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 tuberculosis (TB) in people with HIV (PWH) remains difficult. Since the first pathogen-host interaction in TB occurs in the upper airway, host transcriptomic analysis on nasal specimens may identify novel diagnostic biomarkers. We aimed to demonstrate differences in nasal gene expression in PWH and TB disease versus PWH without TB, evaluate the performance of nasal signatures in predicting TB and compare nasal gene profiles with blood.

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 (DEGs) and enriched pathways between PWH and TB disease and PWH without TB. Supervised machine-learning of gene expression data was used to predict TB.

Results: 40 PWH were enrolled (median age: 34 years, median CD4 count: 182), including 20 with TB disease and 20 without. We identified 44 nasal DEGs and 238 blood DEGs, with three overlapping DEGs between samples. Models trained using all 44 nasal DEGs had a cross-validated area under the curve between 0.87-0.90 for predicting TB disease. A simplified signature (*SPIB, SHISA2, TESPA1* and *CD1B*) met WHO criteria for a TB triage test. Among adults with TB, pathways related to the inflammatory response were downregulated in nasal samples and upregulated in blood.

Conclusion: There are distinct nasal gene expression patterns associated with TB, not seen in blood. Differences in nasal gene expression in PWH who have TB disease, versus those without TB, highlight their potential as diagnostic biomarkers.

Manuscript

Introduction

Tuberculosis (TB) is the primary reason for hospitalization and death in people with HIV (PWH)[1], with the highest TB-HIV coinfection in the African Region[2]. TB-related mortality in PWH is partly due to difficulties in diagnosis. Autopsy studies in PWH demonstrated that TB was undiagnosed in nearly half of cases identified post-mortem[3]. PWH 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[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 PWH[6,7], particularly in antiretroviral therapy (ART) naive individuals[8], or when differentiating TB from other respiratory diseases[9]. Lower specificity in PWH 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. 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 of non-TB respiratory illnesses, including respiratory syncytial virus[11,12], SARS-CoV-2[13-15], pneumonia[16] and asthma[17]. 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[18]. Amongst SARS-CoV-2 patients, suppressed interferon responses in nasal samples were linked to more severe disease[14]. In children, nasal gene signatures could accurately predict respiratory viral infections[11,19], and in some cases, performed better than blood classifiers[11]. 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 PWH and pulmonary TB disease compared with PWH 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, pathways and signatures with paired blood samples and evaluated the diagnostic performance of existing blood TB signatures in our nasal dataset.

Methods

Figure 1 and the text below summarize the methods. A detailed description of methods and bioinformatic packages are described in the Supplementary Methods.

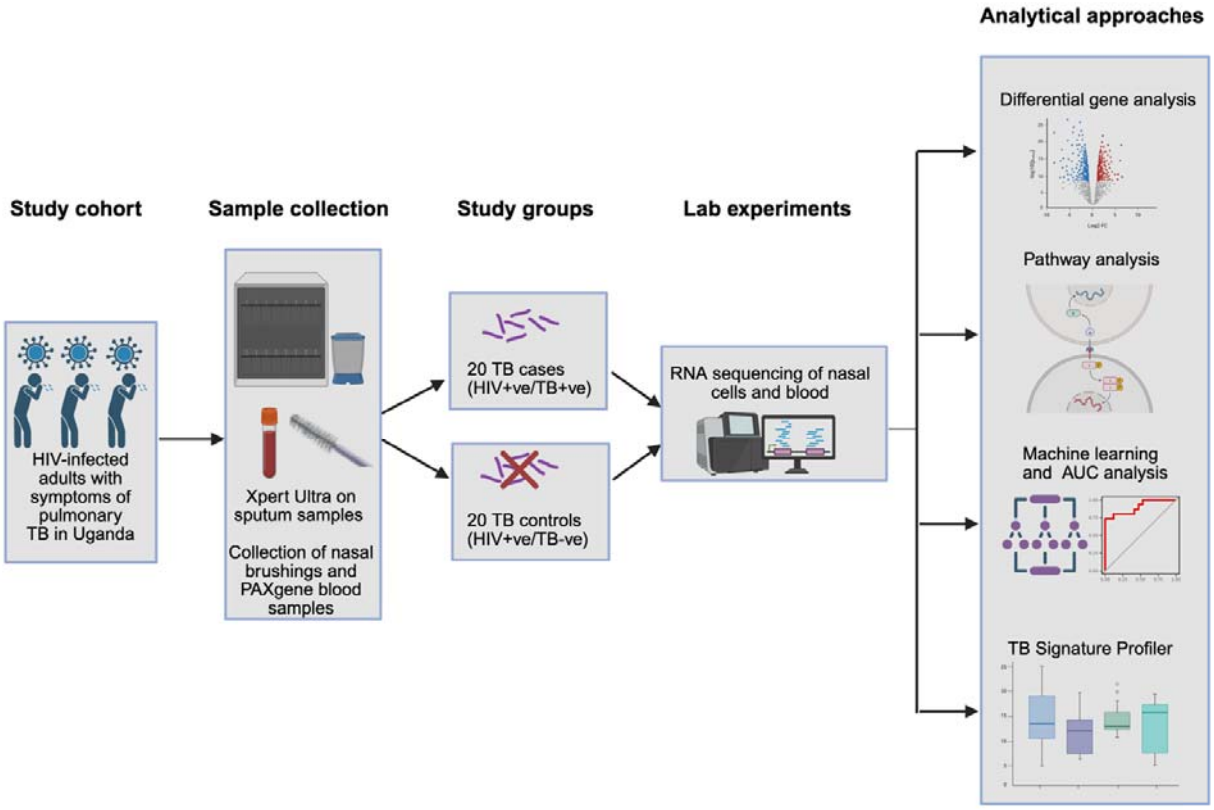

Figure 1: Overview of the methods. Adults in Uganda with newly diagnosed HIV and TB symptoms were enrolled and nasal and blood samples were collected for RNA sequencing. Analysis aimed to determine if nasal gene expression differed in PWH and pulmonary TB disease compared with PWH without TB disease. 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; PWH=people with HIV

Study cohort

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

Definitions and sample collection

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

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

RNA sequencing and quality control

Nasal and blood samples were shipped on dry ice to Rutgers University for sample processing and RNA sequencing. Following alignment of reads to the human genome (GRCh38.p14), sample outliers were identified by examining the Euclidean distances between pairwise sample combinations and using principal component analysis (PCA). Additional information is in Supplementary Methods.

Differential gene expression, machine learning and enrichment analysis

Differential gene expression analysis between PWH and TB disease and PWH without TB disease (controls) was performed in the nasal and blood datasets. Models were trained based on all DEGs using 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. A four-gene nasal signature was also derived based on the variable importance of nasal models. Receiver operating characteristic (ROC) curves for predicting TB status were constructed to estimate the area under the curve (AUC), sensitivity and specificity. The INTERFEROME database (https://interferome.org/interferome/) was used to identify which top DEGs were ISGs.

Functional analyses was performed using fast gene set enrichment analysis (fGSEA), over-representation analysis (ORA) and Ingenuity Pathway Analysis (IPA) (QIAGEN Inc., https://digitalinsights.qiagen.com/IPA)[20]. We also examined immune cell composition with CIBERSORTX and the LM22 signature matrix[21]. See Supplementary Methods.

Performance of published blood RNA signatures

The TBSignatureProfiler package[22] was used to evaluate the performance of previously published diagnostic TB blood signatures in our nasal and blood datasets. 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. See Supplementary Methods.

Results

Study participants

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

Quality control and PCA

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

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

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).**

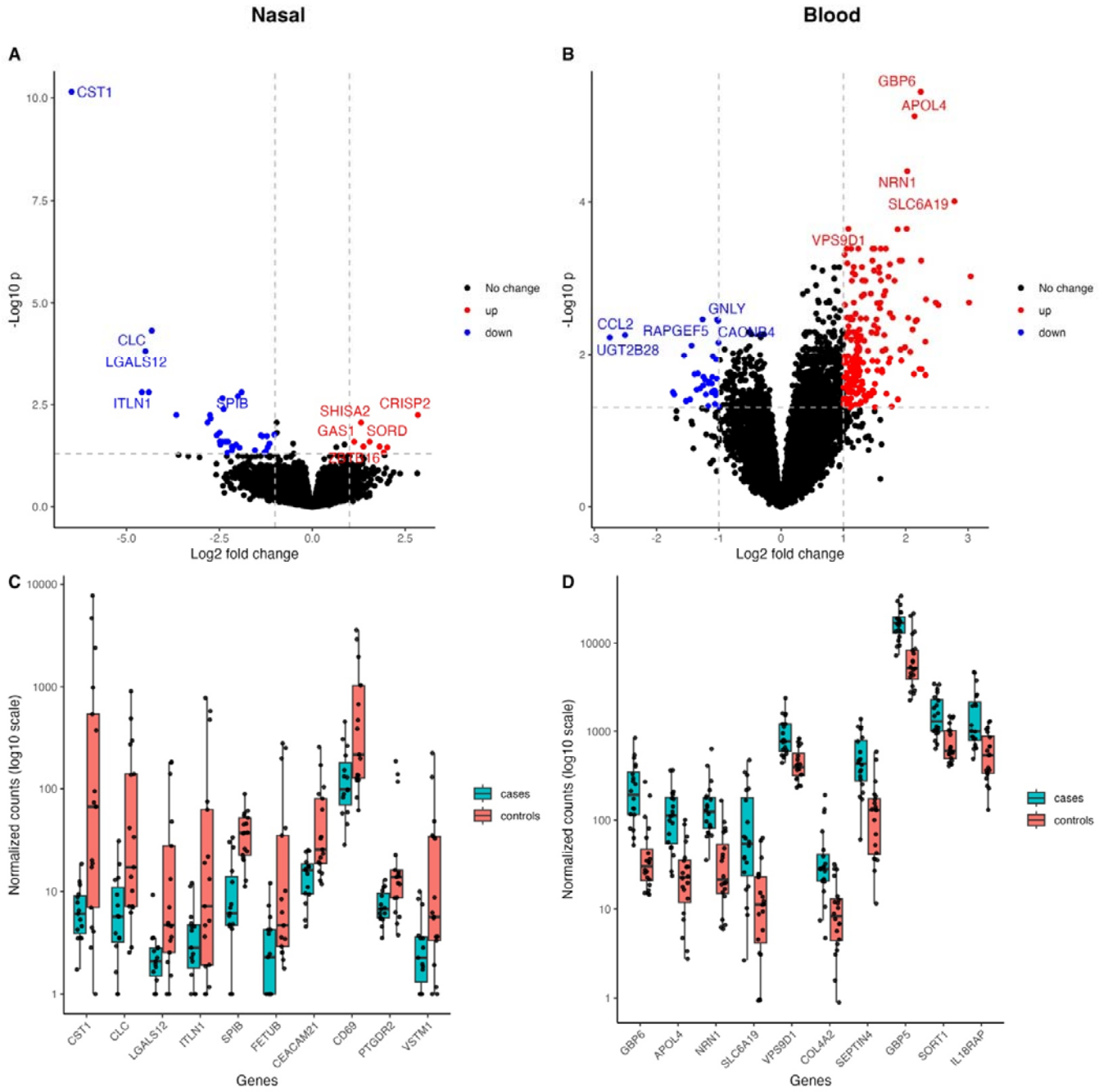

Figure 2: Volcano plots and box plots illustrating the top DEGs in PWH and TB compared to PWH without TB (controls). Volcano plots show the effect size by the -$\log_{10}$P value for nasal (A) and blood (B) samples. DEGs are highlighted in red (upregulated) or blue (downregulated), with gene names shown 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.

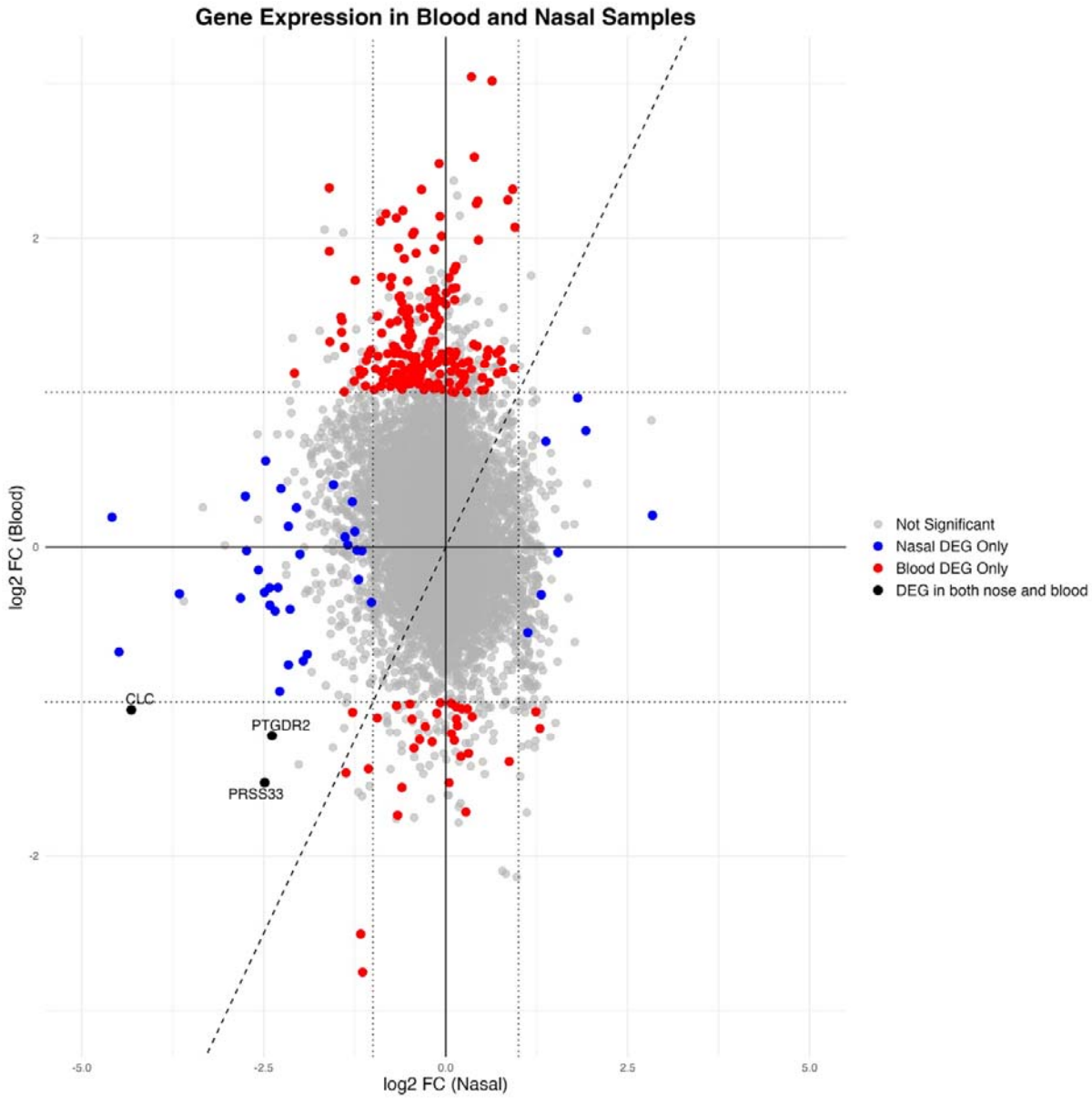

Figure 3: Agreement plot of log2FC for genes in blood samples (x-axis) and nasal samples (y-axis). Colored features are genes which were differentially expressed with log2 FC ≥1 and an adjusted p-value of <0.05. Dashed x=y line represents agreement. Dotted lines are shown at +/-1 log2FC. Labelled DEGs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

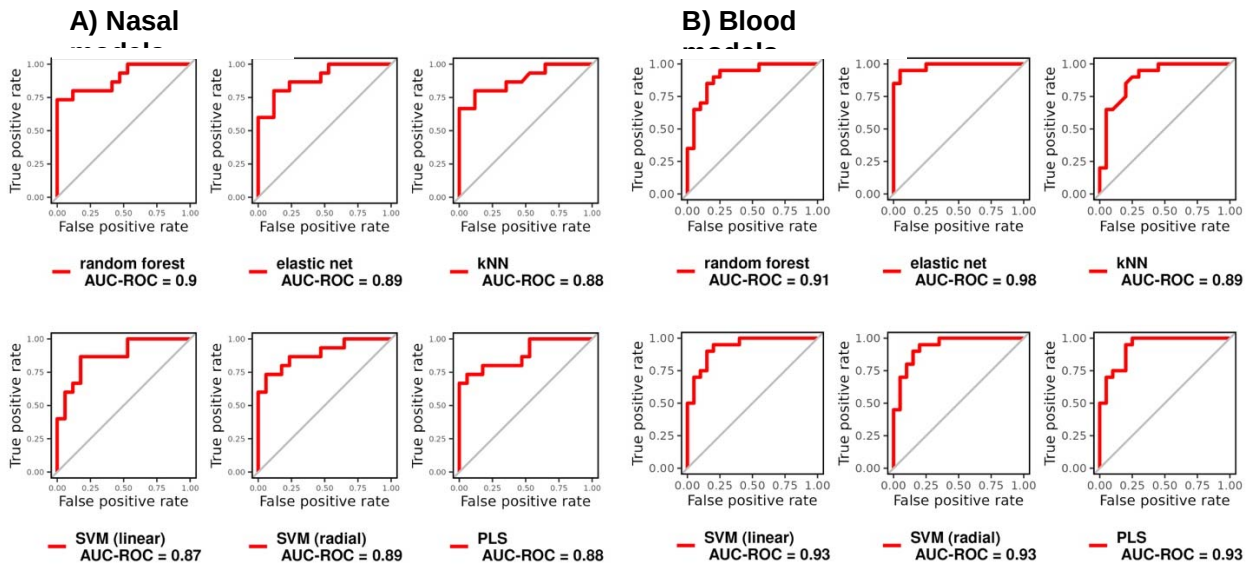

Figure 4: ROC curves and the respective AUC of different machine learning models in distinguishing TB from controls, trained using the (A) 44 nasal DEGs and (B) 238 blood DEGs on the respective datasets, with 10-fold cross-validation repeated 10 times.
DEGs= differentially expressed genes, SVM= support vector machine; KNN= k-nearest neighbor, ROC=receiver operating characteristic; AUC= area under the curve, PLS= partial least squares

	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)

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

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

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

Immune cell composition

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

TB Signature Profiler

In the RNA sequencing dataset, only 6/59 published TB diagnostic blood signatures distinguished TB from controls with an AUC ≥0.7 and p-value <0.05 in the nasal dataset (Maertzdorf_100, Lee_5, Sloot_2, Natarajan_7, Jacobensen_3 and Gjoen_10[22]) **(Figure 5A, 5C).** In contrast, most published blood signatures distinguished TB from controls in our blood dataset (44/59 and 41/59 had an AUC ≥0.7 and p-value <0.05 using ssGSEA and GSVA respectively) **(Figure 5B, 5D).** Only one signature which performed well in the nasal dataset (Jacobsen_3) performed well in the blood dataset. Full AUC results are in **Supplementary Table 5 and 6.**

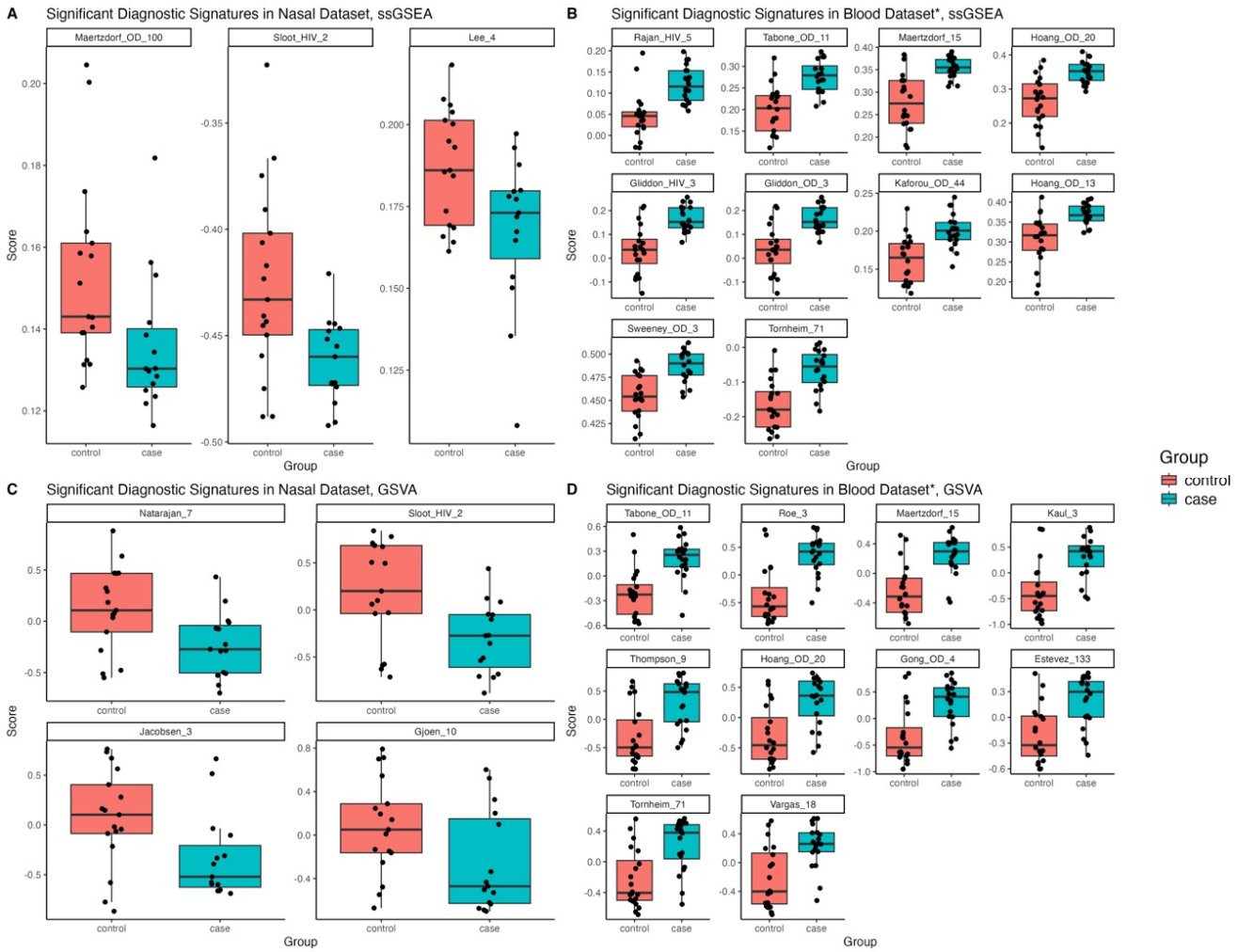

Figure 5: Boxplots of published blood diagnostic signatures that reached statistical significance (p value <0.05) and had a bootstrapped AUC of >0.7 in distinguishing TB from controls using ssGSEA in the A) nasal and B) blood datasets and using GSVA methods in the C) nasal and D) blood datasets.
**For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

Discussion

In a cohort of treatment-naive PWH with presumptive TB in Uganda, we identified 44 DEGs in nasal cells between adults with microbiologically confirmed TB disease compared with individuals without TB. These 44 DEGs yielded a promising AUC of > 0.87 in predicting TB across six models in cross-validation within our dataset. Amongst those with TB disease, we also found a blunted inflammatory nasal response, with downregulation of several innate immune pathways. Compartmental differences were seen, with contrasting patterns of inflammation in TB between the nose and blood and little overlap in nasal and blood DEGs. Our work highlights that there are differences in nasal gene expression associated with TB disease that are distinct from blood.

There are no published nasal transcriptomic studies in TB; however, half of the top ten nasal DEGs have been associated with TB in blood. *SPIB,* which ranked amongst the most important predictive genes across all models, is important for plasmacytoid dendritic cell function[23] and memory B cell maintenance[24]. *SPIB* downregulation was noted in active TB patients in multiple studies[25-27]. *CLC,* which is linked to neutrophil recruitment[28], was part of a four-marker blood signature for differentiating adults with TB and non-TB febrile infection[29]. *ITLN,* which has a role in pathogen recognition, was downregulated in newly diagnosed adults with TB, increasing in expression with treatment[25]. Changes in expression of T cell markers *PTGDR2* and *CD69* have also been associated with early signs of pulmonary TB[30] and TB infection, respectively[31]. Top DEGs not previously linked to TB include *CEACAM21*, which facilitates opsonin-independent recognition and clearance of specific gram-negative bacterial pathogens[32]. *CST1, CLC, FETUB* and *VSMT1* were upregulated in asthma and allergic diseases in nasal studies[33-35], which could reflect overrepresentation of allergic diseases in our non-TB control group.

GSEA, ORA and IPA suggested that TB was associated with suppressed nasal innate immune responses. IFN-γ activity was the top enriched downregulated pathway from the Hallmark gene-sets using fGSEA . IFN-γ is an important cytokine for immunity against mycobacteria. It facilitates the host defense to *Mtb* through stimulating macrophage activity and producing reactive nitrogen species[36], with ineffective IFN-γ activity associated with disseminated mycobacterial infection[37]. Neutrophils were the top downregulated REACTOME enriched pathway. Whilst excessive neutrophil activation can cause tissue damage, they have a protective role in early stages[38] and adult contacts with higher blood neutrophil counts were shown to be less likely to develop TB infection[39]. Chemokine signaling was the most enriched downregulated KEGG pathway. Single nucleotide polymorphisms in chemokines involved in cell migration in TB, were shown to increase susceptibility to active TB[40,41].

The nose is the first line of defense for airborne pathogens and early control of respiratory infection is dependent on robust nasal mucosal immune responses[42]. Dampening of IFN-γ activity, neutrophils, and chemokines in addition to other enriched nasal pathways like complement, phagocytosis, and toll-like receptor signaling, may underlie failure of the early protective nasal response to *Mtb* in TB disease.

The nasal host response in TB differed from blood. Few published TB blood signatures could classify TB in our nasal sequencing dataset. Moreover, downregulated nasal innate immune responses in TB were accompanied by upregulation of inflammatory pathways in blood. Unlike blood, nasal samples have more heterogeneity in cell composition, consisting of both immune and epithelial components[11]. We could not show that nasal immune pathways were driven by immune cell composition. However, neutrophils in the blood were higher in PWH with TB versus those without TB, consistent with our observation of increased neutrophil enrichment in blood.

Our finding of upregulated innate immune pathways in TB blood signatures confirms previously published studies in PWH[10,43,44]. Similar to the downregulation of immune nasal pathways in TB in our study, Mick and colleagues found that toll-like receptor signaling, inflammasomes, neutrophil degranulation and chemokine signaling were attenuated in the nasopharyngeal response in SARS-CoV-2 patients compared to other viral infections[13]. Hurst and colleagues also noted that systemic symptoms in SARS-CoV-2-infected individuals were correlated with suppressed nasal innate and adaptive immune pathways and upregulation of these same pathways in matched blood samples[45]. In contrast, two studies comparing nasal and blood samples in children with viral infections, identified upregulation in immune activity and strong innate responses in both compartments[11,12]. Although our cross-sectional study cannot infer causality, we can hypothesize that reduced mucosal immune responses and inadequate control of *Mtb* in the nose could promote progression to lung disease and consequent systemic inflammation. Longitudinal studies using paired nasal and blood samples are needed to understand the temporal relationship between the nasal and blood immune response and TB disease.

A streamlined nasal gene signature based on variable importance (NASAL_4) met the WHO criteria for a TB triage test (>90% sensitivity and 70% specificity) in cross-validation. Three genes within NASAL_4 relate to immune activity: *SPIB* is linked to interferon production and proinflammatory cytokines[23], *CD1b* is associated with antigen presentation[46] and *TESPA* relates to T cell maturation[47]. Whilst the AUC for nasal models based on all DEGs was slightly lower than blood, most of the top ten nasal DEGs were not ISGs, unlike blood. This is promising given that the performance of many ISG-

dominant RNA blood signatures in differentiating TB from viral infections or in PWH has been inadequate[8,10]. Moreover, adults with TB missed by blood models were classified correctly by nasal models and vice versa, suggesting potential value of combined nasal-blood diagnostic biomarkers.

All adults in our study had HIV. Studies of the nasal transcriptome in HIV are limited, however HIV causes mucosal immune dysfunction[48], which will shape nasal transcriptional changes, independently of TB. Based on flow cytometry and single cell sequencing, Phiri and colleagues found increased nasal neutrophilic abundance and inflammation in PWH compared with non-infected individuals and that nasal neutrophils had transcriptional signatures of stress and senescence and nasal T cells had signatures of exhaustion and apoptosis[49]. Differing numbers of nasal CD3 and CD4 cells based on HIV status have also been noted[50]. Further studies looking at the nasal TB response among individuals without HIV would help explore the extent to which nasal gene profiles in our study are driven by HIV.

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

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

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

In conclusion, our exploratory study suggests that nasal transcriptomics is feasible and informative for the study of TB, whereby elucidating gene expression profiles at the initial pathogen encounter reveals novel insights into host-TB interaction not captured in blood. We found distinct nasal gene expression patterns associated with TB disease in PWH, with DEGs able to distinguish TB from non-TB controls, supporting further study of the nose for TB diagnostic biomarkers.

Data availability

Computational code and data are available at:

https://github.com/nisreenkhambati/uganda_nasal_cells.

Raw expression data for nasal and blood samples are publicly available on the SRA database (PRJNA1379661) on https://www.ncbi.nlm.nih.gov/sra/PRJNA1379661

Conflict of interest

None to declare for all authors

Funding statement

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 and editing: all authors

Visualisation- NK

Supervision: RS, AJP, EMB, DOC, WEJ

Funding acquisition: JE, LN, WEJ, SG

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. 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
12. 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
13. 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
14. 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
15. 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
16. 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

17. 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
18. 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
19. 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
20. 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
21. 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
22. 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
23. 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
24. 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
25. 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
26. 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
27. 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
28. 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
29. 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
30. 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
31. 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
32. 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
33. 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
34. 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

35. 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
36. 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
37. 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
38. 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
39. 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
40. 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
41. 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.
42. 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
43. 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
44. 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
45. 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
46. 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
47. 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
48. Paiardini M, Frank I, Pandrea I, Apetrei C, Silvestri G. Mucosal immune dysfunction in AIDS pathogenesis. AIDS Rev. 2008;10(1):36-46.
49. 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
50. 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-#