, and this reveals potentially new diagnosis, therapeutic and prognostic targets.

**Author affiliations**
[1]DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, Western Cape, South Africa
[2]African Microbiome Institute, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, Western Cape, South Africa
[3]Division of Pulmonary and Critical Care Medicine, New York University Grossman School of Medicine, NYU Langone Health, New York, NY, USA
[4]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[5]National Health Laboratory Service, Tygerberg Hospital, Cape Town, Western Cape, South Africa
[6]Division of Medical Microbiology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, Western Cape, South Africa
[7]Division Anatomical Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, Western Cape, South Africa

**Acknowledgements** The authors thank study participants and Tygerberg Hospital FNA clinic staff especially Sr Cupido. Additionally, we thank CLIME research group staff, especially Sr Ruth Wilson and Roxanne Higgit. GRN acknowledges funding from L'Oréal-UNESCO For Women in Science Sub-Saharan Africa Young Talents Award, and the International Rising Talents Award. The content is the sole responsibility of the authors and does not necessarily represent the official views of the funders. Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: hpc.uct.ac.za.

**Contributors** GRN, CCN and GT contributed to conceptualisation and design of the study and supervised the study, funding acquisition, data collection, and wrote the manuscript. GRN, CCN, IS, BW, IS, JCC, LS and GT contributed to data and statistical analysis, and figures. All authors contributed to interpretation of data and editing of the manuscript. As the study guarantor, GT is responsible for the overall content of this manuscript.

**Funding** This work and authors were supported by the European & Developing Countries Clinical Trials Partnership (EDCTP; project numbers SF1041, TMA2017CDF-1914-MOSAIC and TMA2019CDF-2738-ESKAPE-TB), National Research Foundation (NRF), the South African Medical Research Council (SAMRC), the Harry Crossley Foundation and Stellenbosch University Faculty of Health Sciences, and the National Institutes of Health under award numbers (R01AI136894; U01AI152087; U54EB027049; D43TW010350; K43TW012302).

**Disclaimer** The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Competing interests** None declared.

**Patient consent for publication** Not applicable.

**Ethics approval** This study involves human participants and was approved by Health Research and Ethical Committee of Stellenbosch University (N16/04/050), Tygerberg Hospital (Project ID:4134),Western Cape Department of Health (WC_2016RP15_762). Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available on reasonable request. Data are available on reasonable request. Deidentified patient data, the study protocol, informed consent and datasets generated in this study may be requested from the corresponding author. Sequencing data reported in this paper will be released and made publicly available for sharing on publication (NCBI Bioproject PRJNA738676).



**ORCID iD**
Grant Theron http://orcid.org/0000-0002-9216-2415

## REFERENCES

1 WHO. Global tuberculosis report 2020. Geneva, Switzerland. World Health Organization; 2020.
2 Gupta RK, Lawn SD, Bekker L-G, *et al*. Impact of human immunodeficiency virus and CD4 count on tuberculosis diagnosis: analysis of city-wide data from Cape town, South Africa. *Int J Tuberc Lung Dis* 2013;17:1014–22.
3 Mohapatra PR, Janmeja AK. Tuberculous lymphadenitis. *J Assoc Physicians India* 2009;57:585–90.
4 Mekonnen D, Derbie A, Abeje A, *et al*. Epidemiology of tuberculous lymphadenitis in Africa: a systematic review and meta-analysis. *PLoS One* 2019;14:e0215647.
5 Maji A, Misra R, Kumar Mondal A, Mondal AK, *et al*. Expression profiling of lymph nodes in tuberculosis patients reveal inflammatory milieu at site of infection. *Sci Rep* 2015;5:1–10.
6 Ganchua SKC, White AG, Klein EC, *et al*. Lymph nodes-The neglected battlefield in tuberculosis. *PLoS Pathog* 2020;16:e1008632.
7 Martin CJ, Cadena AM, Leung VW, *et al*. Digitally Barcoding *Mycobacterium tuberculosis* Reveals *In Vivo* Infection Dynamics in the Macaque Model of Tuberculosis. *mBio* 2017;8. doi:10.1128/mBio.00312-17. [Epub ahead of print: 09 May 2017].

8 Yik Wang C, Wang CY. An experimental study of latent tuberculosis. *The Lancet* 1916;188:417–9.
9 Ganchua SKC, Cadena AM, Maiello P, *et al*. Lymph nodes are sites of prolonged bacterial persistence during Mycobacterium tuberculosis infection in macaques. *PLoS Pathog* 2018;14:e1007337.
10 Behr MA, Waters WR. Is tuberculosis a lymphatic disease with a pulmonary portal? *Lancet Infect Dis* 2014;14:250–5.
11 Esmail H, Lai RP, Lesosky M, *et al*. Characterization of progressive HIV-associated tuberculosis using 2-deoxy-2-[$^{18}$F]fluoro-D-glucose positron emission and computed tomography. *Nat Med* 2016;22:1090–3.
12 Blacher E, Levy M, Tatirovsky E, *et al*. Microbiome-modulated metabolites at the interface of host immunity. *J Immunol* 2017;198:572–80.
13 Segal LN, Clemente JC, Li Y, *et al*. Anaerobic bacterial fermentation products increase tuberculosis risk in Antiretroviral-Drug-Treated HIV patients. *Cell Host Microbe* 2017;21:530–7.
14 Kiernan MG, Coffey JC, McDermott K, *et al*. The human mesenteric lymph node microbiome differentiates between Crohn's disease and ulcerative colitis. *Journal of Crohn's and Colitis* 2019;13:58–66.
15 O'Brien CL, Pavli P, Gordon DM, *et al*. Detection of bacterial DNA in lymph nodes of Crohn's disease patients using high throughput sequencing. *Gut* 2014;63:1596–606.
16 Naidoo CC, Nyawo GR, Wu BG, *et al*. The microbiome and tuberculosis: state of the art, potential applications, and defining the clinical research agenda. *Lancet Respir Med* 2019;7:892–906.
17 Vázquez-Pérez JA, Carrillo CO, Iñiguez-García MA, *et al*. Alveolar microbiota profile in patients with human pulmonary tuberculosis and interstitial pneumonia. *Microb Pathog* 2020;139:103851.
18 Zhou Y, Lin F, Cui Z, *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.
19 Minnies S, Reeve BWP, Rockman L, *et al*. Xpert MTB/RIF ultra is highly sensitive for the diagnosis of tuberculosis lymphadenitis in a High-HIV setting. *J Clin Microbiol* 2021;59:e01316–21.
20 Bolyen E, Rideout JR, Dillon MR, *et al*. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;37:852–7.
21 Callahan BJ, McMurdie PJ, Rosen MJ, *et al*. DADA2: high-resolution sample inference from illumina amplicon data. *Nat Methods* 2016;13:581–3.
22 McDonald D, Price MN, Goodrich J, *et al*. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *Isme J* 2012;6:610–8.
23 McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;8:e61217.
24 Dixon P. Vegan, a package of R functions for community ecology. *J Veg Sci* 2003;14:927–30.
25 Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* 2012;7:e30126.
26 Douglas GM, Maffei VJ, Zaneveld J. PICRUSt2: an improved and customizable approach for metagenome inference. *BioRxiv* 2020;672295.
27 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol* 2014;15:1–21.
28 Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 2003;19:368–75.
29 SSTS. *Stata: data analysis and statistical software. Stata statistical software: release 16*. College Station, TX, 2022.
30 Estaki M, Jiang L, Bokulich NA, *et al*. QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. *Curr Protoc Bioinformatics* 2020;70:e100.
31 Krishna P, Jain A, Bisen PS. Microbiome diversity in the sputum of patients with pulmonary tuberculosis. *Eur J Clin Microbiol Infect Dis* 2016;35:1205–10.
32 Ticlla MR, Hella J, Hiza H, *et al*. The sputum microbiome in pulmonary tuberculosis and its association with disease manifestations: a cross-sectional study. *Front Microbiol* 2021;12:633396.
33 Pires MdeM, Pereira GR, Barbosa MS, *et al*. Association of Xpert MTB/RIF cycle threshold values with tuberculosis treatment outcomes. *Lung* 2020;198:985–9.
34 Martyn EM, Bangdiwala AS, Kagimu E, *et al*. Cerebrospinal fluid bacillary load by Xpert MTB/RIF ultra polymerase chain reaction cycle threshold value predicts 2-week mortality in human immunodeficiency virus-associated tuberculous meningitis. *Clin Infect Dis* 2021;73:e3505–10.
35 Peel MM, Palmer GG, Stacpoole AM, *et al*. Human lymphadenitis due to Corynebacterium pseudotuberculosis: report of ten cases from Australia and review. *Clin Infect Dis* 1997;24:185–91.
36 Barroso LF, Pegram SP. Clinical manifestations, diagnosis, and treatment of diphtheria. In: *UpToDate*. Post TW ed. Waltham, MA: UpToDate, 2022.
37 Denis Spelman M, Baddour LM. Cellulitis and skin abscess: Epidemiology, microbiology, clinical manifestations, and diagnosis. In: *UpToDate*. Post TW ed. Waltham, MA: UpToDate, 2021.
38 Naidoo CC, Nyawo GR, Sulaiman I, *et al*. Anaerobe-enriched gut microbiota predicts pro-inflammatory responses in pulmonary tuberculosis. *EBioMedicine* 2021;67:103374.
39 Somboro AM, Diallo D, Holl JL, *et al*. The role of the microbiome in inflammation during tuberculosis. *EBioMedicine* 2021;68:103435.
40 Sulaiman I, Wu BG, Li Y, *et al*. Evaluation of the airway microbiome in nontuberculous mycobacteria disease. *Eur Respir J* 2018;52. doi:10.1183/13993003.00810-2018. [Epub ahead of print: 25 Oct 2018].

**Supplementary Material**





Nyawo GR, *et al. Thorax* 2022;0:1–12. doi: 10.1136/thorax-2022-219103

## Methods

*FNAB collection and TB microbiology*

Needle passes were done on the largest (using surface area recorded in $cm^2$) distinct node with a 23-gauge needle and a 10 mL syringe as described [1]. The first two passes were used to prepare standard microscope slides for cytological examination using Rapidiff and Papanicolaou staining. A flush of the needle was collected in 1.5 mL of TB transport medium [2] media and sent to the National Health Laboratory Services (NHLS) microbiology laboratory for Xpert MTB/RIF (Xpert) or Xpert MTB/RIF Ultra (Ultra), Mycobacteria Growth Indicator Tube 960 liquid culture (MGIT960; BD), and acid-fast bacilli (AFB) staining.

*Definitions*

We used a reference standard to designate patients as definite-TBLs (dTBLs), probable-TBLs (pTBLs), or non-TBLs (nTBLs) based on bacteriological or cytological evidence of TB as previously described [1]. dTBLs had at least one *Mycobacterium tuberculosis* complex (MTBC)-positive extrapulmonary or pulmonary specimen by acid-fast bacilli (AFB) staining microscopy, Xpert MTB/RIF (Xpert) and/or Xpert MTB/RIF Ultra (Ultra), or Mycobacteria Growth Indicator Tube (MGIT) 960 liquid culture (culture). pTBLs did not meet dTBL criteria but commenced treatment empirically. nTBLs had no microbiological TB, were not placed on treatment, and/or had an alternative diagnosis.

*Clustering*

We then evaluated for presence of distinct groups of samples based on identification of distinct microbial communities in lymph nodes which we called lymphotypes. Dirichlet multinomial mixture modelling (DMM) was performed using the R package *DirichletMultinomial* to establish clustering within groups [3]. Using genus tables, the number of clusters was determined by selecting the number of Dirichlet components that reduced the Laplace approximation of the model [3] (i.e. lower values indicate better fits). Clustering profiles indicate unique groupings, interpreted as “lymphotypes”.

## Results

*Environmental and background controls*

It is important to evaluate possible sources of microbial DNA contamination in low biomass samples such lymph fluid. Given the concern for potential carry-over of skin commensals and DNA present on medical apparatus (i.e., syringe used for biopsy), microbiome readouts from patients' lymph fluid were analysed in parallel with that from skin (sampled at puncture site) and saline flush of the syringe used for aspiration (done for 1 in 5 participants). Pairwise comparisons of α-diversity were similar between saline and skin, and lymph fluid and saline (**Figure S1A**). β-diversity was different between the three fluid types (*p=0.001*; **Figure S1B**), with lymph enriched in the respiratory pathogen *Mycobacterium* (**Figure S1C**) vs. skin, and vs. saline (**Figure S1D**). Skin was enriched with *Psychrobacter* and *Corynebacterium* vs. lymph and saline, respectively (**Figures S1C and S1E**), whilst no taxa were enriched in saline (**Figures S1D-E**).

*Cohort characteristics*

We had 89 dTBLs, 61 nTBLs (**Figure 1**) and 8 pTBLs (latter subsequently excluded due to small numbers), from which we collected FNABs from the head, neck or thorax regions. dTBLs were more likely to have supraclavicular or head lymph node involvement than nTBLs, if HIV-positive were more likely to have a lower CD4 count (**Table 1**) and were more likely to have a FNAB that appeared bloody rather than chylous.

*Microbial comparisons including probable TBL patients*

We compared the microbiome of pTBLs (n=9) to dTBLs and nTBLs. There were no differences in α-diversity (**Figure S2A**). β-Diversity differed between pTBLs and dTBLs, but not between pTBLs and nTBLs (**Figure S2B**). We excluded the pTBLs from the primary analysis due to few patients meeting this definition.

*α- and β-diversities according to demographic, clinical, and microbiological characteristics **(Table S2)***

Overall: Females had a higher α-diversity than males (*p=0.016*), patients who used antibiotics within a year, and at recruitment had a lower α-diversity than those who did not (*p=0.042; p=0.003 respectively*), and patients with smaller lymph nodes had a higher α-diversity than those with larger nodes (*p=0.001*). β–diversity was different in patients with antibiotic use at recruitment versus none (*p=0.032*) and antibiotic use within one year versus later use (*p=0.020*). Furthermore, within PLHIV, β–diversity differed by ART status (*p=0.042*) and CD4 count stratum (*p=0.038*). In those patients who received non-TB antibiotics (i.e., current or within one year of enrolment), α-diversities by TB status were similar but β-diversity differed (*p=0.001*). In patients without antibiotics (n=109), α-diversity was lower in dTBLs vs. nTBLs (*p=0.035*) and β-diversity differed (*p=0.035*).

dTBLs: α-diversity was decreased with antibiotic use at recruitment (*p=0.025*) and within one year (*p=0.007*) as well as in larger nodes lymph node size (*p=0.034*). β-diversity also differed by antibiotics usage (current and within one year) and CD4 count stratum in PLHIV (*p=0.034*).

nTBLs: α-diversity was less in males than females (*p=0.003*) and in smokers than non-smokers (*p=0.002*). β-diversity was only associated with specimen appearance in nTBLs (*p=0.047*).

*Correlation between 16S rRNA gene sequencing and TB diagnostic tests*

*Mycobacterium* relative abundance in dTBLs showed a positive correlation with bacillary load (based on Xpert and Ultra cycle threshold values; $r_s$=-0.774, 95% CI [-0.777, -0.514], *p<0.0001*; **Figure S4A**), and culture days-to-positivity; $r_s$=-0.684, 95% CI [-0.859, -0.367.], *p=0.003*; **Figure S4B**). However, there was no correlation between lymph node size and mycobacterial load (relative abundance) (**Figure S4C**).

*Differences by HIV status*

Overall: α-diversity did not differ by HIV status (**Figure 3A**) and although β-diversity did (**Figure 3B**) no differentially enriched taxa were found, however, the relative abundance of *Mycobacterium* was higher in PLHIV (**Figure S5C**).

Comparisons within dTBLs or nTBLs by HIV status: There were 55% (49/89) and 39% (23/59) HIV-positive dTBLs and nTBLs, respectively. Within dTBLs or nTBLs, α-diversities did not differ by HIV status (**Figure 3A**) and β-diversity differed by HIV status within dTBLs (*p=0.017,* **Figure 3C**) but not within nTBLs. HIV-positive dTBLs had higher *Mycobacterium* relative abundance than HIV-negative dTBLs **(Figure S5C)**.

Comparisons within HIV-positives or -negatives by TB status: In people of the same HIV status, α-diversity did not differ by TB status (**Figure 3A**) and β-diversity only differed between dTBLs vs. nTBLs in HIV-positives (*p=0.009*, **Figure 3D**) where dTBLs were enriched in *Mycobacterium* (**Figure S5D**). In HIV-negatives, there were no differences between dTBLs and nTBLs (**Figure S5B**).

*Lymphotype identification and their associations with clinical characteristics*

We further explored this data using DMM to identify potential clusters in the TBL microbiome. These clusters were termed “lymphotypes”. We then looked for associations between each lymphotype(s) and patients’ clinical characteristics.

Overall: We first examined whether all patients could be grouped into distinct lymphotypes; these were termed overall lymphotypes (oLs). Five oLs with differing α- and β-diversities were identified (**Figure 4A-C, Table S3**). oL1 had no dominant taxa (**Figure 4D**), whilst oL4 was *Mycobacterium*-dominated and had the least α-diversity, and oL2, oL3 and oL5 were *Corynebacterium*-, *Prevotella*- and *Streptococcus*-dominated, respectively. While no taxa were differentially abundant in oL1 vs. other oLs (**Figure S6A-C**), oL2, oL3, and oL5 were enriched relative to oL4 in *Corynebacterium*, *Prevotella*, and *Streptococcus,* respectively (**Figure 4E-G**). The proportions of dTBLs in oL1, oL2, oL3, oL4, and oL5 were 35% (17/48), 63% (28/44), 57% (12/21), 100% (21/21), and 69% (11/16), respectively (**Table S4**). The patients in these lymphotypes are associated with distinct clinical characteristics. The majority of nTBLs occurred in highly diverse oLs with a heterogenous mixtures of taxa; likely reflecting the spectrum of pathologies in people with TBL ruled out. oL1 was associated with characteristics indicative of less severe lymphadenitis. Compared separately to oL2, oL4, and oL5, oL1s were less likely to have dTBL. Furthermore, oL1s were less likely to be HIV-positive vs. oL4s but, oL1 PLHIVs had lower CD4 counts vs. oL2 and oL3 PLHIVs. In contrast, oL4 was associated with characteristics resembling more severe lymphadenitis. oL4 was more likely to contain dTBL patients than each other oL. Furthermore, compared with oL2s, oL4s were more likely to have a bigger lymph node, chylous FNABs and, of PLHIV, a smaller proportion on ART. Compared with oL3s, oL4s were more likely to have previous TB and HIV, and those with HIV were more likely to have lower CD4 counts. Compared with oL5s, oL3s with HIV had lower CD4 counts. Therefore, in summary, oL1 appears to be associated with less severe lymphadenitis forms, whereas oL4 was associated more severe forms (**Table S4**).



Nyawo GR, *et al. Thorax* 2022;0:1–12. doi: 10.1136/thorax-2022-219103

Within patients of the same TB status: We then examined whether patients within each TB group could be grouped into distinct lymphotypes. Within dTBLs, three lymphotypes (termed definite TBL lymphotypes; dL) with differing β-diversities were identified (**Figure 5A-B**). dL1 was abundant in *Prevotella* and *Corynebacterium*, dL2 in *Prevotella* and *Streptococcus*, and dL3 in *Mycobacterium* (**Figure 5C**) and these taxa were differentially abundant (**Figure 5D-F**). These lymphotypes were termed *Prevotella-Corynebacterium*, *Prevotella-Streptococcus* and *Mycobacterium,* respectively. dL3s were more likely to be HIV-positive, with larger lymph nodes, compared with dL1s. In addition, dL3s were more likely to have larger lymph nodes than dL2s. Lastly, dL2s are more likely to be female than dL1s (**Table S5**). Together, these differences suggest dL3 is associated with more severe TBL than other lymphotypes. Within nTBLs, no lymphotypes were identified (**Figure S7**).

*Predictive metagenome profiling shows increased short chain fatty acid metabolism*

We further predicted the bacterial metagenome content and made functional inferences of the microbiome using the PICRUSt algorithm. The overall differences among pathways between groups were evaluated and visualised by DESeq2 analysis.

dTBLs vs. nTBLs: 139 inferred microbial metabolic pathways were differentially enriched (75 in dTBLs, 64 in nTBLs). In dTBLs, “fatty acid metabolism”, “benzoate degradation”, “propanoate metabolism” and “butanoate metabolism” were enriched, suggesting increased SCFA production (**Figure 6**).

HIV-positive vs. negatives: The above SCFA-related pathways were enriched in HIV-positive vs. -negative patients overall and, within dTBLs, in HIV-positives vs. -negatives (**Figure 7A-B**). Within nTBLs, HIV-positives were enriched in the “cell cycle – *Caulobacter*”, “bacterial secretion system” and “oxidative phosphorylation” vs. -negatives (**Figure S8**).

In different lymphotypes: When comparing inferred pathways in oLs, a similar core of pathways was enriched in oL4. These included the “propanoate metabolism”, “tuberculosis”, “lipid biosynthesis”, “butanoate metabolism”, “fatty acid metabolism” and “PPAR signalling pathway” (most-to-least enriched) (**Figure 8A-B)**. In contrast, vs. oL4, oL1 was enriched in “epithelial cell signalling in *Helicobacter pylori* infection”, oL2 was enriched in “carbohydrate digestion and absorption”, oL3 was enriched in “dioxin degradation”, and oL5 was enriched in “carbohydrate digestion and absorption” (**Figure S9A-H)**. When comparing the three dTBL lymphotypes, *Mycobacterium*-dominated dL3 was, compared with each other dLs, enriched in the similar core pathways seen for the *Mycobacterium*-dominated oL4 in all patients (**Figure 8C; Figure S10**).

**Table S1: Reference standard definition used in the study.** Due a small number of pTBLs, they were excluded from analyses.

| | dTBLs | nTBLs | pTBLs |
|---|---|---|---|
| | **Site-of-disease fluid** | | |
| Xpert | ✓ | × | × |
| Ultra | ✓ | × | × |
| MGIT960 Culture | ✓ | × | × |
| Smear microscopy | ✓ | × | × |
| Cytology | ✓ | × | × |
| | **Non-site-of-disease fluid** | | |
| Smear microscopy | ✓ | × | × |
| Xpert | ✓ | × | × |
| Ultra | ✓ | × | × |
| MGIT960 | ✓ | × | × |
| | **Treatment information** | | |
| TB treatment initiated | × | × | ✓ |
| Response to treatment self-reported by patient | × | × | ✓ |

Abbreviations: dTBLs: definite tuberculous lymphadenitis; nTBLs: non-tuberculous lymphadenitis; pTBLs: probable-tuberculous lymphadenitis; Xpert: Xpert MTB/RIF; Ultra: Xpert MTB/RIF Ultra; MGIT960 Culture: Mycobacteria Growth Indicator Tube 960 liquid culture.

**Table S2: α- and β-diversities in presumptive TBL patients when patients with different demographic and clinical characteristics were compared.** Several characteristics, described in the Supplementary Results text, were associated with differing diversities.

204

| Characteristics | Overall (n=150) | | | dTBLs (n=89) | | | nTBLs (n=61) | | |
|---|---|---|---|---|---|---|---|---|---|
| | α-diversity *p-value* (Shannon's Index) | β-diversity | | α-diversity *p-value* (Shannon's Index | β-diversity | | α-diversity *p-value* (Shannon's Index | β-diversity | |
| | | *p-value* (PERMANOVA) | $R^2$ *value* | | *p-value* (PERMANOVA) | $R^2$ *value* | | *p-value* (PERMANOVA) | $R^2$ *value* |
| dTBL | 0.110 | **0.001** | 0.037 | - | - | - | - | - | - |
| Sex | **0.016** | 0.121 | 0.010 | 0.406 | 0.616 | 0.035 | **0.003** | **0.012** | 0.008 |
| HIV | 0.860 | **0.004** | 0.023 | 0.179 | **0.008** | 0.043 | 0.312 | 0.731 | 0.432 |
| *CD4+ <200 cells/µl* | 0.459 | **0.038** | 0.032 | 0.053 | **0.034** | 0.055 | 0.140 | 0.455 | 0.045 |
| *On ART* | 0.662 | **0.042** | 0.030 | 0.267 | 0.344 | 0.022 | 0.306 | 0.267 | 0.055 |
| Previous TB | 0.337 | 0.072 | 0.012 | 0.426 | 0.141 | 0.018 | 0.501 | 0.603 | 0.015 |
| Tobacco smoking | 0.084 | 0.189 | 0.009 | 0.636 | 0.658 | 0.008 | **0.002** | 0.276 | 0.020 |
| Antibiotic use within 1 year of recruitment | **0.042** | **0.020** | 0.015 | **0.025** | **0.012** | 0.036 | 0.547 | 0.212 | 0.022 |
| Antibiotic use at recruitment | **0.003** | **0.032** | 0.061 | **0.007** | **0.025** | 0.141 | 0.062 | 0.064 | 0.115 |
| Site (neck vs. thorax) | 0.220 | 0.134 | 0.010 | 0.128 | 0.142 | 0.018 | 0.809 | 0.830 | 0.011 |
| Specimen appearance (bloody vs. chylous) | 0.213 | 0.068 | 0.012 | 0.771 | 0.198 | 0.016 | **0.020** | **0.047** | 0.778 |
| Lymph node characteristics: size, $cm^2$ | **0.011** | 0.128 | 0.012 | **0.034** | 0.065 | 0.265 | 0.197 | 0.612 | 0.017 |

*$R^2$ provides the proportion of variation explained (e.g., a factor that has a $R^2 = 0.037$, explains 3.7% of the variation in community composition) by β-diversity.

Abbreviations: TB: tuberculosis; TBL: tuberculous lymphadenitis; ART: antiretroviral therapy; dTBLs: definite tuberculous lymphadenitis; nTBLs: non-tuberculous lymphadenitis

**Table S3: Adjusted p-values for α-diversity comparisons between lymphotypes measured by Shannon's diversity index.**

| Comparison | Lymphotype with highest α-diversity | Adjusted *p-value* |
|---|---|---|
| **Lymphotype comparisons in overall lymphotypes** | | |
| oL1 vs. oL2 | oL2 | **<0.0001** |
| oL1 vs. oL3 | oL3 | **0.0012** |
| oL1 vs. oL4 | oL1 | >0.9999 |
| oL1 vs. oL5 | oL5 | **<0.0001** |
| oL2 vs. oL3 | oL2 | >0.9999 |
| oL2 vs. oL4 | oL2 | **<0.0001** |
| oL2 vs. oL5 | oL5 | **0.0329** |
| oL3 vs. oL4 | oL3 | **<0.0001** |
| oL3 vs. oL5 | oL5 | 0.1088 |
| oL4 vs. oL5 | oL5 | **<0.0001** |
| **Lymphotype comparisons in all dTBL lymphotypess** | | |
| dL1 vs. dL2 | oL2 | **0.001** |
| dL1 vs. oL3 | oL3 | **<0.0001** |
| dL2 vs. oL3 | oL3 | **0.001** |

Definition of abbreviations: oL: overall lymphotype;; dL: dTBL lymphotype.

205
206

**Table S4: Demographic, clinical, and microbiological differences in each lymphotype (overall in all patients) showing oL1 is likely associated with less severe forms of lymphadenitis whereas oL4 is associated with more severe forms.** Amongst other differences, oL1s were less likely to have dTBL than oL2s, oL4s, and oL5s. Furthermore, oL1s were less likely to be HIV-positive vs. oL4s. oL1 PLHIV had lower CD4 counts vs. oL2 and oL3 PLHIVs. In contrast, oL4s were more likely to be dTBLs than other lymphotypes. Furthermore, compared to oL2, oL4s had bigger lymph nodes and were more likely to have chylous FNABs and a smaller proportion of PLHIVs on ART. Compared to oL3, oL4s were more likely to have previous TB and HIV, and oL3 PLHIVs were more likely to have lower CD4 counts. Compared to oL5, oL3 PLHIV had lower CD4 counts.

| Characteristic[¶] | Total (n=150) | L1 (n=48) (No dominant taxa) | L2 (n=44) *Corynebacterium* ) | L3 (n=21) (*Prevotella*) | L4 (n=21) (*Mycobacterium*) | L5 (n=16) (*Streptococcus*) | *p-value* (L1 vs. L2) | *p-value* (L1 vs. L3) | *p-value* (L1 vs. L4) | *p-value* (L1 vs. L5) | *p-value* (L2 vs. L3) | *p-value* (L2 vs. L4) | *p-value* (L2 vs. L5) | *p-value* (L3 vs. L4) | *p-value* (L3 vs. L5) | *p-value* (L4 vs. L5) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age, years | 36 (30-45) | 35 (29-47) | 37 (32-47) | 31 (28-46) | 37 (34-43) | 36 (28-45) | >0.999 | >0.999 | >0.999 | >0.999 | >0.999 | >0.999 | >0.999 | >0.999 | >0.999 | >0.999 |
| dTBLs | 89/150 (59) | 17/48 (35) | 28/44 (64) | 12/21 (57) | 21/21 (100) | 11/16 (69) | **0.007** | 0.093 | **<0.001** | **0.020** | 0.615 | **0.001** | 0.713 | **0.001** | 0.471 | **0.006** |
| Female | 83/150 (55) | 25/48 (52) | 26/44 (59) | 8/21 (38) | 12/21 (57) | 12/16 (75) | 0.499 | 0.284 | 0.698 | 0.108 | 0.113 | 0.882 | 0.258 | 0.217 | **0.026** | 0.260 |
| HIV | 72/148 (49) | 20/48 (42) | 24/42 (57) | 6/21 (29) | 15/21 (71) | 7/16 (44) | 0.143 | 0.302 | **0.023** | 0.884 | **0.032** | 0.271 | 0.361 | **0.006** | 0.338 | 0.089 |
| *CD4+* | 166 (90-308) | 35 (29-48) | 171 (86-332) | 255 (154-387) | 83 (17-163) | 136 (54-334) | **<0.0001** | **0.001** | 0.733 | 0.103 | >0.999 | 0.620 | >0.999 | 0.223 | >0.999 | >0.999 |
| *CD4+ <200 cells/µl* | 43/72 (60) | 10/20 (50) | 14/24 (58) | 2/6 (33) | 12/15 (80) | 5/15 (33) | 0.580 | 0.473 | 0.069 | 0.324 | 0.272 | 0.163 | 0.129 | **0.040** | >0.999 | **0.010** |
| *On ART* | 35/71 (49) | 9/20 (45) | 16/24 (67) | 3/6 (50) | 5/15 (33) | 2/6 (33) | 0.149 | 0.829 | 0.486 | 0.612 | 0.449 | **0.042** | 0.136 | 0.477 | 0.558 | >0.999 |
| Previous TB | 33/148 (22) | 10/48 (21) | 9/42 (21) | 2/21 (10) | 9/21 (43) | 3/16 (19) | 0.945 | 0.254 | 0.060 | 0.858 | 0.241 | 0.076 | 0.822 | **0.014** | 0.416 | 0.121 |
| Tobacco smoking | 43/149 (29) | 18/48 (38) | 12/44 (27) | 8/21 (38) | 3/21 (14) | 2/15 (13) | 0.296 | 0.936 | 0.054 | 0.062 | 0.377 | 0.245 | 0.232 | 0.079 | 0.082 | 0.875 |
| Antibiotic use within 1 year of recruitment | 38/147 (26) | 11/47 (23) | 9/9 (100) | 4/20 (20) | 9/20 (45) | 5/16 (31) | **<0.001** | 0.760 | 0.077 | 0.533 | **<0.001** | **0.005** | **0.001** | 0.091 | 0.439 | 0.400 |
| *At recruitment* | 21/38 (55) | 8/11 (73) | 5/9 (56) | 1/4 (25) | 6/9 (67) | 1/5 (20) | 0.423 | 0.095 | 0.769 | **0.049** | 0.308 | 0.629 | 0.198 | 0.164 | 0.858 | 0.094 |
| Lymph node characteristics: sites | | | | | | | | | | | | | | | | |
| Neck | 133/150 (89) | 46/48 (96) | 37/44 (84) | 18/21 (86) | 20/21 (95) | 12/16 (75) | 0.058 | 0.136 | 0.911 | **0.013** | 0.865 | 0.201 | 0.421 | 0.293 | 0.410 | 0.074 |
| *Deep anterior cervical* | 60/133 (45) | 19/46 (41) | 16/37 (43) | 8/18 (44) | 13/20 (65) | 4/12 (33) | 0.859 | 0.819 | 0.077 | 0.615 | 0.933 | 0.117 | 0.544 | 0.203 | 0.543 | 0.082 |
| *Deep lateral cervical* | 25/133 (19) | 13/46 (28) | 8/37 (22) | 2/18 (11) | 2/20 (10) | 0/12 (0) | 0.489 | 0.145 | 0.104 | **0.037** | 0.343 | 0.271 | 0.078 | 0.911 | 0.232 | 0.258 |
| *Superficial* | 15/133 (11) | 8/46 (17) | 2/37 (5) | 3/18 (17) | 2/20 (10) | 0/12 (0) | 0.095 | 0.945 | 0.442 | 0.120 | 0.173 | 0.517 | 0.411 | 0.544 | 0.136 | 0.258 |
| *Supraclavicular* | 20/133 (15) | 2/46 (4) | 7/37 (19) | 3/18 (17) | 3/20 (15) | 5/12 (42) | **0.034** | 0.099 | 0.133 | **<0.001** | 0.839 | 0.710 | 0.111 | 0.888 | 0.129 | 0.092 |
| *Head* | 13/133 (10) | 4/46 (9) | 4/37 (11) | 2/18 (11) | 0/20 (0) | 3/12 (25) | 0.746 | 0.766 | 0.174 | 0.123 | 0.973 | 0.127 | 0.222 | 0.126 | 0.317 | **0.019** |
| Thorax | 17/150 (11) | 2/48 (4) | 7/44 (16) | 3/21 (14) | 1/21 (5) | 4/16 (25) | 0.058 | 0.136 | 0.911 | **0.013** | 0.865 | 0.201 | 0.421 | 0.293 | 0.410 | 0.074 |
| *Axillary (vs. breast)* | 13/17 (76) | 1/2 (50) | 7/7 (100) | 3/3 (100) | 1/1 (100) | 1/4 (25) | **0.047** | 0.171 | 0.387 | 0.540 | - | - | **0.007** | - | **0.047** | 0.171 |
| Lymph node characteristics: size, $cm^2$ | 4 (2-9) | 4 (4-9) | 3 (1-4) | 5 (3-10) | 6 (4-29) | 4 (1-9) | **0.0288** | >0.9999 | >0.9999 | >0.9999 | 0.1069 | **0.002** | >0.9999 | >0.9999 | >0.9999 | 0.420 |
| Specimen appearance | | | | | | | | | | | | | | | | |
| Bloody (vs. chylous) | 123/150 (82) | 40/48 (83) | 39/44 (89) | 19/21 (90) | 14/21 (67) | 11/16 (69) | 0.466 | 0.438 | 0.122 | 0.209 | 0.823 | **0.033** | 0.068 | 0.060 | 0.095 | 0.893 |

Abbreviations: TB: tuberculosis; TBLs: tuberculous lymphadenitis; HIV: human immunodeficiency virus; ART: antiretroviral therapy; L: lymphotype; oL: overall lymphotype; dTBLs: definite tuberculous lymphadenitis; nTBLs: non-tuberculous lyphadenitis;

**Table S5: Demographic, clinical, and microbiological differences between dTBL lymphotypes.** dL3s had characteristics associated with more severe TBL. dL3s were more likely to have HIV and larger lymph nodes compared to dL1s and dL2s. dL2s were more likely to be female than dL1s.

| Characteristic | Total (n=89) | dL1 (n=48) (*Prevotella*-*Corynebacterium*) | dL2 (n=21) (Prevotella-*Streptococcus*) | dL3 (n=20) (*Mycobacterium*) | *p-value* (dL1 vs. dL2) | *p-value* (dL1 vs. dL3) | *p-value* (dL2 vs. dL3) |
|---|---|---|---|---|---|---|---|
| Age, years | 35 (29-40) | 33 (28-38) | 36 (28-46) | 37 (34-44) | >0.999 | 0.197 | 0.873 |
| Female | 48/89 (54) | 22/48 (46) | 15/21 (71) | 11/20 (55) | **0.050** | 0.491 | 0.275 |
| HIV | 49/89 (55) | 23/48 (48) | 11/21 (52) | 15/20 (75) | 0.733 | **0.040** | 0.133 |
| *CD4+* | 155 (76-251) | 157 (106-250) | 212 (64-385) | 92 (17-226) | >0.999 | 0.254 | 0.172 |
| *CD4+ <200 cells/µl* | 32/49 (65) | 16/23 (70) | 5/11 (45) | 11/15 (73) | 0.180 | 0.800 | 0.150 |
| *On ART* | 21/49 (43) | 11/23 (48) | 5/11 (45) | 5/15 (33) | 0.900 | 0.380 | 0.530 |
| Previous TB | 24/88 (27) | 11/47 (23) | 4/21 (19) | 9/20 (45) | 0.689 | 0.077 | 0.074 |
| Tobacco smoking | 21/89 (24) | 13/48 (27) | 4/21 (19) | 4/20 (20) | 0.480 | 0.540 | 0.940 |
| within 1 year of recruitment | 22/87 (25) | 7/47 (15) | 6/20 (30) | 9/20 (45) | 0.153 | **0.008** | 0.327 |
| *At recruitment* | 10/22 (45) | 2/7 (29) | 2/6 (33) | 6/9 (67) | 0.850 | 0.130 | 0.200 |
| Lymph node characteristics: sites | | | | | | | |
| Neck | 78/89 (88) | 42/48 (88) | 17/21 (81) | 19/20 (95) | 0.480 | 0.350 | 0.170 |
| *Deep anterior cervical* | 36/78 (46) | 16/42 (38) | 7/17 (41) | 13/19 (68) | 0.826 | **0.028** | 0.101 |
| *Deep lateral cervical* | 15/78 (19) | 11/42 (26) | 2/17 (12) | 2/19 (11) | 0.230 | 0.170 | 0.910 |
| *Superficial* | 6/78 (8) | 5/42 (12) | 0/17 (0) | 1/19 (5) | 0.140 | 0.420 | 0.340 |
| *Supraclavicular* | 17/78 (22) | 7/42 (17) | 7/17 (41) | 3/19 (16) | **0.045** | 0.920 | 0.090 |
| *Head* | 4/78 (5) | 3/42 (7) | 1/17 (6) | 0/19 (0) | 0.860 | 0.230 | 0.280 |
| Thorax | 11/89 (12) | 6/48 (13) | 4/21 (19) | 1/20 (5) | 0.480 | 0.350 | 0.170 |
| *Axillary (vs. breast)* | 9/11 (82) | 6/6 (100) | 2/4 (50) | 1/1 (100) | 0.053 | - | 0.361 |
| :haracteristics: size, $cm^2$ | 4 (2-9) | 4 (1-7) | 4 (1-4) | 8 (4-12) | 0.827 | **0.030** | **0.005** |
| Specimen appearance | | | | | | | |
| Bloody (vs. chylous) | 66/89 (74) | 38/48 (79) | 14/21 (67) | 14/20 (70) | 0.270 | 0.420 | 0.820 |

Abbreviations: TB: tuberculosis; TBLs: tuberculous lymphadenitis; HIV: human immunodeficiency virus; ART: antiretroviral therapy; dTBLs: definite tuberculous lymphadenitis; dL: dTBL lymphotype.

**Figure S1: Paired analysis of controls and lymph fluid (n=33) indicates that environmental cross contamination is highly unlikely.** (A) α-diversity analyses show skin has higher diversity than lymph fluid. (B) β-diversity of lymph fluid differs to saline and skin. DESeq2 volcano plots depicting differentially abundant taxa show that (C) lymph was enriched in *Mycobacterium* vs. (C) skin and (D) saline, and there were more differentially abundant taxa in skin vs. (C) lymph and (E) saline. Significantly more discriminatory taxa appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0.2) as the degree of significance increases. Relative taxa abundance is indicated by circle size.

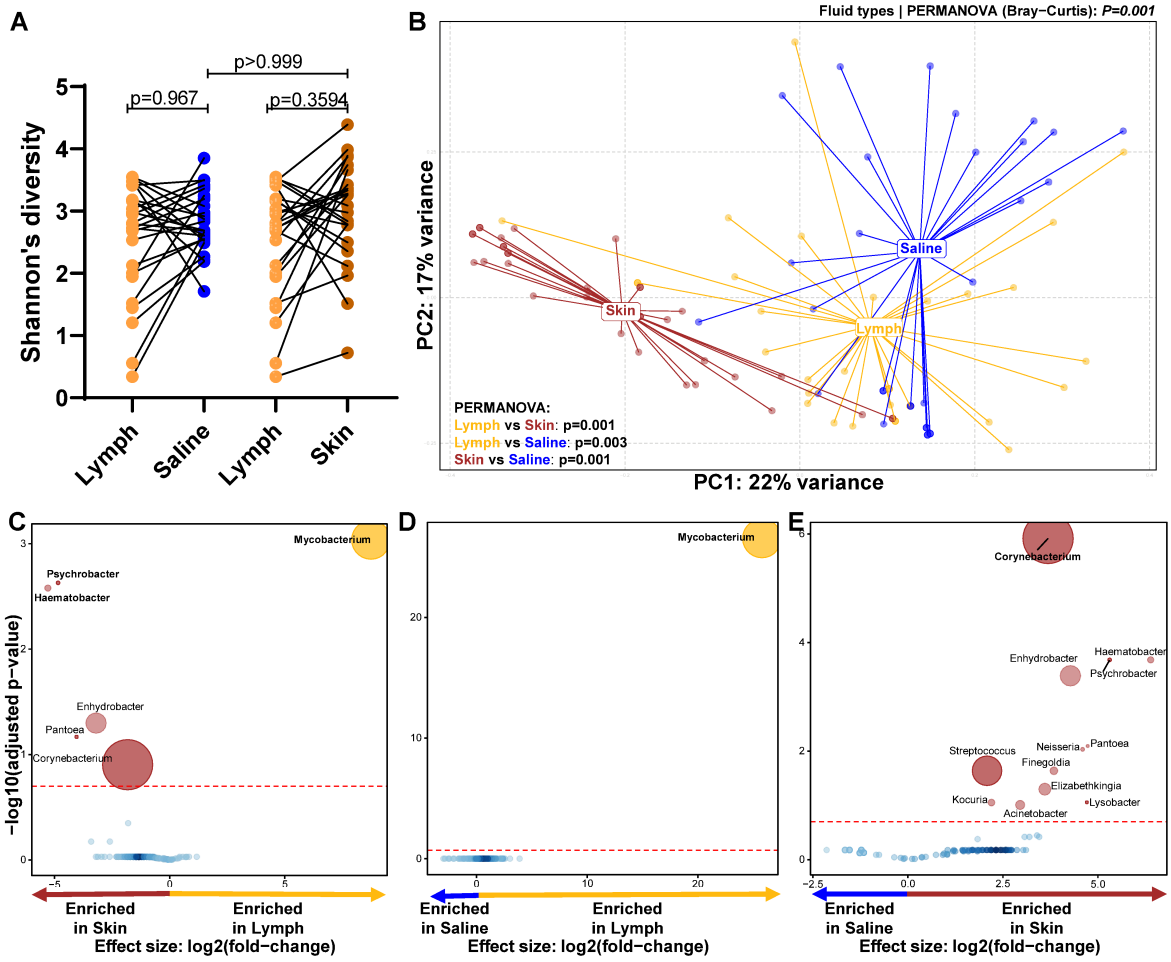
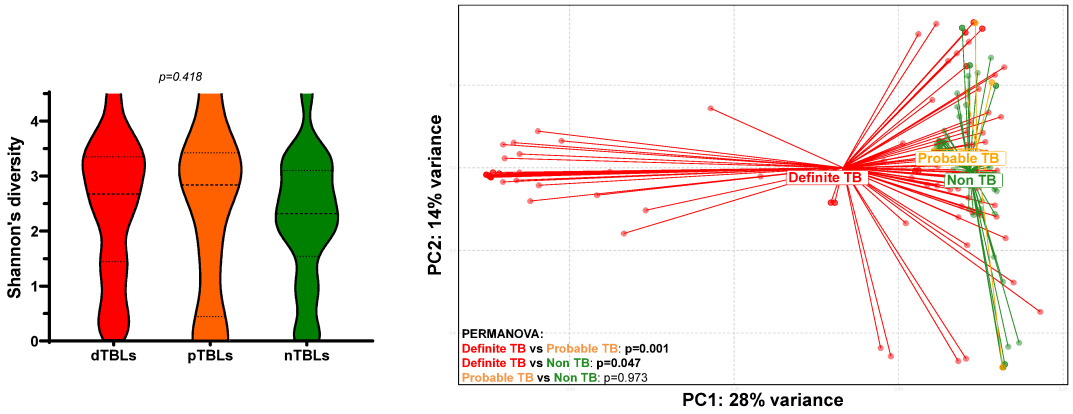


**Figure S2: dTBLs have a distinct microbiome to pTBLs and nTBLs s. (A) Microbial diversity is similar, but (B) microbial composition of dTBLs is different from both pTBLs and nTBLs**.






Nyawo GR, *et al. Thorax* 2022;0:1–12. doi: 10.1136/thorax-2022-219103

**Figure S3: *Mycobacterium* reads in FNABs of participants showing some nTBLs with Mycobacterium reads.** Relative abundance of *Mycobacterium* per participant stratified by TB status shows *Mycobacterium* in some nTBLs. Furthermore, not all dTBLs had detected *Mycobacterium* reads.

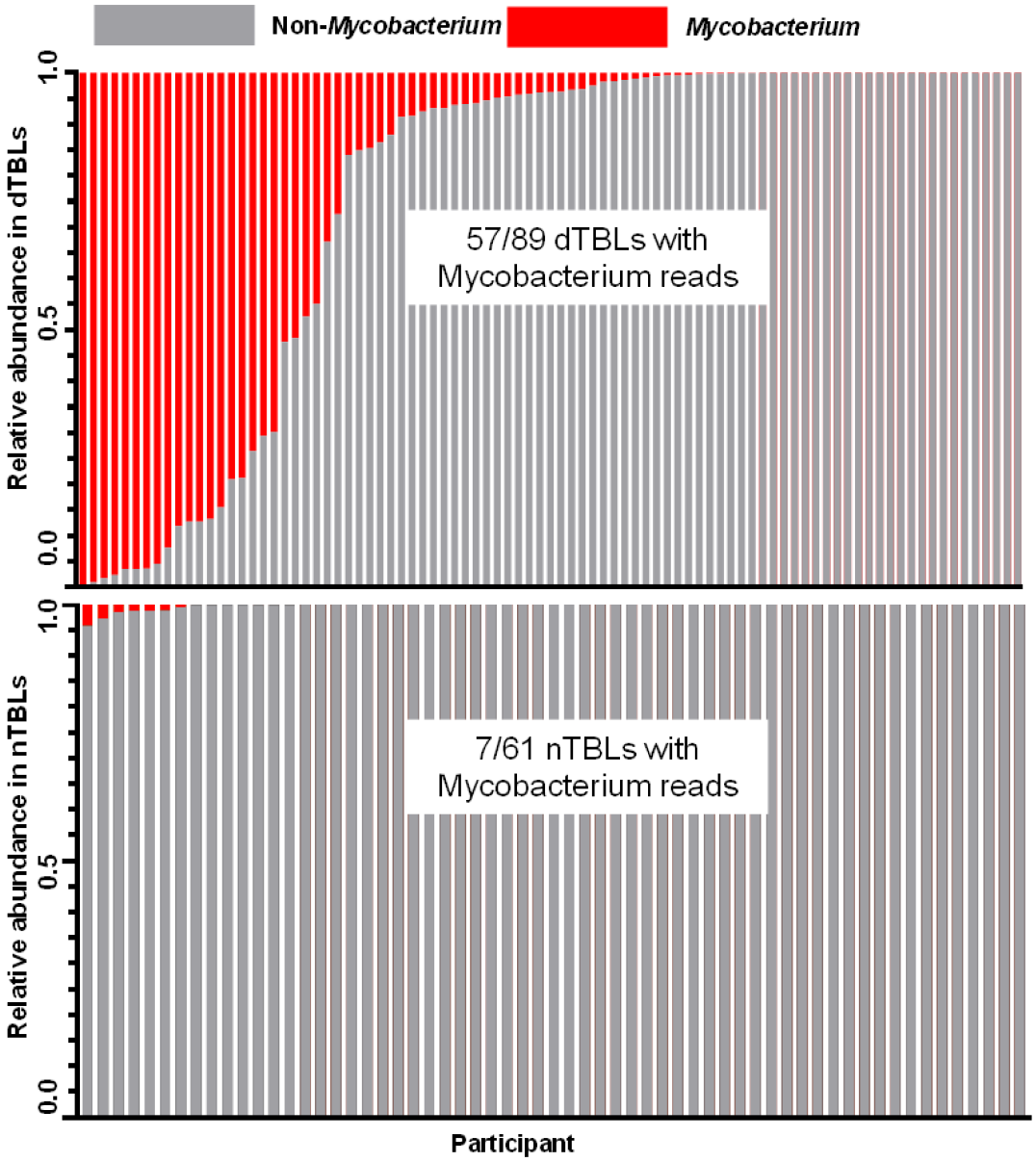


**Figure S4: 16S rRNA gene sequencing positively correlated to TB diagnostic tests. Mycobacterial reads positively correlated with *Mtb* load: (A) Xpert and Ultra and (B) culture (days-to-positivity); but not with (C) lymph node size. Xpert: Xpert MTB/RIF; Ultra: Xpert MTB/RIF Ultra; MGIT960: Mycobacteria Growth Indicator Tube 960 liquid culture; rs: Spearman correlation coefficient.**

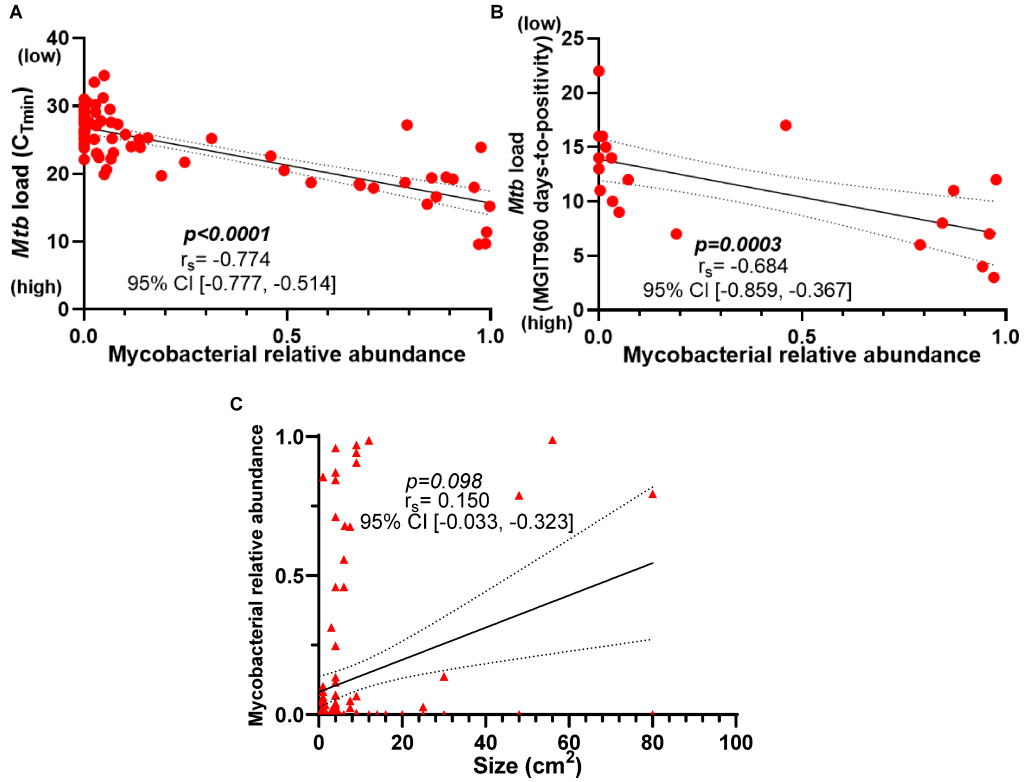


**Figure S5: HIV has a greater effect on the microbiome in patients co-infected with TB**. β-diversity did not differ (A) by HIV status in nTBLs, or by TBL status in (B) HIV-negatives. Circle sizes represent relative abundances. dTBLs: definite tuberculous lymphadenitis; nTBLs: non-tuberculous lymphadenitis**. (C)** *Mycobacterium* abundance did not differ by HIV status within dTBLs and within nTBLs, and (D) HIV-positive dTBLs were enriched in *Mycobacterium* compared to nTBLs. Circle sizes represent relative abundances. dTBLs: definite tuberculous lymphadenitis; nTBLs: nontuberculous lymphadenitis.

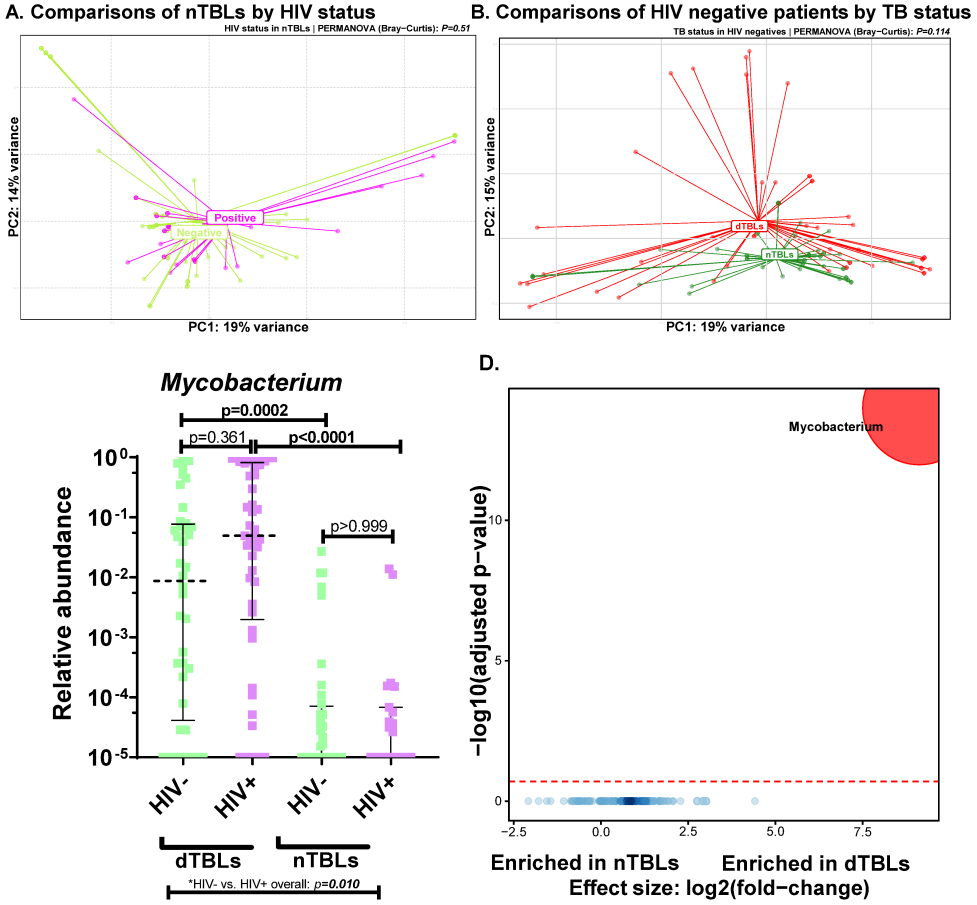


**Figure S6: Five microbial community states observed in presumptive TBL patients are enriched with distinct taxa. o**L1 had no enriched taxa, and was depleted in (A) *Enhydrobacter*, (B) *Mycobacterium*, and (C) *Streptococcus*, *Anaerosinus*, *Neisseria* and *Kocuria*. oL3 was enriched in (D) *Acinetobacter* and depleted of *Prevotella*. oL5 was enriched in *Streptococcus* accompanied with (E) *Anaerosinus*, *Neisseria*, *Kocuria* and *Prevotella* vs. oL2, and with (F) *Bacteroides* and *Kocuria* vs. oL3. Significantly more discriminatory taxa (bolded) appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0·2) as significance increases. Relative abundance of taxa is indicated by circle size. oL: Lymphotype.

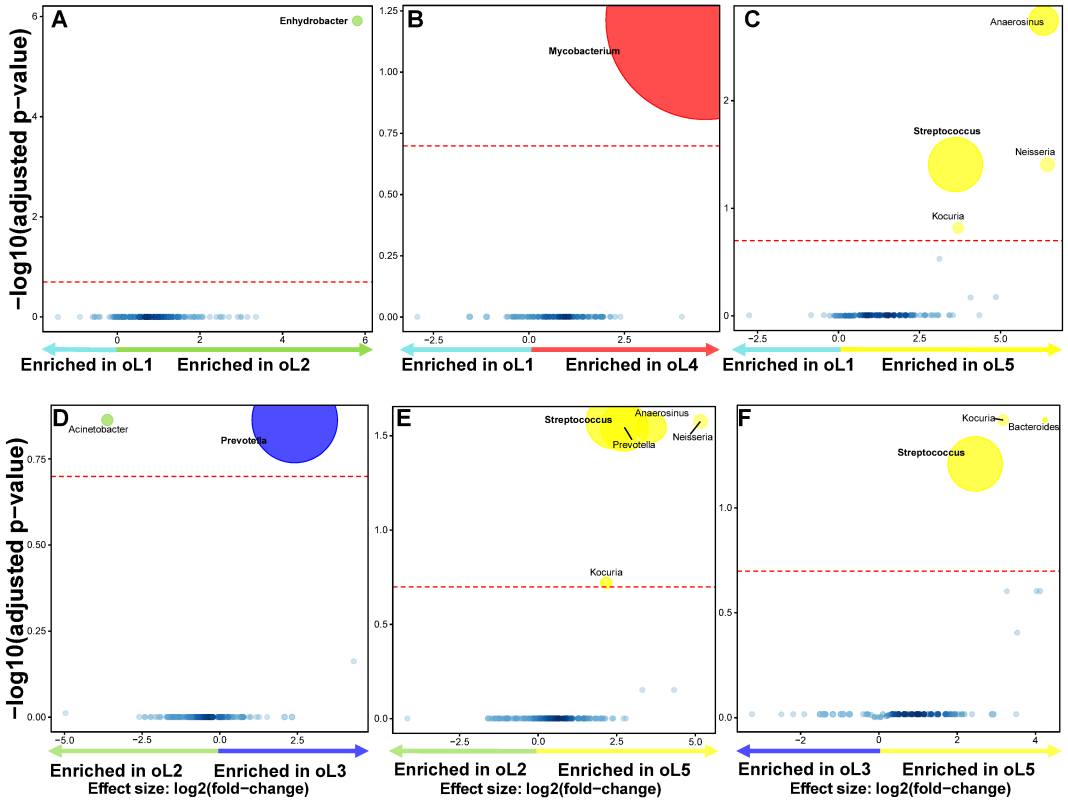


**Figure S7: The Laplace approximation of model evidence is a measure of the model fit**. Laplace approximation predicts no clustering for nTBL patients. Lower values indicate better fit. nTBLs: non-tuberculous lymphadenitis.

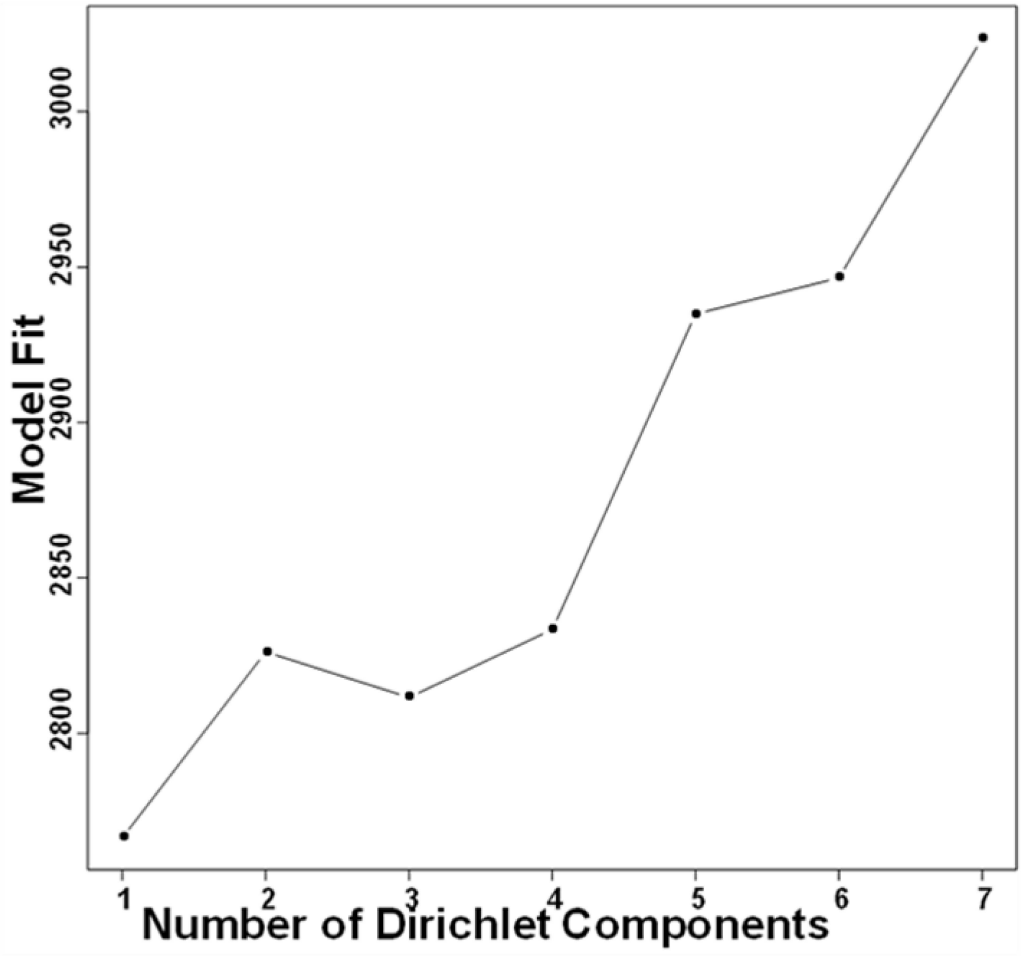


**Figure S8: Predicted metagenome function in HIV-positive nTBLs versus HIV-negative nTBLs.** Volcano plot depicting functional pathways differing between HIV-positive and HIV-negative nTBLs. Significantly more discriminatory pathways appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0.05). Key pathways of interest include "cell cycle - *Caulabacter*", "bacterial secretion system", "taurine and hypotaurine metabolism", and "histidine metabolism". Relative gene abundance is indicated by circle size.

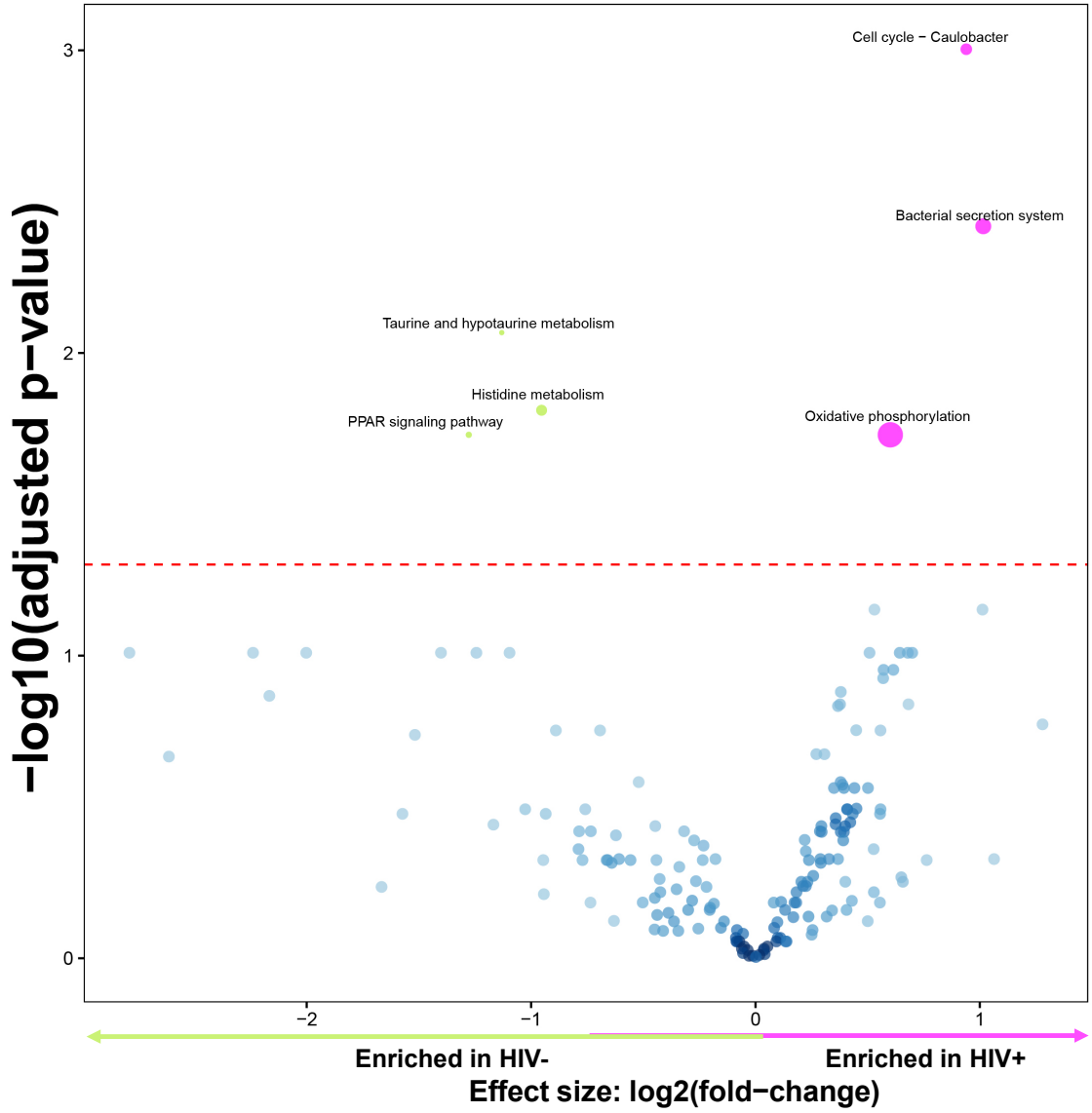


**Figure S9: Inferred metagenomes of lymphotypes in all patients.** Volcano plot depicting differentially enriched pathways in oL4 included pathways involving lipid biosynthesis, fatty acids, and SCFA metabolism i.e. lipid biosynthesis proteins, propanoate metabolism, benzoate degradation, and valine, leucine and isoleucine degradation. Significantly more discriminatory pathways appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0.05). Relative gene abundance is indicated by circle size. oL: overall Lymphotype.

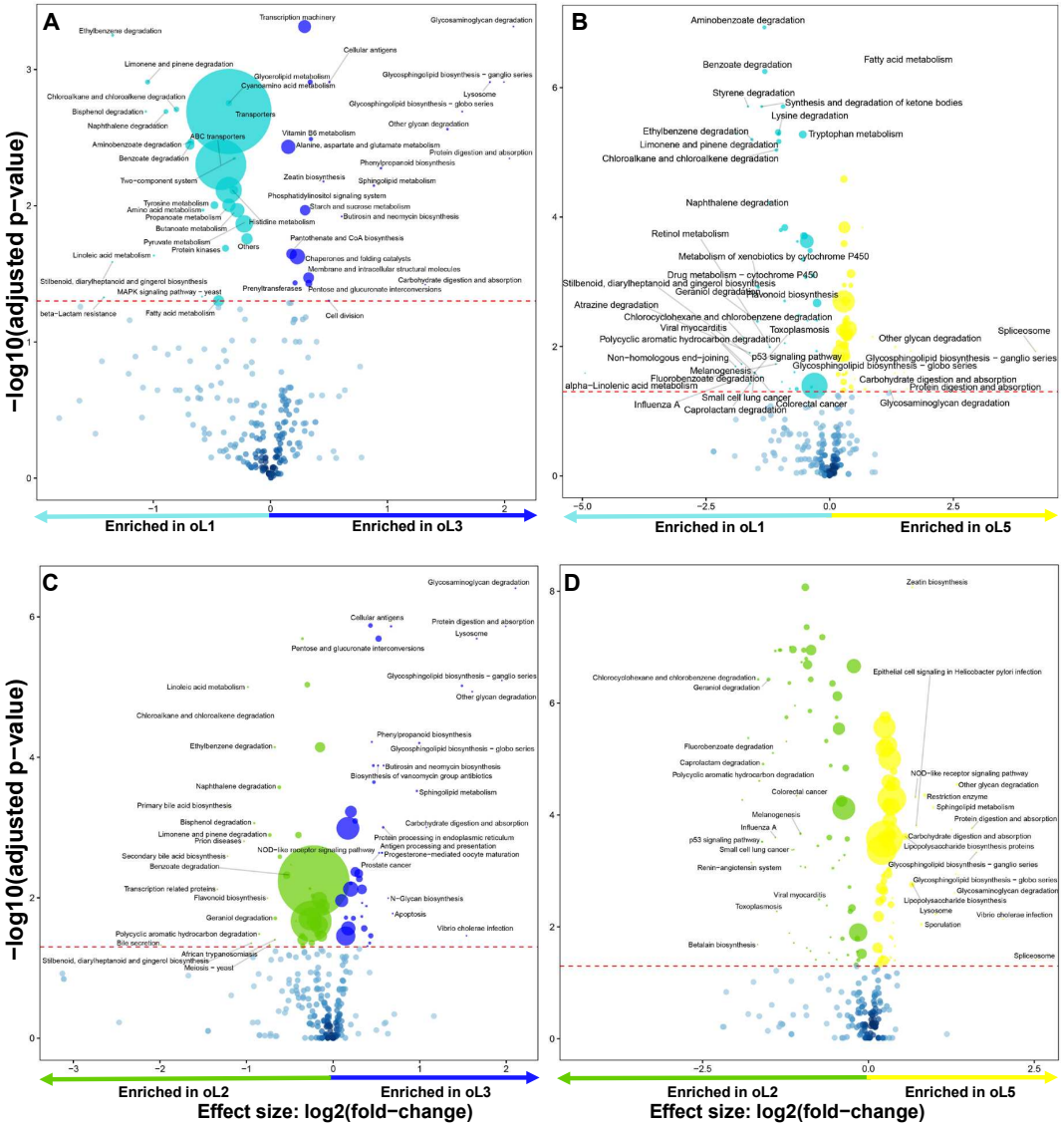


**Figure S9 cont.**

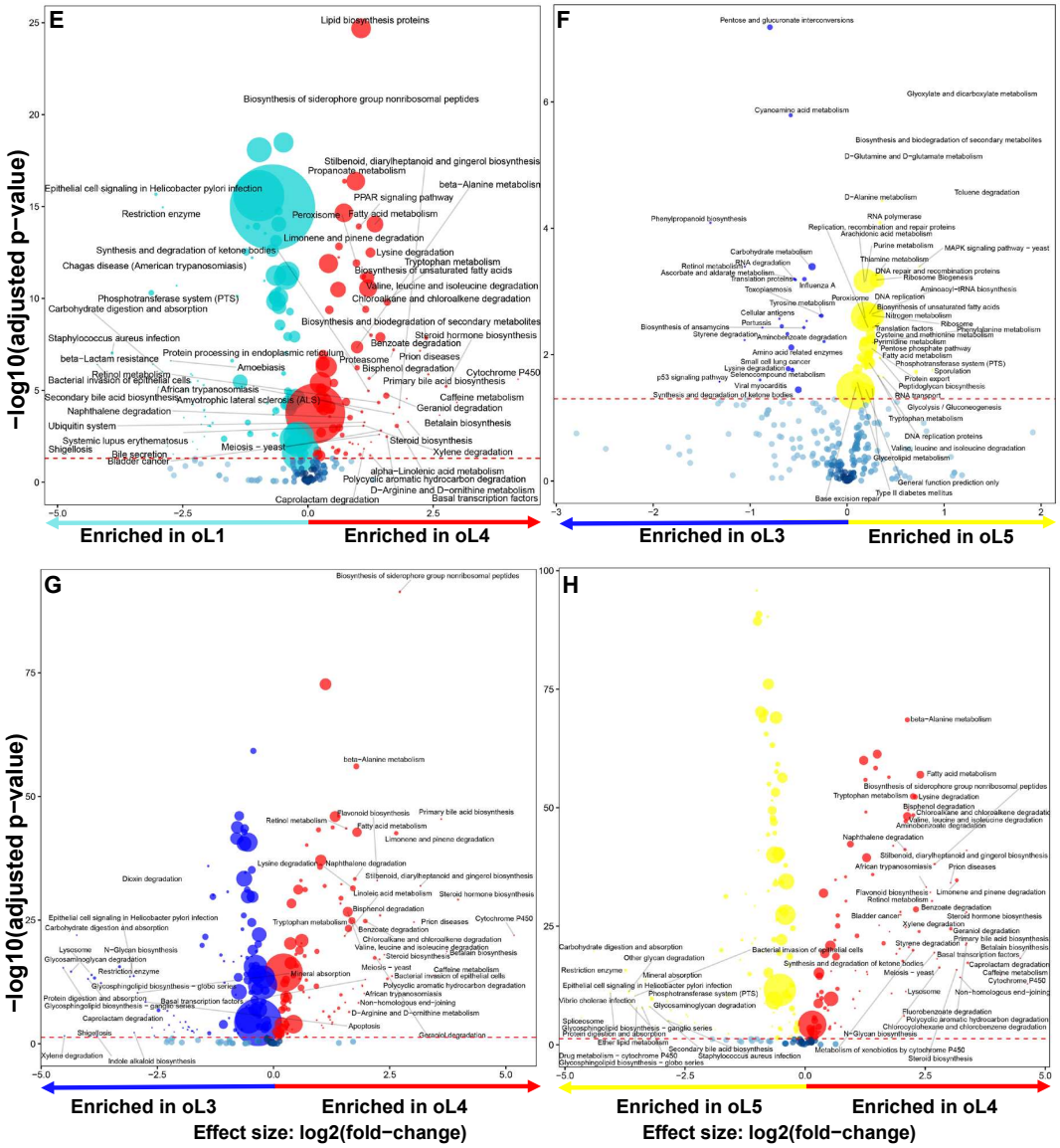
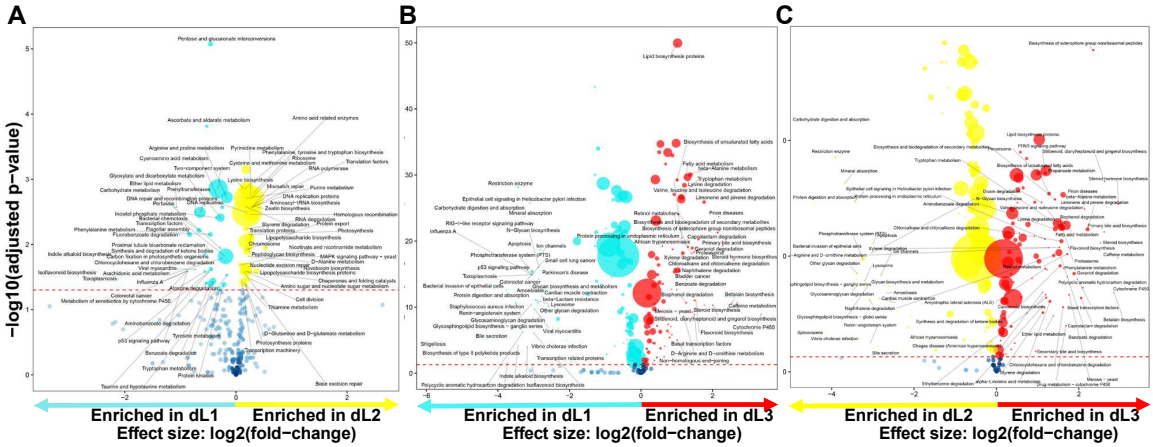


**Figure S10: Inferred metagenomes of lymphotypes in dTBLs.** Volcano plot depicting differentially enriched pathways in dL3 included pathways involving lipid biosynthesis, fatty acids, and SCFA metabolism i.e. lipid biosynthesis proteins, propanoate metabolism, benzoate degradation, and valine, leucine and isoleucine degradation. Significantly more discriminatory pathways appear closer to the left or right, and higher above the threshold (red dotted line, FDR=0.05). Relative gene abundance is indicated by circle size. dL: dTBL Lymphotype.




1
2

## References

1. Minnies S, Reeve BW, Rockman L, et al. Xpert MTB/RIF Ultra Is Highly Sensitive for the Diagnosis of Tuberculosis Lymphadenitis in a High-HIV Setting. *Journal of clinical microbiology* 2021;59(12):e01316-21.
2. Wright CA, Bamford C, Prince Y, et al. Mycobacterial transport medium for routine culture of fine needle aspiration biopsies. *Archives of disease in childhood* 2010;95(1):48-50.
3. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS ONE* 2012;7(2):e30126.