Title: Nasal Gene Expression in ART-Naive Adults with HIV and Pulmonary Tuberculosis in Uganda

Authors

Nisreen Khambati*[1]

Kattya Lopez Tamara*[2]

Elizabeth Nakabugo[3]

Arthur Van Valkenburg[4]

Jessica K Anderson[4]

Sean Lu[4]

Rinn Song[1]

Suryaram Gummuluru[5]

Andrew J Pollard[1]

Jerrold Ellner[2]

Padmini Salgame[2]

Else Margreet Bijker[1,6]

Lydia Nakiyingi[3]

Daniel O Connor[1]

W Evan Johnson[4]

*shared first author

Affiliations

1. Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom
2. Department of Medicine, Centre for Emerging Pathogens, Rutgers-New Jersey Medical School, Newark, NJ, USA
3. Makerere University College of Health Sciences, Kampala, Uganda

4. Division of Infectious Disease, Center for Data Science, Rutgers New Jersey Medical School, Newark, NJ, 07103, USA
5. Department of Virology, Immunology & Microbiology, Boston University Chobanian and Avedisian School of Medicine, Boston MA 02118
6. Department of Paediatrics, Maastricht University Medical Center, MosaKids Children's Hospital, Maastricht, The Netherlands

Corresponding author

Nisreen Khambati, nisreen.khambati@paediatrics.ox.ac.uk

Department of Paediatrics, Oxford University, Children's Hospital, John Radcliffe

Headington, Oxford, OX3 9DU

Abstract

Background: Diagnosis of pulmonary tuberculosis (TB) in people living with HIV remains difficult. Since the first pathogen-host interaction in TB occurs in the upper airway, we hypothesized that host transcriptomic analysis on nasal specimens may identify novel diagnostic biomarkers. We aimed to demonstrate differences in nasal gene expression between people with HIV and TB disease versus people with HIV without TB, evaluate the performance of nasal signatures in classifying TB and compare nasal gene profiles with blood gene profiles from the same cohort.

Methods: We enrolled adults in Uganda with newly diagnosed HIV and symptoms of pulmonary TB disease We collected nasal cells and blood for RNA sequencing to identify differentially expressed genes and enriched pathways between people with HIV and TB disease and people with HIV without TB. Supervised machine-learning of gene expression data was used to predict TB status.

Results: 40 adults living with HIV were enrolled (median age: 34 years, median CD4 count: 182), including 20 with TB disease and 20 without. We identified 44 nasal differentially expressed genes and 238 blood differentially expressed genes, with three overlapping genes between sample types. Models trained using all 44 nasal differentially expressed genes had a cross-validated area under the curve between 0.87-0.90 for predicting TB disease amongst adults living with HIV. A simplified four-gene signature (*SPIB, SHISA2, TESPA1* and *CD1B*) met the World Health Organization criteria for a TB triage test. Among adults with TB, pathways related to the inflammatory response and innate immune system were downregulated in nasal samples and upregulated in blood.

Conclusion: This proof-of-concept study demonstrated that there were distinct nasal gene expression patterns associated with TB, not seen in blood. Differences in nasal gene expression in people with HIV who have TB disease, versus those without TB, highlight their potential as diagnostic biomarkers.

Further validation studies of gene signatures using minimally invasive nasal samples are recommended in other difficult to diagnose groups.

Keywords

tuberculosis, HIV, diagnosis, nasal, transcriptomics

Manuscript

Background

Tuberculosis (TB) is the primary reason for hospitalization and death in people living with HIV (PLHIV)[1], with the highest TB-HIV coinfection in the African Region[2]. TB-related mortality in PLHIV is partly due to difficulties in diagnosis. Autopsy studies in PLHIV demonstrated that TB was undiagnosed in nearly half of cases identified post-mortem[3]. PLHIV are more likely to have paucibacillary disease and difficulty in expectorating sputum than those without HIV[1], limiting sputum-based tests such as culture or Xpert MTB/RIF Ultra (Ultra), a cartridge-based polymerase-chain-reaction (PCR) assay. Recognizing this, the World Health Organization (WHO) has advocated for development of non-sputum-based biomarker tests for diagnosis of pulmonary TB[4].

Several blood-based mRNA biomarkers have shown promise as triage tests for TB[5]. However, their potential has been limited by a poorer performance in PLHIV[6,7], particularly in antiretroviral therapy (ART) naive individuals[8], or when differentiating TB from other respiratory diseases[9]. Lower specificity in PLHIV could be due to the dominance of interferon-stimulated genes (ISGs) in blood and HIV-related viraemia or opportunistic infections upregulating interferon-driven pathways[8-10]. Therefore, alternative sample types for biomarker discovery are worth exploring. With the first pathogen-host interaction in TB occurring in the upper airway, transcriptomic analysis of nasal specimens may identify host gene profiles and pathways not previously identified in blood. Overlap in patterns of gene expression across the upper and lower airways in respiratory inflammatory conditions[11-13] suggests that nasal samples could be proxies for studying lung disease. Moreover, nasal specimens are also minimally invasive and easy to collect.

A rich transcriptomic response in nasal samples has been demonstrated for biomarker discovery and understanding pathophysiology for several non-TB respiratory illnesses, including respiratory syncytial virus[14,15], SARS-CoV-2[16-18], pneumonia[19], asthma[20] and lung cancer[12]. Pandya *et al* analyzed

1,555 nasal samples across ten cohorts and developed a 33-gene nasal signature that could classify viral acute respiratory infections from healthy or non-viral samples with over 80% specificity and sensitivity[21]. Amongst SARS-CoV-2 patients, suppressed interferon responses in nasal samples were linked to more severe disease[17]. In children, nasal gene signatures could accurately predict respiratory viral infections[14,22], and in some cases, performed better than blood classifiers[14]. Such studies highlight the value of studying the local nasal response to respiratory infections.

Despite these potential benefits, to our knowledge, nasal gene expression profiling has not been studied in TB. To address this gap, we sought to determine if nasal gene expression differs in ART-naive PLHIV and pulmonary TB disease compared with PLHIV without TB disease. We identified differentially expressed genes (DEGs) and biological pathways from nasal cells and assessed the performance of nasal signatures in classifying TB disease. As secondary objectives, we compared nasal-derived DEGs, biological pathways and gene signatures with paired blood samples from the same cohort and evaluated the diagnostic performance of existing published blood TB signatures in our nasal dataset.

Methods

Figure 1 provides an overview of the methods. Computational code for the methodology is available at: https://github.com/nisreenkhambati/uganda_nasal_cells.

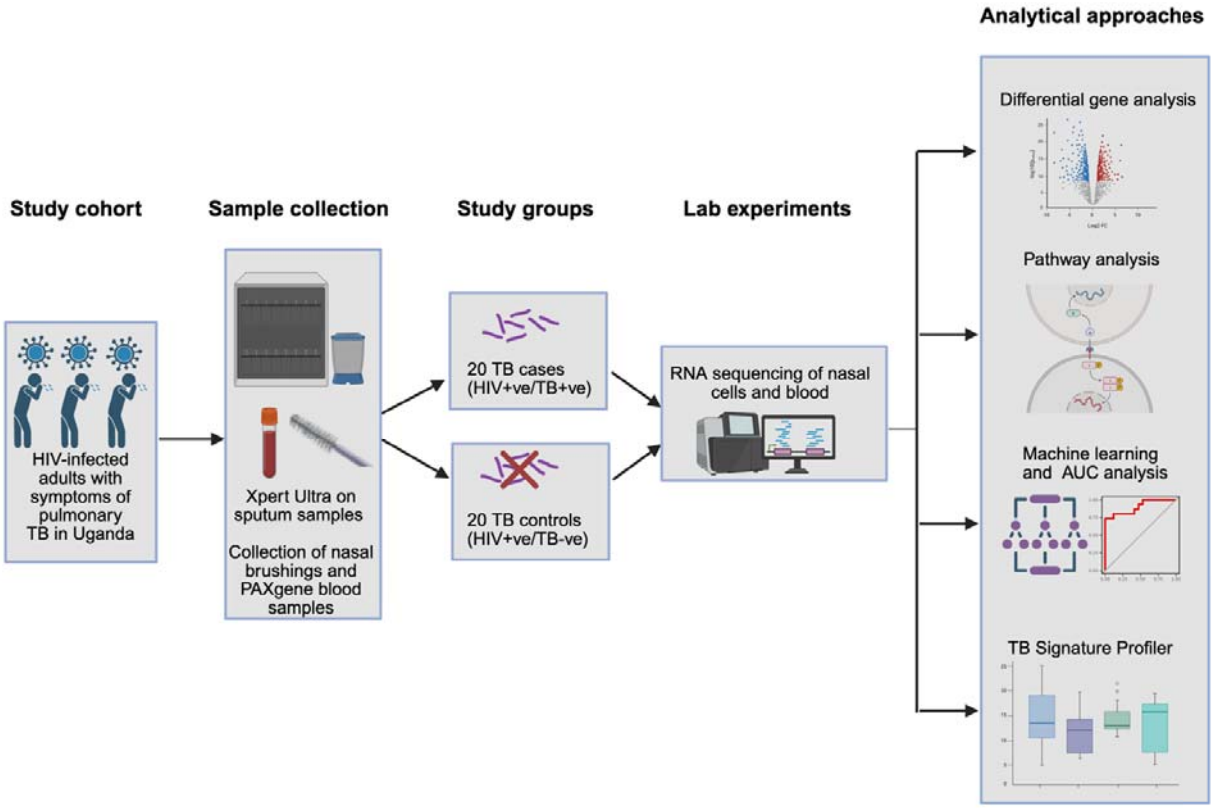

Figure 1: Overview of the methods. Adults in Uganda with newly diagnosed HIV and TB symptoms were enrolled and nasal and blood samples were collected for RNA sequencing. Analysis aimed to determine if nasal gene expression differed in PLHIV and pulmonary TB disease compared with PLHIV without TB disease. Figure created in BioRender.com, license obtained by Anderson, J., 2026, https://BioRender.com/3fuig8f)

Ultra= Xpert MTB/RIF Ultra; AUC= area under the curve; PLHIV=people with HIV

168 **Study cohort**

169 This study was conducted at Mulago National Referral Hospital and Infectious Diseases Institute-
170 affiliated clinics in Kampala, Uganda between 2020 and 2021. Individuals between 18-50 years with
171 presumptive TB (defined as having symptoms of pulmonary TB including cough, fever, night sweats,
172 weight loss and hemoptysis), no previous HIV or TB treatment (including isoniazid prophylaxis) and
173 newly confirmed HIV were consecutively enrolled. The HIV status of those with presumptive TB was
174 determined as part of study screening, i.e. HIV was diagnosed due to the presentation with TB
175 symptoms. Those who were current or former smokers, had diabetes, severe malnutrition, a history
176 of TB, or had used corticosteroids or antibiotics in the past three months were excluded. Informed
177 consent was obtained from all participants, and the study was approved by institutional review
178 boards of all partnering institutions (Boston University Medical Campus, Joint Clinical Research
179 Centre and Uganda National Council for Science and Technology).
180

181 **Definitions and sample collection**

182 Each adult had one early morning expectorated sputum tested with Ultra (Cepheid, CA, USA). Adults
183 were classified as having active TB disease based on a positive sputum Ultra. Control subjects were
184 those with no microbiological evidence of TB.
185

186 Study personnel followed a standardized protocol to collect nasal cells and blood. A cytology brush
187 was used to collect nasal cells by insertion into each nostril and rotating it against the lateral wall for
188 three seconds. The brush head was placed into a cryovial containing RNA-preserving agent. Blood
189 was collected into PAXgene RNA tubes. Samples were obtained before TB or HIV treatments were
190 initiated and frozen at -80°C.
191

192 **RNA sequencing and quality control**

193 Nasal and blood samples were shipped on dry ice to Rutgers University. Total cellular RNA from nasal
194 and PAXgene blood samples was extracted using the RNeasy Plus Micro Kit protocol (QIAGEN Inc.,
195 Hilden, Germany) at Rutgers University according to the manufacturer's instructions. The integrity
196 of RNA was analyzed using an Agilent TapeStation. To generate the cDNA library, the NEBNext®
197 rRNA Depletion Kit was used to remove the ribosomal RNA. Illumina-compatible RNA-seq library was
198 then prepared using the NEB Next Ultra II RNASeq library preparation kit. The cDNA libraries were
199 purified with AmpureXP beads and quantified using an Agilent TapeStation and a Qubit 4
200 Fluorometer. Equimolar amounts of barcoded libraries were pooled and sequenced on the
201 Illumina NovaSeq S2 FC 6000 (Illumina, San Diego, CA), producing 40-50 million paired-end 150 base-
202 pair reads per sample. All samples were run across two lanes in a single flow cell in the same
203 sequencing run.
204
205 Raw FASTQ data were assessed for quality using FastQC[23]. Trimmomatic was used to remove low
206 quality bases and adaptors (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36)[24]. Reads were
207 aligned to the human genome (GRCh38.p14) using the Spliced Transcripts Alignment to a Reference
208 (STAR) tool[25] and annotated using the GENCODE v44 primary assembly.[26]
209
210 Gene count data were exported to R (version 4.3.1) and subset for protein coding genes using the
211 biomaRt package[27]. Mitochondrial, hemoglobin, ribosomal and RNA genes, as well as low-count
212 genes, were filtered out. Outliers were identified by examining the Euclidean distances between
213 pairwise combinations of samples, visualized using ComplexHeatmap[28] and principal component
214 analysis (PCA). TB status, age, sex, CD4 count, and HIV stage were investigated as potential
215 expression-associated variables using PCA and corrected if needed (excluding TB status) for
216 downstream analyses. Differences in principal components for variables separating TB and controls
217 were statistically tested using the Student's t-test, given normality of data distribution. All PCA and

heatmap analyses were performed on transformed normalized counts (variance stabilizing transformation).

Differential gene expression, machine learning and ROC analysis

Differential gene expression analysis between PLHIV and TB disease and PLHIV without TB disease (controls) was performed in the nasal and blood datasets using DESeq2[29]. DEGs were defined as having an adjusted p-value of <0.05 and log2 fold-change ≥1. Identified DEGs were used to cluster samples, visualized with ComplexHeatmap[28], using Euclidean distance as a measure of dissimilarity and complete linkage for between-cluster separation. Gene set enrichment analysis (GSEA) was performed using fgsea[30] to evaluate whether DEGs from one tissue were enriched in the ranked list of genes from the other tissue. Genes were ranked based on the test statistic from DESeq2[29].

Models were trained with six machine learning algorithms based on all DEGs using caret[31]: random forest (RF), support vector machine (SVM) with radial and linear kernels, elastic net, partial least squares (PLS) discriminant analysis and k-nearest neighbors (KNN), with 10-fold cross-validation repeated 10 times. These models incorporate both linear and non-linear data relationships and are commonly used in TB diagnostic transcriptomic research[32-36]. Prior to model training, gene counts of nasal and blood DEGs underwent variance stabilizing transformation. Predictors were pre-processed by removing near zero-variance variables and centering and scaling for models sensitive to feature scaling (elastic net, SVM, KNN and PLS) were also performed. Receiver operating characteristic (ROC) curves for each model in diagnosing TB were constructed with the MLeval package[37] to estimate the area under the curve (AUC), sensitivity, and specificity with 95% confidence intervals. Cut-off thresholds were determined by the maximum Youden index[38].

Models with a reduced gene panel are potentially easier to translate to a diagnostic test. Variable importance for each nasal model trained using all DEGs was estimated using caret[31]. The top 10

predictive genes for each model were identified and genes common to all six nasal models were retained to refit a new model using only these genes. The AUC, sensitivity and specificity of this parsimonious nasal signature in classifying TB was calculated using MLeval[37].

Finally, the INTERFEROME database[39] was used to identify which top DEGs were interferon-stimulated genes, which we defined as genes that were significantly up or down regulated (>2.0-fold change, $p<0.05$) in human samples treated with interferon versus controls.

Pathway analysis and sample deconvolution

Pathway analyses for nasal and blood samples were performed with GSEA using the fgsea package[30], and over-representation analysis (ORA) using the clusterProfiler package[40]. For fGSEA, ranked gene lists using the test statistic were compared to Hallmark gene sets and KEGG and Reactome pathways from the Molecular Signatures Database (MSigDB)[41]. For ORA, gene ontology biological processes[42] (GO BP) were evaluated in the differentially expressed upregulated and downregulated genes separately in adults with TB. Pathways with adjusted p-value < 0.05 were considered functionally enriched.

Nasal and blood gene lists from DESeq2 were also uploaded to the Qiagen's Ingenuity Pathway Analysis (IPA) tool[43] and filtered by adjusted p-value (false discovery rate) < 0.1 to determine canonical pathways and upstream regulators in TB. IPA uses a Fisher's exact test to compute p-values and uses the log2 fold-change to compute z-scores. For visualization, canonical pathways and upstream regulators with a p-value <0.05 and IPA z-score≥ ±1.5 were shown.

Nasal samples consist of both immune and epithelial cell components. The CIBERSORTX tool and the validated LM22 gene signature matrix[44,45] were used to examine the immune cell composition in

nasal and blood samples. CIBERSORTX is a deconvolution tool that employs support vector regression to estimate the relative proportion of immune cell subtypes from mixed bulk tissue transcriptome profiles[45]. The LM22 matrix consists of 547 genes that can accurately distinguish 22 immune cell populations, which can further be summarized into 11 major cell types based on shared lineage, including B cells, CD4 cells, CD8 cells, gamma delta T cells, eosinophils, plasma cells, NK cells, macrophages, neutrophils, dendritic cells and mast cells[44]. Counts data were first normalized as transcripts per million (TPM) to match the signature matrix. CIBERSORTX was performed with B-mode batch correction, quantile normalization disabled, and the number of permutations set to 500. Results were expressed as fractions of total leukocyte content. Proportions of estimated immune cell populations were compared by TB status using either Student's t-test or Mann-Whitney U test depending on normality of data distribution, with analyses corrected for multiple comparisons.

Performance of published blood RNA signatures

The TBSignatureProfiler package[46] was used to apply published blood signatures that diagnose TB disease to our nasal and blood RNA sequencing datasets. We selected blood signatures that were shown to distinguish TB disease from latent TB infection, TB from some combination of other diseases or TB from HIV. Raw counts were first converted into log counts per million (CPM). Signature performance scores were calculated using gene set variation analysis (GSVA) and single-sample GSEA (ssGSEA) and compared using two-sample t-tests and bootstrapped AUCs. Only published blood signatures with two or more genes that coincided with genes in our sequencing data were used in the analysis.

Results

295 **Study participants**

296 Forty PLHIV were enrolled, including 20 with pulmonary TB disease and 20 without (controls). The
297 median age was 35 years, 45% were male, 55% had HIV stage 3 and the median CD4 count was 182.
298 PAXgene samples were collected from all, whereas nasal samples were available from 35 (17 with TB
299 and 18 controls). Adults with TB and controls did not differ in age, sex, CD4 count or HIV stage
300 **(Supplementary Table 1).**
301

302 **Quality control and PCA**

303 67% of nasal and 57% of blood reads uniquely mapped to the human genome **(Supplementary Table**
304 **2).** Three nasal samples (S4, S20, S31) demonstrated high average sample-to-sample distances from
305 most other samples and had low reads (<5 million reads) **(Supplementary Figure 1).** These samples
306 were removed, with 15 adults with TB and 17 controls remaining for downstream nasal analyses. No
307 blood samples failed quality control, hence none were excluded.

308

309 Using PCA, nasal samples from adults with TB and controls clustered differently on PC4 ($p= 0.048$)
310 **(Supplementary Figure 2A, 2C).** In blood, differences in PC1 between TB and controls were
311 significant ($p = 6.853\times10^{-4}$) **(Supplementary Figure 2B, 2D).** Males clustered separately from females
312 for PC3 ($p= 2.798 \times10^{-3}$) and PC4 ($p= 4.943 \times10^{-7}$) in nasal samples and for PC3 ($p=5.449 \times10^{-5}$) and
313 PC4 ($p=1.303 \times10^{-5}$) for blood, whereas HIV-related variables and age did not separate samples
314 **(Supplementary Figure 3)**. After sex-correction, PC3 in nasal samples and PC1 in blood samples
315 demonstrated significant differences between TB and controls (nasal PC3: $p = 0.046$, blood PC1: $p =$
316 2.706×10^{-4}) **(Supplementary Figure 4).**
317

Differential gene expression and ROC curve analysis

44 nasal DEGs and 238 blood DEGs were identified between TB and controls **(Figure 2, Supplementary Table 3).** Nine of the top ten blood DEGs were classified as ISGs compared to four of the top nasal DEGs. Hierarchical clustering based on the DEGs showed grouping of samples by TB status, although overlap remained between the two groups **(Supplementary Figure 5).**

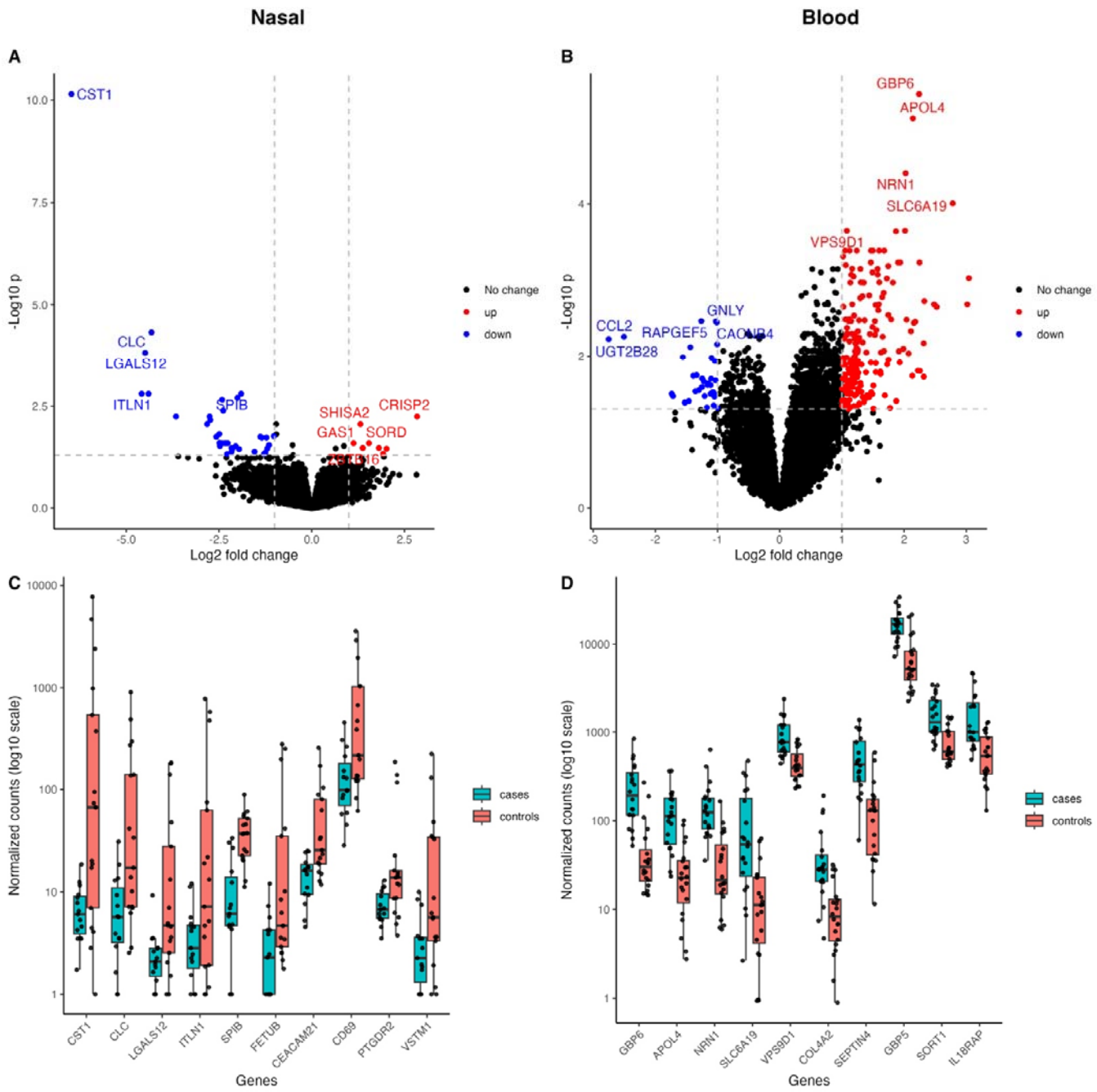

Figure 2: Volcano plots and box plots illustrating the top DEGs in PLHIV and TB compared to PLHIV without TB (controls). Volcano plots show the effect size by the -$\log_{10}$P value for nasal (A) and blood (B) samples. DEGs are highlighted in red (upregulated) or blue (downregulated), with gene names shown for the top 5 upregulated and downregulated DEGs. Boxplots are of normalized counts ($\log_{10}$ scale) for the top ten DEGs, which were all downregulated for nasal samples (C) and upregulated for blood samples (D) in TB cases. Individual data points are shown as jittered dots. DEGs=differentially expressed genes

In TB compared with controls, most nasal DEGs were downregulated whereas most blood DEGS were upregulated, with only three overlapping DEGs **(Figure 3).** Based on fGSEA, nasal DEGs were significantly negatively enriched in the ranked blood gene list and blood DEGs were also negatively enriched in the ranked nasal gene list **(Supplementary Figure 6),** suggesting that genes respond differently to TB disease in the two tissues, with compartment-specific immune responses.

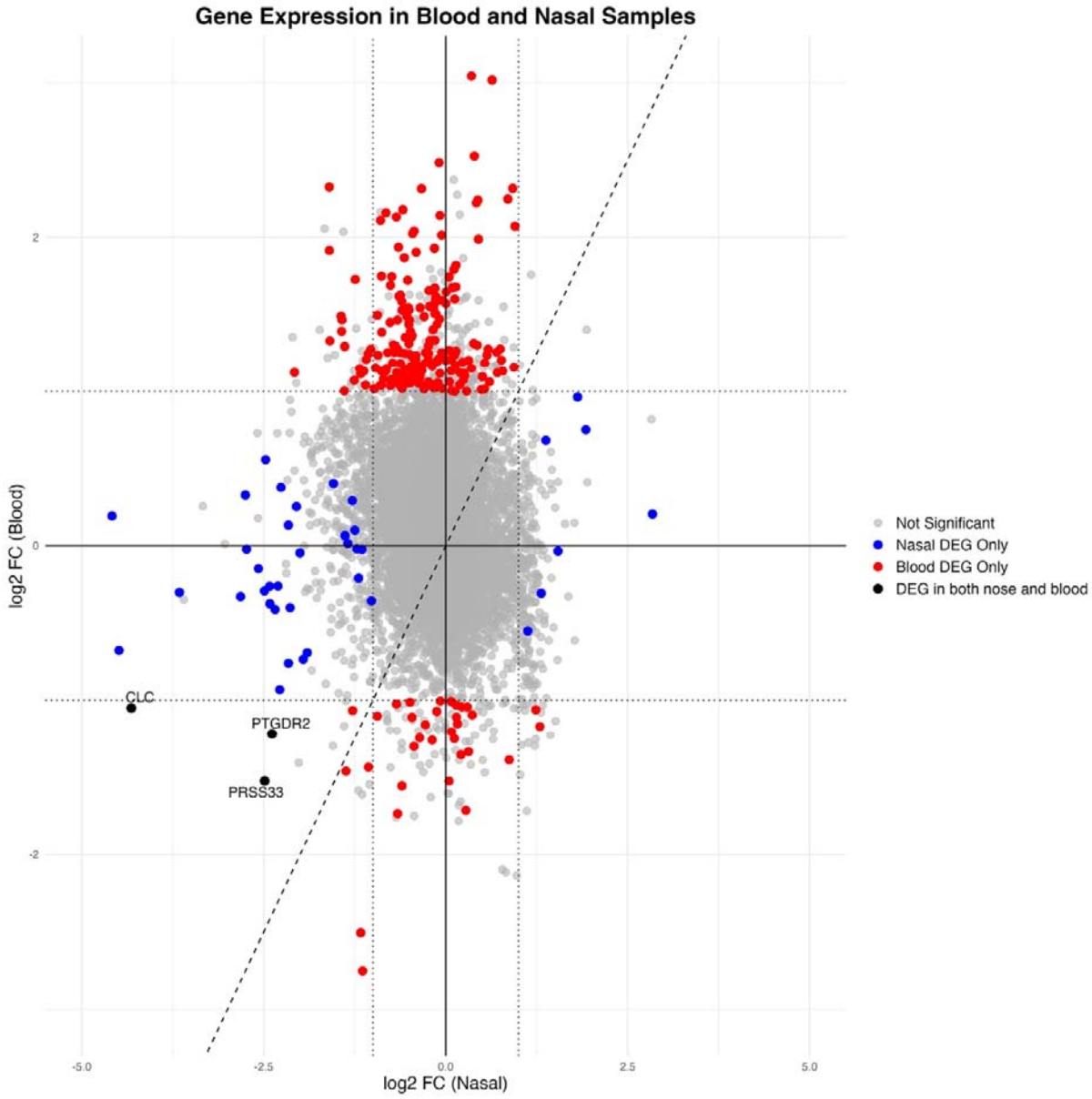

Figure 3: Agreement plot of log2FC for genes in blood samples (x-axis) and nasal samples (y-axis). Colored features are genes which were differentially expressed with log2 FC ≥1 and an adjusted p-value of <0.05. Dashed x=y line represents agreement. Dotted lines are shown at +/-1 log2FC. Labelled DEGs are those which are common to both nasal cells and blood and were downregulated in TB.

DEGs= differentially expressed genes; log2 FC= log2 fold-change.

Models trained using all 44 nasal DEGs had a cross-validated AUC between 0.87-0.90 for predicting TB status in the nasal dataset, which was similar to models trained with the 238 blood DEGs (AUC: 0.89-0.98) in blood **(Figure 4, Table 1).** The highest AUC for the nasal models was RF. All four adults with TB that were missed by the nasal RF model were correctly identified as TB by the blood RF model. Likewise, the two adults with TB missed by the blood RF model were correctly classified by the nasal RF model.

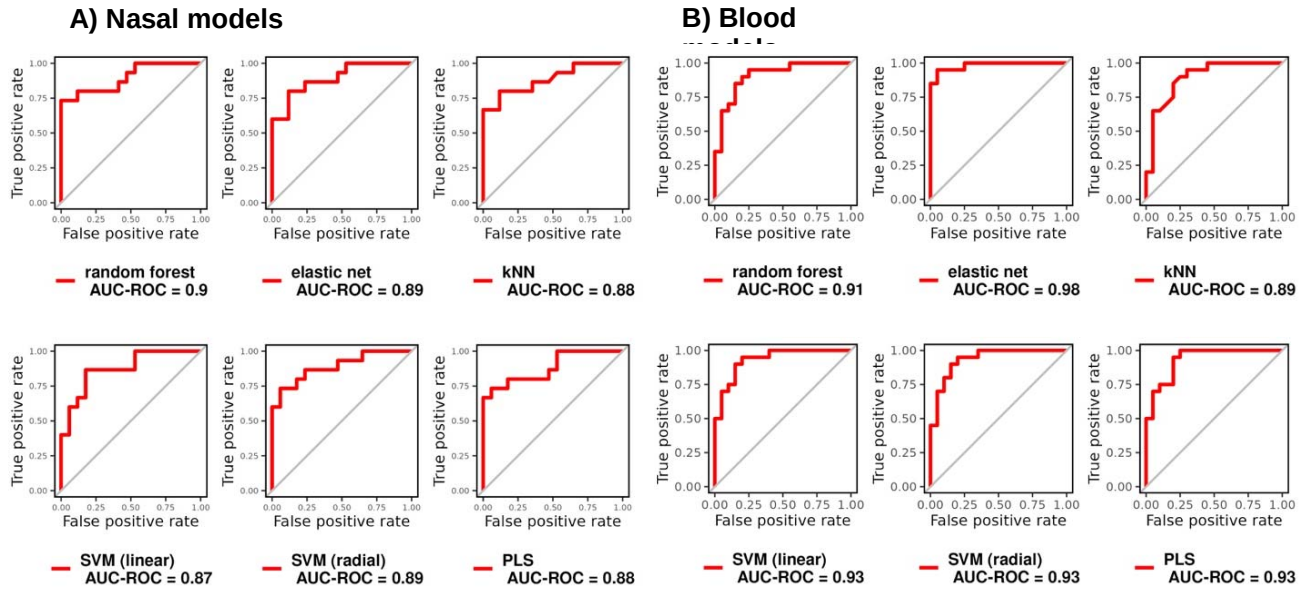

Figure 4: ROC curves and the respective AUC of different machine learning models in distinguishing TB from controls, trained using the (A) 44 nasal DEGs and (B) 238 blood DEGs on the respective datasets, with 10-fold cross-validation repeated 10 times.

DEGs= differentially expressed genes, SVM= support vector machine; KNN= k-nearest neighbor, ROC=receiver operating characteristic; AUC= area under the curve, PLS= partial least squares

Table 1: Diagnostic performance with 95% CI of each model

	NASAL DATA			BLOOD DATA		
	AUC (95%CI)	sensitivity (95%CI)	specificity (95%CI)	AUC (95%CI)	sensitivity (95%CI)	specificity (95%CI)
k-Nearest Neighbors	0.88 (0.75-1.01)	0.80 (0.55-0.93)	0.88 (0.66-0.97)	0.89 (0.78-1)	0.85 (0.64-0.95)	0.80 (0.58-0.92)
Support Vector Machine (radial)	0.89 (0.77-1.01)	0.73 (0.48-0.89)	0.94 (0.73-0.99)	0.93 (0.85-1.01)	0.90 (0.7-0.97)	0.85 (0.64-0.95)
Support Vector Machine (linear)	0.87 (0.74-1)	0.87 (0.62-0.96)	0.82 (0.59-0.94)	0.93 (0.85-1.01)	0.90 (0.7-0.97)	0.85 (0.64-0.95)
Partial Least Squares	0.88 (0.75-1.01)	0.73 (0.48-0.89)	0.94 (0.73-0.99)	0.93 (0.85-1.01)	0.95 (0.76-0.99)	0.80 (0.58-0.92)
Elastic Net	0.89 (0.77-1.01)	0.80 (0.55-0.93	0.88 (0.66-0.97)	0.98 (0.93-1.03)	0.95 (0.76-0.99)	0.95 (0.76-0.99)
Random Forest	0.90 (0.78-1.02)	0.73 (0.48-0.89)	1.00 (0.82-1)	0.91 (0.81-1.01)	0.90 (0.7-0.97)	0.80 (0.58-0.92)

All models were trained using the 44 nasal DEGs and 238 blood DEGs on the respective datasets in distinguishing TB from controls. Sensitivity and specificity were estimated at the optimal threshold defined by the Youden index.

DEGs= differentially expressed genes, AUC= area under the curve ; CI= confidence intervals

Amongst the top ten nasal DEGs with the highest variable importance for each model, *SPIB, SHISA2, TESPA1* and *CD1B* were common to all nasal models (**Supplementary Figure 7).** Models trained using this parsimonious four-gene signature (NASAL_4) had an AUC between 0.87-0.94 for predicting TB status in the nasal data, with 5/6 models meeting the WHO target criteria for a TB triage test (>90% sensitivity and >70% specificity)[4] **(Supplementary Figure 8 and Supplementary Table 4)**.

Enrichment pathway analysis

GSEA results are summarized in **Table 2.** The top GSEA enriched nasal pathways were interferon gamma (IFN-γ), neutrophil degranulation and chemokine signaling, all downregulated in adults with TB compared with controls. Enriched downregulated pathways related to cytokine signaling and the inflammatory response, particularly the innate immune system. Upregulated enriched pathways were associated with cilia function, organelle maintenance and cholesterol metabolism. In contrast, in blood, upregulated enriched pathways in TB related to immune activity and hem metabolism,

whereas downregulated pathways were associated with transcription and RNA processing. Enriched pathways in interferon and interleukin signaling, neutrophil degranulation, TNF-alpha signaling, complement, inflammatory response, phagocytosis, and leishmania infection were downregulated in nasal samples and upregulated in blood samples amongst adults with TB.

Using ORA, no enriched nasal pathways were identified with an adjusted p-value of <0.05. With an adjusted p-value of <0.1, upregulated nasal DEGs in TB pertained mainly to metabolic biological processes and cell division and differentiation whereas inflammatory and immune pathways were enriched in the downregulated genes **(Supplementary Figure 9A, 9C).** In blood, upregulated DEGs were enriched in the inflammatory and immune response, whereas biological processes in the downregulated DEGs included synaptic transmission, odontogenesis and regulation of cell migration (adjusted p-value <0.05) **(Supplementary Figure 9B, 9D).**

Canonical pathways in IPA also demonstrated suppressed immune activity in nasal samples from adults with TB, including reduced interferon signaling, phagosome formation, and S100 family and G-protein-coupled receptor signaling. In contrast, blood showed activation of innate immune pathways, such as neutrophil degranulation, IFN-γ, toll-like receptor signaling, inflammasomes, pyroptosis and iNOS signaling **(Supplementary Figure 10A, 10B)**. This opposing pattern of inflammation and interferon response was supported by upstream regulator analysis which predicted inhibition of type I and type III interferon expression (IFNA2, IFNL1), interferon regulatory factors (IRF1), JAK- STAT components and pathogen-recognition receptors (TLR9) in the nose and activation of inflammatory regulators (IFN-γ, CSF2, IL5, TNF, IL15, TLR2, SELPG) in blood **(Supplementary Figure 10C, 10D)**. Overall, GSEA, ORA and IPA suggested there was reduced nasal mucosal inflammation and increased systemic inflammation in TB disease.

424

Table 2 should appear here. Table 2 is larger than one A4 page in length and as per BMC Infectious Diseases Submission Guidelines is found at the end of the document text file.

428 **Immune cell composition**

429 Enrichment analysis suggested that the immune response, particularly the innate immune system,
430 was downregulated in the nasal samples and upregulated in blood in adults with TB. Heterogeneity
431 in immune cell composition was explored using CIBERSORTX and LM22. There were no statistically
432 significant differences in the proportions of immune cell subtypes in nasal samples based on TB
433 status. However, for blood, the proportion of neutrophils was higher in TB (adjusted p-value =
434 0.0376), whereas natural killer cells were lower in TB (adjusted p-value = 0.0376). Other types of
435 immune cells in blood showed no significant differences **(Supplementary Figure 11).**
436

437 **TB Signature Profiler**

438 In the RNA sequencing dataset, only 6/59 published TB diagnostic blood signatures distinguished TB
439 from controls with an AUC ≥0.7 and p-value <0.05 in the nasal dataset (Maertzdorf_100, Lee_5,
440 Sloot_2, Natarajan_7, Jacobensen_3 and Gjoen_10[46]) **(Figure 5A, 5C).** In contrast, most published
441 blood signatures distinguished TB from controls in our blood dataset (44/59 and 41/59 had an AUC
442 ≥0.7 and p-value <0.05 using ssGSEA and GSVA respectively) **(Figure 5B, 5D).** Only one signature
443 which performed well in the nasal dataset (Jacobsen_3) performed well in the blood dataset. Full
444 AUC results are in **Supplementary Table 5 and 6.**

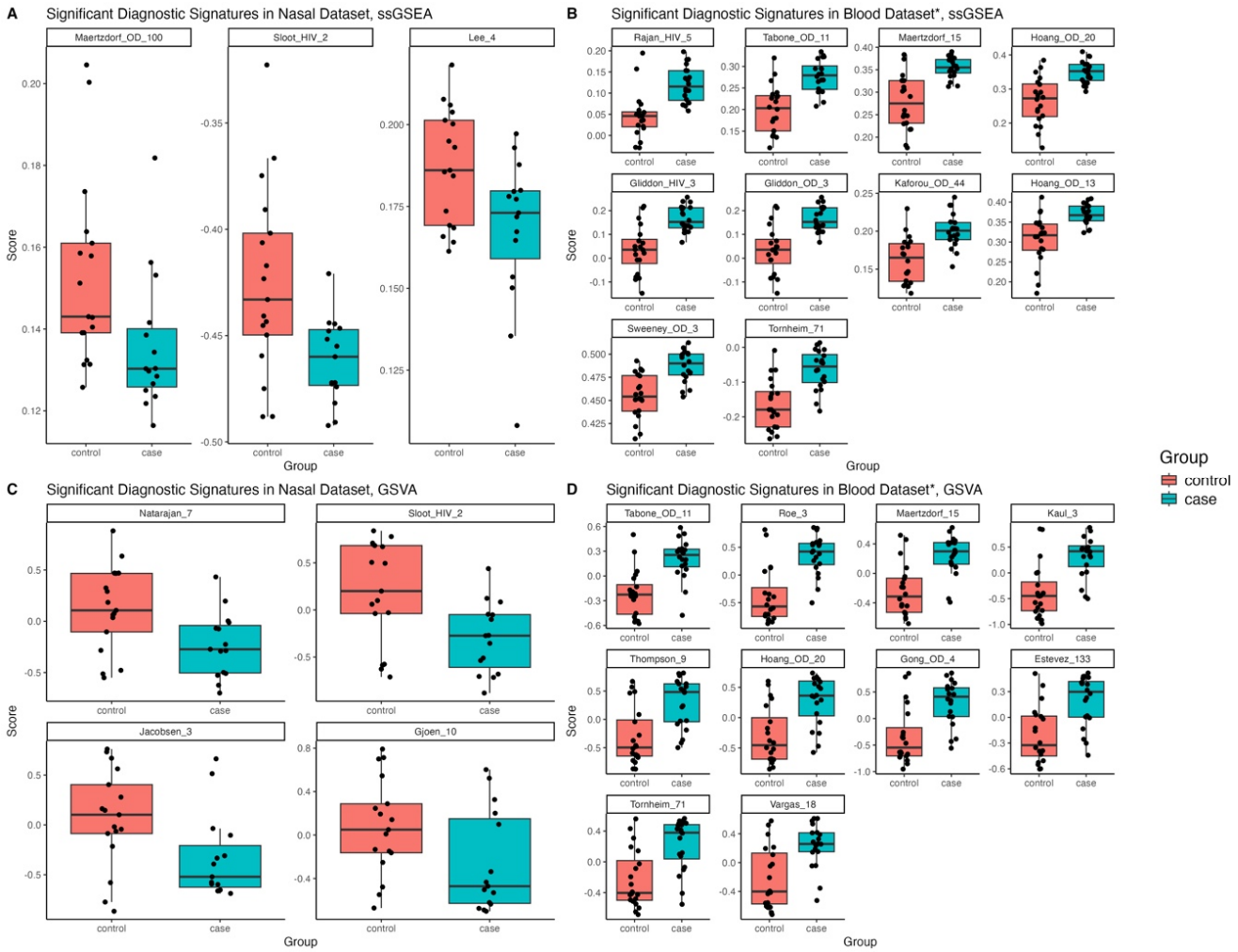

Figure 5: Boxplots of published blood diagnostic signatures that reached statistical significance (p value <0.05) and had a bootstrapped AUC of >0.7 in distinguishing TB from controls using ssGSEA in the A) nasal and B) blood datasets and using GSVA methods in the C) nasal and D) blood datasets.

**For blood, only the top ten (based on AUC) published blood diagnostic signatures are shown.*

GSVA= gene set variation analysis; ssGSEA= single-sample gene set enrichment analysis.

457 **Discussion**

458 In a cohort of treatment-naive PLHIV with presumptive TB in Uganda, we identified 44 DEGs in nasal
459 cells between adults with microbiologically confirmed TB disease compared with individuals without
460 TB. These 44 DEGs yielded a promising AUC of > 0.87 in predicting TB across six models in cross-
461 validation within our dataset. Amongst those with TB disease, we also found a blunted inflammatory
462 nasal response, with downregulation of several innate immune pathways. Compartmental
463 differences were seen, with contrasting patterns of inflammation in TB between the nose and blood
464 and little overlap in nasal and blood DEGs. Our proof-of-concept study highlights that there are
465 differences in nasal gene expression associated with TB disease that are distinct from blood.
466
467 There are no published nasal transcriptomic studies in TB; however, half of the top ten nasal DEGs
468 have been associated with TB in blood. *SPIB,* which ranked amongst the most important predictive
469 genes across all models, is important for plasmacytoid dendritic cell function[47] and memory B cell
470 maintenance[48]. *SPIB* downregulation was noted in active TB patients in multiple studies[49-51]. *CLC,*
471 which is linked to neutrophil recruitment[52], was part of a four-marker blood signature for
472 differentiating adults with TB and non-TB febrile infection[53]. *ITLN,* which has a role in pathogen
473 recognition, was downregulated in newly diagnosed adults with TB, increasing in expression with
474 treatment[49]. Changes in expression of T cell markers *PTGDR2* and *CD69* have also been associated
475 with early signs of pulmonary TB[54] and TB infection, respectively[55]. Top DEGs not previously linked to
476 TB include *CEACAM21*, which facilitates opsonin-independent recognition and clearance of specific
477 gram-negative bacterial pathogens[56]. *CST1, CLC, FETUB* and *VSMT1* were upregulated in asthma and
478 allergic diseases in nasal studies[57-59], which could reflect overrepresentation of allergic diseases in
479 our non-TB control group.
480
481 GSEA, ORA and IPA suggested that TB was associated with suppressed nasal innate immune
482 responses. IFN-γ activity was the top enriched downregulated pathway from the Hallmark gene-sets

483 using fGSEA . IFN-γ is an important cytokine for immunity against mycobacteria. It facilitates the host
484 defense to *Mtb* through stimulating macrophage activity and producing reactive nitrogen species[60],
485 with ineffective IFN-γ activity associated with disseminated mycobacterial infection[61]. Neutrophils
486 were the top downregulated REACTOME enriched pathway. Whilst excessive neutrophil activation
487 can cause tissue damage, they have a protective role in early stages[62] and adult contacts with higher
488 blood neutrophil counts were shown to be less likely to develop TB infection[63]. Chemokine signaling
489 was the most enriched downregulated KEGG pathway. Single nucleotide polymorphisms in
490 chemokines involved in cell migration in TB, were shown to increase susceptibility to active TB[64,65].
491 The nose is the first line of defense for airborne pathogens and early control of respiratory infection
492 is dependent on robust nasal mucosal immune responses[66]. Dampening of IFN-γ activity,
493 neutrophils, and chemokines in addition to other enriched nasal pathways like complement,
494 phagocytosis, and toll-like receptor signaling, may underlie failure of the early protective nasal
495 response to *Mtb* in TB disease.
496
497 The nasal host response in TB differed from blood. Few published TB blood signatures could classify
498 TB in our nasal sequencing dataset. Moreover, downregulated nasal innate immune responses in TB
499 were accompanied by upregulation of inflammatory pathways in blood. Unlike blood, nasal samples
500 have more heterogeneity in cell composition, consisting of both immune and epithelial
501 components[14]. We could not show that nasal immune pathways were driven by immune cell
502 composition. However, neutrophils in the blood were higher in PLHIV with TB versus those without
503 TB, consistent with our observation of increased neutrophil enrichment in blood.
504
505 Our finding of upregulated innate immune pathways in TB blood signatures confirms previously
506 published studies in PLHIV[10,67,68]. Similar to the downregulation of immune nasal pathways in TB in
507 our study, Mick and colleagues found that toll-like receptor signaling, inflammasomes, neutrophil
508 degranulation and chemokine signaling were attenuated in the nasopharyngeal response in SARS-

509 CoV-2 patients compared to other viral infections[16]. Hurst and colleagues also noted that systemic
510 symptoms in SARS-CoV-2-infected individuals were correlated with suppressed nasal innate and
511 adaptive immune pathways and upregulation of these same pathways in matched blood samples[69].
512 In contrast, two studies comparing nasal and blood samples in children with viral infections,
513 identified upregulation in immune activity and strong innate responses in both compartments[14,15].
514 Although our cross-sectional study cannot infer causality, we can hypothesize that reduced mucosal
515 immune responses and inadequate control of *Mtb* in the nose could promote progression to lung
516 disease and consequent systemic inflammation. Longitudinal studies using paired nasal and blood
517 samples are needed to understand the temporal relationship between the nasal and blood immune
518 response and TB disease.

519 A streamlined nasal gene signature based on variable importance (NASAL_4) met the WHO criteria
520 for a TB triage test (>90% sensitivity and 70% specificity) in cross-validation. Three genes within
521 NASAL_4 relate to immune activity: *SPIB* is linked to interferon production and proinflammatory
522 cytokines[47], *CD1b* is associated with antigen presentation[70] and *TESPA* relates to T cell maturation[71].
523 Whilst the AUC for nasal models based on all DEGs was slightly lower than blood, most of the top ten
524 nasal DEGs were not ISGs, unlike blood. This is promising given that the performance of many ISG-
525 dominant RNA blood signatures in differentiating TB from viral infections or in PLHIV has been
526 inadequate[8,10]. Moreover, adults with TB missed by blood models were classified correctly by nasal
527 models and vice versa, suggesting potential value of combined nasal-blood diagnostic biomarkers.

528 All adults in our study had HIV. Studies of the nasal transcriptome in HIV are limited, however HIV
529 causes mucosal immune dysfunction[72], which will shape nasal transcriptional changes,
530 independently of TB. Based on flow cytometry and single cell sequencing, Phiri and colleagues found
531 increased nasal neutrophilic abundance and inflammation in PLHIV compared with non-infected
532 individuals and that nasal neutrophils had transcriptional signatures of stress and senescence and
533 nasal T cells had signatures of exhaustion and apoptosis[73]. Differing numbers of nasal CD3 and CD4

534 cells based on HIV status have also been noted[74]. Further studies looking at the nasal TB response
535 among individuals without HIV would help explore the extent to which nasal gene profiles in our
536 study are driven by HIV.

537 Our work has limitations. The reliance on sputum Ultra as the TB reference standard in PLHIV means
538 paucibacillary TB disease could be missed in the control group. We could not follow up controls to
539 identify later development of TB disease and were unable to evaluate the role of nasal signatures in
540 predicting TB progression. Secondly, our ART-naive cohort was at high risk of opportunistic infections
541 and nasopharyngeal bacterial colonization. However, we were unable to ascertain and adjust for
542 these as well as any alternative diagnoses in non-TB controls that would affect gene expression. Case
543 numbers were too small to study the impact of CD4 count on nasal signatures, and other covariates
544 that impact gene expression, like malnutrition and TB infection via TST or IGRA status, were
545 unavailable. Finally, nasal DEG-based classifiers were trained and evaluated on the same dataset
546 since the sample size was too small to allocate a separate test set, and there are no other published
547 nasal transcriptomic datasets for TB. Although we employed k-fold cross-validation to reduce
548 overfitting, the need for external validation of nasal signatures on independent datasets remains.
549

550 Despite these limitations, these are the first data interrogating the nasal transcriptome in TB disease.
551 Whilst developing a nasal diagnostic test will require a more accessible technology than sequencing,
552 detecting gene expression differences is the first step. Our focus on ART-naïve PLHIV who have an
553 elevated risk of TB addresses a major gap in HIV-TB studies and by only studying symptomatic adults,
554 we remove the bias of using healthy controls. Although outside the scope of this work, sequencing
555 nasal samples enable the concurrent study of the respiratory microbiome and opportunity to
556 develop integrated host-microbiome diagnostic biomarkers. Although we focused on adults
557 diagnosed with Ultra on expectorated sputum as our microbiological reference standard, a nasal
558 biomarker could still have additional value for TB diagnosis given the ease of sampling compared

with sputum. Future work to confirm the utility of nasal signatures in TB requires discovery and external validation of nasal biomarkers across larger, independent, geographically distinct cohorts, that include adults with and without HIV across the spectrum of TB disease, as well as other difficult to diagnose groups, such as young children or those without microbiologically confirmed TB.

Conclusion

In conclusion, our exploratory study using minimally invasive nasal samples suggests that host nasal transcriptomics is feasible and informative for the diagnosis of TB, whereby elucidating gene expression profiles at the initial pathogen encounter reveals novel insights into host-TB interaction. We found distinct nasal gene expression patterns associated with TB disease in PLHIV that were not captured in blood, with DEGs able to distinguish TB from non-TB controls. Our findings support further study of the nose for non-sputum-based TB diagnostic biomarkers, which is an urgent research priority to reduce global TB-associated mortality.

Additional tables greater than one A4 in length

Table 2: The top ten enriched pathways in the Hallmark gene sets, KEGG and Reactome pathways based on fGSEA in nasal and blood samples

NASAL			BLOOD		
	Adjusted p-value	**Normalized enrichment score**		**Adjusted p-value**	**Normalized enrichment score**
Hallmark gene set					
*interferon gamma response**	4.58E-33	-3.1391781	heme metabolism	1.44E-17	2.69075782
*interferon alpha response**	1.51E-23	-3.1566226	*interferon gamma response**	1.95E-17	2.65410796
allograft rejection	8.15E-20	-2.7435101	EF2 targets	1.18E-12	-2.504435
*inflammatory response**	5.64E-19	-2.667695	*complement**	4.29E-11	2.33214589
*TNF-A signaling via NF-kB**	1.45E-12	-2.3558235	*IL6 JAK STAT3 signaling**	1.47E-09	2.50731015
IL2 STAT5 signaling	1.49E-10	-2.2420619	MYC targets	1.47E-09	-2.2858968
*complement**	1.00E-08	-2.1158445	*inflammatory response**	1.71E-08	2.15816136
*IL6 JAK STAT3 signaling**	1.52E-08	-2.3716069	hypoxia	5.38E-08	2.11394868
bile acid metabolism	1.77E-05	2.04062867	*interferon alpha response**	5.38E-08	2.34472265
cholesterol homeostasis	4.66E-05	2.04477918	*TNF-A signaling via NF-kB**	7.37E-08	2.10746927
REACTOME pathway					
*neutrophil degranulation**	1.22E-20	-2.3753505	*neutrophil degranulation**	4.60E-48	3.03136454
cilium assembly	2.14E-19	2.82675326	platelet activation signaling and aggregation	1.65E-09	2.24160937
infraglabellar transport	5.73E-18	3.28715505	*signaling by interleukins**	3.23E-09	1.99216824
organelle biogenesis and maintenance	2.01E-16	2.50485971	rrna processing	5.40E-09	-2.5587675
interferon alpha beta signalling	9.70E-14	-2.878876	antigen processing cross presentation	1.41E-07	2.36613708
immunoregulatory interactions between lymphoid and a non-lymphoid cell	1.25E-13	-2.6368349	interferon gamma signaling	1.41E-07	2.41404798
*signaling by interleukins**	1.07E-12	-2.0969441	trna processing	9.24E-07	-2.335618
interferon signaling	1.34E-11	-2.262785	rrna modification in the nucleus and cytosol	1.10E-06	-2.5765856
leishmania infection	3.79E-09	-2.2903078	response to elevated platelet cytosolic ca2	2.32E-06	2.18883777
cholesterol biosynthesis	2.46E-08	2.8140458	diseases of signal transduction by growth factor receptors and second messengers	3.68E-06	1.80933376
KEGG pathway					
chemokine signaling pathway	1.24E-13	-2.5264553	regulation of actin cytoskeleton	9.20E-08	2.15502842
*fc gamma r mediated phagocytosis**	9.01E-10	-2.4900559	leukocyte transendothelial migration	7.01E-06	2.20647871
cytokine cytokine receptor interaction	9.21E-09	-2.1236459	*fc gamma r mediated phagocytosis**	1.67E-05	2.17411173
natural killer cell mediated cytotoxicity	9.21E-09	-2.3593142	lysosome	2.77E-05	2.05385563
*toll like receptor signaling pathway**	9.21E-09	-2.4708031	focal adhesion	3.76E-05	1.92835864
*leishmania infection**	2.48E-07	-2.3806651	*leishmania infection**	6.30E-05	2.13257101

JAK STAT signaling pathway	1.45E-06	-2.1495304	acute myeloid leukemia	8.44E-05	2.17502533
B cell receptor signaling pathway	1.96E-06	-2.2891477	complement and coagulation cascades	8.44E-05	2.2165221
T cell receptor signaling pathway	1.96E-06	-2.1668356	neurotrophin signaling pathway	8.44E-05	1.9836064
MAPK signaling pathway	2.04E-06	-1.8908071	*toll like receptor signaling pathway**	8.86E-05	2.05311921

Negative normalized enrichment scores indicate the pathway is downregulated in TB compared to non-TB controls, and positive enrichment scores indicate the pathway is upregulated in TB.

Enriched pathways which are shared in both nasal and blood samples are italicized and have an asterisk (*).

List of abbreviations

ART: antiretroviral therapy

AUC: area under the curve

DEG: differentially expressed genes

GSEA: gene set enrichment analysis

GSVA: gene set variation analysis

IPA: ingenuity pathway analysis

KNN: k nearest neighbors

ORA: over representation analysis

PCA: principal component analysis

PLHIV: people living with HIV

PLS: partial least squares

RF: random forest

ROC: receiver operating characteristic curve

SVM: support vector machine

ssGSEA: single sample gene set enrichment analysis

TB : tuberculosis

WHO: world health organization

Declarations

Ethics approval and consent to participate

This study used human data. Informed consent was obtained from all participants. The study protocol and informed consent forms was approved, with ethics approval given by institutional review boards of all partnering institutions (Boston University Medical Campus, Joint Clinical Research Centre and Uganda National Council for Science and Technology). This study did not use any animal data.

Consent for publication

Not applicable

Availability of data and materials

The datasets analyzed during the current study are available in the SRA repository (PRJNA1379661) on https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1379661

Computational code and datasets are also available at: https://github.com/nisreenkhambati/uganda_nasal_cells.

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by grants from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health: UO1AI52084 (JE) , R01AI152159 (PS) , R21AI154387 (WEJ), R01GM127430 (WEJ), R01AI175315 (WEJ) and the PROVIDENCE/BOSTON CFAR (P30AI042853).

Author contributions

Conceptualization: JE, LN, WEJ, SG

Methodology: NK, KLT, LN, WEJ

Resources: LN, WEJ, PS

Project administration: LN, WEJ

Investigation: EN, LN, KLT, NK

Software: NK

Data curation: EN, NK

Formal analysis: NK (except IPA done by AVV)

Validation: KLT, SL, JKA, WEJ

Writing–original draft: NK

Writing –review of final manuscript: all authors

Visualisation- NK

Supervision: RS, AJP, EMB, DOC, WEJ

Funding acquisition: JE, LN, WEJ, SG

Acknowledgements

We would like to thank all the participants in Uganda who were part of this study, as well as all local clinical, lab and data teams involved.

References

1. Meintjes G, Maartens G. HIV-Associated Tuberculosis. N Engl J Med. 2024;391(4):343-55. doi:10.1056/NEJMra2308181
2. The World Health Organization. Global tuberculosis report 2024. Geneva: The WHO; 2024.
3. Gupta RK, Lucas SB, Fielding KL, Lawn SD. Prevalence of tuberculosis in post-mortem studies of HIV-infected adults and children in resource-limited settings: a systematic review and meta-analysis. AIDS. 2015;29(15):1987-2002. doi:10.1097/QAD.0000000000000802
4. The World Health Organization. High-priority target product profiles or new tuberculosis diagnostics: report of a consensus meeting. Geneva; 2014.
5. Warsinske H, Vashisht R, Khatri P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. PLoS Med. 2019;16(4):e1002786. doi:10.1371/journal.pmed.1002786
6. Muwanga VM, Mendelsohn SC, Leukes V, Stanley K, Mbandi SK, Erasmus M, et al. Blood transcriptomic signatures for symptomatic tuberculosis in an African multicohort study. Eur Respir J. 2024;64(2). doi:10.1183/13993003.00153-2024
7. Gupta-Wright A, Ha H, Abdulgadar S, Crowder R, Emmanuel J, Mukwatamundu J, et al. Evaluation of the Xpert MTB Host Response assay for the triage of patients with presumed pulmonary tuberculosis: a prospective diagnostic accuracy study in Viet Nam, India, the Philippines, Uganda, and South Africa. Lancet Glob Health. 2024;12(2):e226-e34. doi:10.1016/S2214-109X(23)00541-7
8. Mann T, Gupta RK, Reeve BWP, Ndlangalavu G, Chandran A, Krishna AP, et al. Blood RNA biomarkers for tuberculosis screening in people living with HIV before antiretroviral therapy initiation: a diagnostic accuracy study. Lancet Glob Health. 2024;12(5):e783-e92. doi:10.1016/S2214-109X(24)00029-9
9. Mendelsohn SC, Verhage S, Mulenga H, Scriba TJ, Hatherill M. Systematic review of diagnostic and prognostic host blood transcriptomic signatures of tuberculosis disease in people living with HIV. Gates Open Res. 2023;7:27. doi:10.12688/gatesopenres.14327.2
10. Esmail H, Lai RP, Lesosky M, Wilkinson KA, Graham CM, Horswell S, et al. Complement pathway gene activation and rising circulating immune complexes characterize early disease in HIV-associated tuberculosis. Proc Natl Acad Sci U S A. 2018;115(5):E964-E73. doi:10.1073/pnas.1711853115
11. Kicic A, de Jong E, Ling KM, Nichol K, Anderson D, Wark PAB, et al. Assessing the unified airway hypothesis in children via transcriptional profiling of the airway epithelium. J Allergy Clin Immunol. 2020;145(6):1562-73. doi:10.1016/j.jaci.2020.02.018
12. Perez Rogers JF. Shared Gene Expression Alterations in Nasal and Bronchial Epithelium for Lung Cancer Detection. J Natl Cancer Inst. 2017;109(7). doi:10.1093/jnci/djw327
13. Gower AC, Steiling K, Brothers JF, 2nd, Lenburg ME, Spira A. Transcriptomic studies of the airway field of injury associated with smoking-related lung disease. Proc Am Thorac Soc. 2011;8(2):173-9. doi:10.1513/pats.201011-066MS
14. Yu J, Peterson DR, Baran AM, Bhattacharya S, Wylie TN, Falsey AR, et al. Host Gene Expression in Nose and Blood for the Diagnosis of Viral Respiratory Infection. J Infect Dis. 2019;219(7):1151-61. doi:10.1093/infdis/jiy608
15. Do LAH, Pellet J, van Doorn HR, Tran AT, Nguyen BH, Tran TTL, et al. Host Transcription Profile in Nasal Epithelium and Whole Blood of Hospitalized Children Under 2 Years of Age With Respiratory Syncytial Virus Infection. J Infect Dis. 2017;217(1):134-46. doi:10.1093/infdis/jix519
16. Mick E, Kamm J, Pisco AO, Ratnasiri K, Babik JM, Castaneda G, et al. Upper airway gene expression reveals suppressed immune responses to SARS-CoV-2 compared with other respiratory viruses. Nat Commun. 2020;11(1):5854. doi:10.1038/s41467-020-19587-y

17. Ziegler CGK, Miao VN, Owings AH, Navia AW, Tang Y, Bromley JD, et al. Impaired local intrinsic immunity to SARS-CoV-2 infection in severe COVID-19. Cell. 2021;184(18):4713-33 e22. doi:10.1016/j.cell.2021.07.023
18. Hurst JH, Mohan AA, Dalapati T, George IA, Aquino JN, Lugo DJ, et al. Differential host responses within the upper respiratory tract and peripheral blood of children and adults with SARS-CoV-2 infection. medRxiv. 2023:2023.07.31.23293337. doi:10.1101/2023.07.31.23293337
19. Jochems SP, Marcon F, Carniel BF, Holloway M, Mitsi E, Smith E, et al. Inflammation induced by influenza virus impairs human innate immune control of pneumococcus. Nat Immunol. 2018;19(12):1299-308. doi:10.1038/s41590-018-0231-y
20. Castro-Nallar E, Shen Y, Freishtat RJ, Perez-Losada M, Manimaran S, Liu G, et al. Integrating microbial and host transcriptomics to characterize asthma-associated microbial communities. BMC Med Genomics. 2015;8:50. doi:10.1186/s12920-015-0121-1
21. Pandya R, He YD, Sweeney TE, Hasin-Brumshtein Y, Khatri P. A machine learning classifier using 33 host immune response mRNAs accurately distinguishes viral and non-viral acute respiratory illnesses in nasal swab samples. Genome Med. 2023;15(1):64. doi:10.1186/s13073-023-01216-0
22. Yahya M, Rulli M, Toivonen L, Waris M, Peltola V. Detection of Host Response to Viral Respiratory Infection by Measurement of Messenger RNA for MxA, TRIM21, and Viperin in Nasal Swabs. J Infect Dis. 2017;216(9):1099-103. doi:10.1093/infdis/jix458
23. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data 2010 [Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed: 6th January 2026
24. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. doi:10.1093/bioinformatics/btu170
25. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635
26. GENCODE. Human Release 44 (GRCh38.p14) 2022 [Available from: https://www.gencodegenes.org/human/release_44.html. Accessed: 6th January 2026
27. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184-91. doi:10.1038/nprot.2009.97
28. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847-9. doi:10.1093/bioinformatics/btw313
29. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi:10.1186/s13059-014-0550-8
30. Korotkevich G, Sukhov V, Sergushichev A. Fast gene set enrichment analysis. bioRxiv. 2019.
31. Kuhn M. Building Predictive Models in R Using the caret Package. Journal of Statistical Software. 2008;28(5):1 - 26. doi:10.18637/jss.v028.i05
32. Channon-Wells S, Habgood-Coote D, Vito O, Galassini R, Wright VJ, Brent AJ, et al. Integration and validation of host transcript signatures, including a novel 3-transcript tuberculosis signature, to enable one-step multiclass diagnosis of childhood febrile disease. J Transl Med. 2024;22(1):802. doi:10.1186/s12967-024-05241-4
33. Bobak CA, Titus AJ, Hill JE. Comparison of common machine learning models for classification of tuberculosis using transcriptional biomarkers from integrated datasets. Applied Soft Computing. 2019;74:264-73. doi:https://doi.org/10.1016/j.asoc.2018.10.005
34. Anderson ST, Kaforou M, Brent AJ, Wright VJ, Banwell CM, Chagaluka G, et al. Diagnosis of childhood tuberculosis and host RNA expression in Africa. N Engl J Med. 2014;370(18):1712-23. doi:10.1056/NEJMoa1303657
35. Ding S, Yi X, Gao J, Huang C, Zhou Y, Yang Y, et al. Combining bioinformatics and machine learning to identify diagnostic biomarkers of TB associated with immune cell infiltration. Tuberculosis (Edinb). 2024;149:102570. doi:10.1016/j.tube.2024.102570

36. Turner CT, Gupta RK, Tsaliki E, Roe JK, Mondal P, Nyawo GR, et al. Blood transcriptional biomarkers for active pulmonary tuberculosis in a high-burden setting: a prospective, observational, diagnostic accuracy study. Lancet Respir Med. 2020;8(4):407-19. doi:10.1016/S2213-2600(19)30469-2

37. John CR. MLeval: Machine Learning Model Evaluation 2020 [Computer software version 0.3]. Available from: https://cran.r-project.org/web/packages/MLeval/index.html. Accessed: 10th January 2026

38. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biom J. 2005;47(4):458-72. doi:10.1002/bimj.200410135

39. Rusinova I, Forster S, Yu S, Kannan A, Masse M, Cumming H, et al. Interferome v2.0: an updated database of annotated interferon-regulated genes. Nucleic Acids Res. 2013;41(Database issue):D1040-6. doi:10.1093/nar/gks1215

40. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021;2(3):100141. doi:10.1016/j.xinn.2021.100141

41. Broad Institute. Human MSigDB Collections: Broad Institute, Inc., Massachusetts Institute of Technology, and Regents of the University of California; 2023 [Available from: https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp. Accessed: 6th January 2026

42. Gene Ontology C, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023;224(1). doi:10.1093/genetics/iyad031

43. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014;30(4):523-30. doi:10.1093/bioinformatics/btt703

44. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453-7. doi:10.1038/nmeth.3337

45. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37(7):773-82. doi:10.1038/s41587-019-0114-2

46. Johnson WE, Odom A, Cintron C, Muthaiah M, Knudsen S, Joseph N, et al. Comparing tuberculosis gene signatures in malnourished individuals using the TBSignatureProfiler. BMC Infect Dis. 2021;21(1):106. doi:10.1186/s12879-020-05598-z

47. Sasaki I, Hoshino K, Sugiyama T, Yamazaki C, Yano T, Iizuka A, et al. Spi-B is critical for plasmacytoid dendritic cell function and development. Blood. 2012;120(24):4733-43. doi:10.1182/blood-2012-06-436527

48. Horiuchi S, Koike T, Takebuchi H, Hoshino K, Sasaki I, Fukuda-Ohta Y, et al. SpiB regulates the expression of B-cell-related genes and increases the longevity of memory B cells. Front Immunol. 2023;14:1250719. doi:10.3389/fimmu.2023.1250719

49. Cliff JM, Lee JS, Constantinou N, Cho JE, Clark TG, Ronacher K, et al. Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. J Infect Dis. 2013;207(1):18-29. doi:10.1093/infdis/jis499

50. Bah SY, Forster T, Dickinson P, Kampmann B, Ghazal P. Meta-Analysis Identification of Highly Robust and Differential Immune-Metabolic Signatures of Systemic Host Response to Acute and Latent Tuberculosis in Children and Adults. Front Genet. 2018;9:457. doi:10.3389/fgene.2018.00457

51. Natarajan S, Ranganathan M, Hanna LE, Tripathy S. Transcriptional Profiling and Deriving a Seven-Gene Signature That Discriminates Active and Latent Tuberculosis: An Integrative Bioinformatics Approach. Genes (Basel). 2022;13(4). doi:10.3390/genes13040616

52. Gevaert E, Delemarre T, De Volder J, Zhang N, Holtappels G, De Ruyck N, et al. Charcot-Leyden crystals promote neutrophilic inflammation in patients with nasal polyposis. J Allergy Clin Immunol. 2020;145(1):427-30.e4. doi:10.1016/j.jaci.2019.08.027

53. Roe JK, Thomas N, Gil E, Best K, Tsaliki E, Morris-Jones S, et al. Blood transcriptomic diagnosis of pulmonary and extrapulmonary tuberculosis. JCI Insight. 2016;1(16):e87238. doi:10.1172/jci.insight.87238

54. Shaukat SN, Eugenin E, Nasir F, Khanani R, Kazmi SU. Identification of immune biomarkers in recent active pulmonary tuberculosis. Sci Rep. 2023;13(1):11481. doi:10.1038/s41598-023-38372-7
55. Chen ZY, Wang L, Gu L, Qu R, Lowrie DB, Hu Z, et al. Decreased Expression of CD69 on T Cells in Tuberculosis Infection Resisters. Front Microbiol. 2020;11:1901. doi:10.3389/fmicb.2020.01901
56. Pils S, Gerrard DT, Meyer A, Hauck CR. CEACAM3: an innate immune receptor directed against human-restricted bacterial pathogens. Int J Med Microbiol. 2008;298(7-8):553-60. doi:10.1016/j.ijmm.2008.04.005
57. Xu Z, Forno E, Sun Y, Manni ML, Han YY, Kim S, et al. Nasal epithelial gene expression and total IgE in children and adolescents with asthma. J Allergy Clin Immunol. 2024;153(1):122-31. doi:10.1016/j.jaci.2023.09.014
58. Klingler AI, Stevens WW, Tan BK, Peters AT, Poposki JA, Grammer LC, et al. Mechanisms and biomarkers of inflammatory endotypes in chronic rhinosinusitis without nasal polyps. J Allergy Clin Immunol. 2021;147(4):1306-17. doi:10.1016/j.jaci.2020.11.037
59. Tomizawa H, Yamada Y, Arima M, Miyabe Y, Fukuchi M, Hikichi H, et al. Galectin-10 as a Potential Biomarker for Eosinophilic Diseases. Biomolecules. 2022;12(10). doi:10.3390/biom12101385
60. Shanmuganathan G, Orujyan D, Narinyan W, Poladian N, Dhama S, Parthasarathy A, et al. Role of Interferons in Mycobacterium tuberculosis Infection. Clin Pract. 2022;12(5):788-96. doi:10.3390/clinpract12050082
61. Dorman SE, Holland SM. Mutation in the signal-transducing chain of the interferon-gamma receptor and susceptibility to mycobacterial infection. J Clin Invest. 1998;101(11):2364-9. doi:10.1172/JCI2901
62. de Waal AM, Hiemstra PS, Ottenhoff TH, Joosten SA, van der Does AM. Lung epithelial cells interact with immune cells and bacteria to shape the microenvironment in tuberculosis. Thorax. 2022;77(4):408-16. doi:10.1136/thoraxjnl-2021-217997
63. Martineau AR, Newton SM, Wilkinson KA, Kampmann B, Hall BM, Nawroly N, et al. Neutrophil-mediated innate immune resistance to mycobacteria. J Clin Invest. 2007;117(7):1988-94. doi:10.1172/JCI31097
64. Feng WX, Flores-Villanueva PO, Mokrousov I, Wu XR, Xiao J, Jiao WW, et al. CCL2-2518 (A/G) polymorphisms and tuberculosis susceptibility: a meta-analysis. Int J Tuberc Lung Dis. 2012;16(2):150-6. doi:10.5588/ijtld.11.0205
65. Sanchez-Castanon M, Baquero IC, Sanchez-Velasco P, Farinas MC, Ausin F, Leyva-Cobian F, et al. Polymorphisms in CCL5 promoter are associated with pulmonary tuberculosis in northern Spain. Int J Tuberc Lung Dis. 2009;13(4):480-5.
66. Zhou X, Wu Y, Zhu Z, Lu C, Zhang C, Zeng L, et al. Mucosal immune response in biology, disease prevention and treatment. Signal Transduct Target Ther. 2025;10(1):7. doi:10.1038/s41392-024-02043-4
67. Verma S, Du P, Nakanjako D, Hermans S, Briggs J, Nakiyingi L, et al. "Tuberculosis in advanced HIV infection is associated with increased expression of IFNγ and its downstream targets". BMC Infect Dis. 2018;18(1):220. doi:10.1186/s12879-018-3127-4
68. Kulkarni V, Queiroz ATL, Sangle S, Kagal A, Salvi S, Gupta A, et al. A Two-Gene Signature for Tuberculosis Diagnosis in Persons With Advanced HIV. Front Immunol. 2021;12:631165. doi:10.3389/fimmu.2021.631165
69. Hurst JH, McCumber AW, Aquino JN, Rodriguez J, Heston SM, Lugo DJ, et al. Age-Related Changes in the Nasopharyngeal Microbiome Are Associated With Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection and Symptoms Among Children, Adolescents, and Young Adults. Clin Infect Dis. 2022;75(1):e928-e37. doi:10.1093/cid/ciac184
70. Shahine A. The intricacies of self-lipid antigen presentation by CD1b. Mol Immunol. 2018;104:27-36. doi:10.1016/j.molimm.2018.09.022

71. Wang D, Zheng M, Lei L, Ji J, Yao Y, Qiu Y, et al. Tespa1 is involved in late thymocyte development through the regulation of TCR-mediated signaling. Nat Immunol. 2012;13(6):560-8. doi:10.1038/ni.2301

72. Paiardini M, Frank I, Pandrea I, Apetrei C, Silvestri G. Mucosal immune dysfunction in AIDS pathogenesis. AIDS Rev. 2008;10(1):36-46.

73. Phiri JA, Sibale LL, Kapira G, Mlogoti L, Nyirenda JT, Mitole N, et al. Persistent pneumococcal colonisation in antiretroviral-treated HIV infection is associated with nasal inflammation. Nat Commun. 2025. doi:10.1038/s41467-025-67258-7

74. Moss RB, Scott TA, Goldrich M, Pitale M, Daniel A, Lebovics R, et al. Nasal mucosal cells in human immunodeficiency virus type 1-seropositive patients with sinusitis. J Clin Lab Anal. 1996;10(6):418-22. doi:10.1002/(SICI)1098-2825(1996)10:6<418::AID-JCLA18>3.0.CO;2-#