Communications Medicine

https://doi.org/10.1038/s43856-025-01284-w

Article in Press

Conserved neutrophil degranulation transcripts in HIV-TB coinfected children across East and Southern Africa

Received: 29 April 2025

Accepted: 26 November 2025

Cite this article as: Katagirya, E., Mlotshwa, B., Kyobe, S. *et al.* Conserved neutrophil degranulation transcripts in HIV-TB coinfected children across East and Southern Africa. *Commun Med* (2026). https://doi.org/10.1038/s43856-025-01284-w

Eric Katagirya, Busisiwe Mlotshwa, Samuel Kyobe, Savannah Mwesigwa, Gaone Retshabile, Lesedi Williams, Marion Amujal, John Mukisa, Gerald Mboowa, David P. Kateete, Misaki Wayengera, Sununguko Wata Mpoloka, Angella N. Mirembe, Ishmael Kasvosve, Koketso Morapedi, Makhosazana Dlamini, Betty Nsangi, Grace P. Kisitu, Adeodata R. Kekitiinwa, Gabriel Anabwani, Moses L. Joloba, Eddie Mujjwiga Wampande, Dithan Kiragga, Florence Anabwani-Richter, Chester W. Brown, Graeme Mardon, Neil A. Hanchard, Mogomotsi Matshaba & and for the Collaborative African Genomics Network (CAfGEN)

Conserved neutrophil degranulation transcripts in HIV-TB coinfected children across East and Southern Africa

Eric Katagirya [1], Busisiwe Mlotshwa [2], Samuel Kyobe [1], Savannah Mwesigwa [1], Gaone Retshabile [2], Lesedi Williams [2], Marion Amujal [1], John Mukisa [1], Gerald Mboowa [1], David P. Kateete [1], Misaki Wayengera [1], Sununguko Wata Mpoloka [2], Angella N. Mirembe [3], Ishmael Kasvosve [2], Koketso Morapedi [2], Makhosazana Dlamini [6], Betty Nsangi [3], Grace P. Kisitu [3], Adeodata R. Kekitiinwa [3], Gabriel Anabwani [4], Moses L. Joloba [1], Eddie Mujjwiga Wampande [1], Dithan Kiragga [3], Florence Anabwani-Richter [5], Chester W. Brown [6], Graeme Mardon [7], Neil A. Hanchard [8], Mogomotsi Matshaba [4,7] and for the Collaborative African Genomics Network (CAfGEN)*

[1] College of Health Sciences, Makerere University, Kampala, Uganda.

[2] University of Botswana, Gaborone, Botswana.

[3] Baylor College of Medicine Children's Foundation Uganda (Baylor Uganda), Kampala, Uganda.

[4] Botswana-Baylor Children's Clinical Centre of Excellence, Gaborone, Botswana.

[5] Baylor College of Medicine Children's Foundation, Eswatini, Mbabane, Eswatini.

[6] University of Tennessee Health Science Center, Le Bonheur Children's Hospital, Memphis, TN, USA.

[7] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA.

[8] National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA.

*A list of authors and their affiliations appears at the end of the paper.

Keywords: Differential transcript expression, Tuberculosis, HIV-TB coinfection, Pediatric TB, Bulk RNAseq

Corresponding Author: Mogomotsi Matshaba, Email: matshaba@bcm.edu

Abstract

Background HIV-tuberculosis (HIV-TB) coinfection poses a significant public health challenge among children in high-burden African regions. Most previous transcriptomic studies have concentrated on adults and non-African populations, primarily analyzing gene-level differential expression. This approach overlooks multi-isoform complexity and may obscure both inherent and pathogen-induced intragenic heterogeneity. This multi-center case-control study aimed to identify and characterize the transcript-level landscape of HIV-TB coinfection in children from different African regions.

Methods We analyzed whole-blood RNA sequencing data from 97 children with and without tuberculosis from Uganda (East Africa) and from Botswana and Eswatini (Southern Africa). Reads were quality-controlled, and low-abundance transcripts filtered out. Differential transcript expression was estimated using models that adjusted for batch, age, and sex, with multiple testing controlled by the Benjamini–Hochberg procedure. Pathway enrichment was performed on the set of differentially expressed transcripts.

Results Our analyses show geographic heterogeneity in immune responses; however, the top three gene pathways – immune system, innate immune system, and neutrophil degranulation are consistently conserved across regions. Although there is limited overlap among upregulated transcripts, four of the six shared differentially expressed transcripts (DETs) are enriched in neutrophil degranulation pathways, indicating a conserved transcriptional signature of HIV-TB coinfection. Additionally, we identify five genes with region-specific, non-overlapping isoforms, a distinction not detectable through gene-level analysis.

Conclusions These findings demonstrate a conserved whole-blood transcriptomic signature in pediatric HIV-TB coinfection, while also highlighting regional variation at the isoform level. This supports the use of transcript-level analyses to identify biomarkers and enhance understanding of host responses in diverse African settings.

Plain language summary

Tuberculosis (TB) is difficult to diagnose in children with HIV, yet they are at a high risk of contracting it. We analyzed gene expression in the blood of children living with HIV with and without TB in Uganda, Botswana, and Eswatini. We focused on transcripts with at least a fourfold increase in abundance in HIV-TB coinfection. While children from different regions had different gene expression, many were involved in similar cellular processes, such as certain parts of the innate immune system, with four of the six shared transcripts. Our findings highlight

the conserved nature of the immune response and underscore the need for studies across different regions to better understand these infections in children.

Introduction

Human Immunodeficiency Virus (HIV) and tuberculosis (TB) co-infection remains an important pandemic syndrome or 'syndemic' in Sub-Saharan Africa, where the dual burden of these diseases is among the highest in the world [1,2]. In 2023, of the 39.9 million people living with HIV globally, 25.9 million (65%) were in Sub-Saharan Africa [3]. Similarly, of the 10.8 million new TB infections in the same year, 2.6 million (24%) occurred in the WHO African Region [4]. The co-infection presents unique diagnostic and therapeutic challenges due to the interaction between the two pathogens. The immunocompromised state following chronic HIV infection exacerbates *Mycobacterium tuberculosis* (MTB) pathogenesis, resulting in atypical TB presentations [5].

Children living with HIV who acquire TB are confronted with a disproportionately high risk of morbidity and mortality compared to their HIV-negative peers, particularly those under three years of age. Co-infection frequently complicates clinical presentation and management and is associated with rapid progression to severe or disseminated disease and suboptimal treatment outcomes. The immature immune systems of young children reduce their ability to control *Mycobacterium tuberculosis* infection, resulting in increased susceptibility to TB. Additionally, there are atypical, nonspecific symptoms that can be easily mistaken for those of other illnesses. This is particularly challenging in children due to the difficulty in obtaining adequate respiratory specimens for microbiological confirmation of *Mycobacterium tuberculosis*.

Furthermore, TB presents as a spectrum, including latent infection and subclinical infection, which limits the utility of current pathogen-centric diagnostic approaches [6,7]. Moreover, there is still an incomplete understanding of the pathophysiological mechanisms in coinfected children whose immune systems are still developing [8]. There is therefore a need to understand better the pathobiology of *M. tuberculosis* – HIV coinfection with the ultimate aim of developing new diagnostic approaches, therapies, and vaccines [9].

The interaction between HIV and TB affects gene expression, leading to unique transcriptional signatures in coinfected individuals [10,11]. However, gene-level differential expression analysis collapses multiple gene transcripts or isoforms into a single entity, inherently obscuring induced and intragenic heterogeneity. Alternative splicing (AS) is one of the post-transcriptional processes that enables the generation of different gene isoforms from a single gene by selectively including or excluding exons (or parts of exons/introns) during messenger RNA (mRNA) maturation [12]. It is well established that both HIV and *M. tuberculosis* affect AS in infected cells, such as in macrophages [13,14] and CD4+ T cells [15,16]. The gene isoforms generated may be structurally and functionally distinct, often featuring different protein domains critical to immune regulation (e.g., cytokine receptor signaling motifs) or pathogen recognition (e.g., antigen-binding domains), among other differences. Isoform shifts have been linked to TB-specific outcomes such as interferon-gamma resistance and tuberculin skin test conversion [17]. Thus, analyzing differentially expressed transcripts (DETs) offers a more refined resolution of transcriptome changes, enabling the decoding of mechanistic drivers of pathogenesis and potentially uncovering variations that may be overlooked in traditional DGE analysis.

Most prior transcriptomic studies have focused on predicting progression from latent TB infection (LTBI) to active disease or on treatment response, predominantly in adults or non-African populations [18]. These studies, alongside early adult work such as Berry et al. (2010) [19], consistently highlight, among other enriched pathways, neutrophil-driven signatures, particularly interferon-related pathways, as central to TB pathogenesis. However, the extent to which such mechanisms are conserved in children with HIV-TB co-infection remains poorly understood, especially in those of African ancestry.

In this exploratory study, we identify differentially expressed transcripts (DETs) distinguishing HIV-TB co-infected from HIV mono-infected children across geographically distinct African cohorts. While not driven by a single predefined hypothesis, the unbiased analysis highlights neutrophil-associated pathways, particularly those related to degranulation. These findings generate hypotheses for future work, including the potential of conserved neutrophil degranulation transcripts as biomarkers of TB co-infection in HIV-infected children and the role of regional transcriptional differences in reflecting host or pathogen diversity.

Methods

Study design and participant recruitment

This was a multi-center case-control study and was conducted on RNA sequences obtained from children infected with HIV recruited prospectively and retrospectively at three Baylor Centers of Excellence in Uganda, Botswana, and Eswatini as part of the Collaborative African Genomics Network (*CAfGEN*) of the Human Heredity and Health in Africa (H3Africa) consortium [20]. A major goal of the *CAfGEN* project was to understand mechanisms of TB Disease among HIV-infected children in Sub-Saharan African countries. Active TB was diagnosed by bacteriological confirmation and clinical acumen, according to the National Institutes of Health (NIH) classification of intrathoracic TB in pediatric populations [21,22]. This classified unconfirmed TB into possible and probable TB. Probable TB was defined as a child having at least one sign or symptom suggestive of TB, a suggestive chest radiography, and a positive clinical response to anti-tuberculosis treatment or a documented exposure to *M. tuberculosis* or immunological evidence of *M. tuberculosis* infection. Possible TB, on the other hand, required the presence of at least one sign or symptom suggestive of TB and a suggestive chest radiography, or a positive clinical response to anti-tuberculosis treatment, or a documented exposure to *M. tuberculosis* or immunological evidence of *M. tuberculosis* infection [23]. The bacteriological diagnosis of TB was made using either GeneXpert, sputum smear microscopy, or culture [24]. The controls, consisting of HIV mono-infected children, were those with confirmed HIV receiving care at the Centers of Excellence (COEs), who showed no signs or symptoms of TB and were examined by clinicians to rule out any features of TB. It is worth noting, however, that latent TB infection (LTBI) was not ruled out, as tuberculin skin tests (TST) or interferon-gamma release assays (IGRA) were not routinely performed at the time of recruitment. All children were on anti-retroviral therapy (ART) at the time of sampling, and none were receiving TB therapy.

Ethical considerations

The study was conducted in accordance with the Declaration of Helsinki and approved by the School of Biomedical Sciences Research and Ethics Committee (SBS-REC) at Makerere University, the Botswana Ministry of Health Research and Development Committee, and the Eswatini Health and Human Research Review Board (EHHRRB). *CAfGEN* also obtained IRB clearance from the Baylor College of Medicine (BCM) IRB. Additionally, in Uganda, the study was registered with the Uganda National Council of Science and Technology (UNCST). Informed consent was obtained from the parents/legal guardians of the children, and assent was obtained from children aged 11 years or older. Consent forms were in both English and the main local language at each study site (Luganda in Uganda, and Setswana and siSwati in Botswana and Eswatini, respectively).

Blood sample collection and processing

PAXgene RNA stabilization tubes (QIAGEN Inc., Valencia, CA), which stabilize intracellular RNA at the time of blood collection, were used to collect peripheral whole blood samples from children at the time of recruitment. Tubes were stored at -80 °C prior to RNA extraction, using the Manual PAXgene blood RNA extraction procedure according to the manufacturer's instructions.

RNA sequencing (RNA-seq)

RNA was isolated from whole blood using the PAXgene Blood RNA Kit (Qiagen, USA) at the Genomics, Molecular and Immunology (GMI) laboratory in Uganda and the University of Botswana for samples from Botswana and Eswatini. The RNA was then shipped on dry ice to the Laboratory for Translational Genetics (LTG) at the Children's Nutrition Research Center (CNRC) at Baylor College of Medicine, Houston, Texas. Complementary DNA (cDNA) libraries were prepared at the LTG lab from messenger RNA (mRNA) using the TruSeq Stranded mRNA (Illumina, San Diego, CA) sample preparation kit, as per the manufacturer's instructions. In brief, poly-T oligo-attached magnetic beads were used to purify poly(A) tails containing mRNA molecules. This was followed by fragmentation into small pieces. The fragments were then copied into cDNA using reverse transcriptase and random primers, resulting in the first and second strands. The cDNA underwent end repair with the addition of an "A" base and adapter ligation. Purification and PCR enrichment were performed prior to paired-end RNA sequencing. The samples were sequenced in two batches using a 98-base-pair read length protocol on a HiSeq 2500 (batch 1) and subsequently on a NovaSeq 6000 Illumina platform (batch 2) (Illumina, San Diego, CA), with an average read length of 139 bp. Sequencing was performed

according to the platform manufacturer's instructions, and base-calling files were converted to FASTQ files using bclToFastq.

Bioinformatics analysis

FastQC [25] and MultiQC [26] were used to assess the sequencing quality of the raw reads, and FastP [27] was used to clean out adapter sequences and low-quality reads using the default settings. The cleaned reads were mapped using Salmon v1.2.1 to the human reference genome (GRCh37). To ensure accurate quantification, mapping validation, GC bias correction, and sequence bias correction were enabled. Additionally, bootstrapping of 30 iterations was performed to estimate uncertainty. The sequencing depth of the cleaned reads ranged from 14 to 50 million reads (mean ± SD: 29.9 ± 9.6 million) for Uganda and from 17 to 46 million (mean ± SD: 29.8 ± 8.1 million) reads for Southern Africa. Filtering was performed on the identified transcripts, retaining only those with at least 10 reads in at least three samples. Differential transcript expression between the conditions was performed on the filtered transcript counts using DESeq2 (v1.42.1) [28] in R (v4.3.1) while correcting for batch, sex, and age (linearly scaled).

Additionally, correction for hidden batch effects was performed using ComBat-Seq, using the surrogate variation analysis (SVA v3.50.0) R package [29]. Normalization and batch correction were evaluated using MA plots (pre- and post-VST transformation) and unsupervised PCA plots generated before and after ComBat-Seq correction. To formally assess batch effects, we also performed PERMANOVA on VST counts, testing variance explained by sequencing batch before and after ComBat-Seq adjustment. A country-stratified PCA of the Southern Africa cohort was computed on VST-transformed, ComBat-Seq–adjusted counts using DESeq2 plotPCA (grouped by country and condition) to assess potential country structure. Functional enrichment analysis was performed using the STRING database (v12.0) (https://string-db.org/) [30] with a similarity setting of 1.0 and default settings for all other parameters. Ensembl transcript IDs were mapped to their corresponding HUGO names, which were then used for the pathway analyses in STRING. All visualizations, except those from STRINGdb, were created in R using ggplot2 v3.5.1 [31].

Statistics and Reproducibility

Descriptive statistics were used to analyze the participant study characteristics, specifically age and sex. Fisher's exact test was used for categorical characteristics between the cases (HIV-

TB coinfected) and controls (HIV mono-infected). The Mann-Whitney U-test was used for continuous variables. The differential analysis by DESeq2 uses a negative binomial to test for significance. Statistical significance for DET analysis was considered at a log2-fold change of 2 and an adjusted p-value < 0.05. Our study was adequately powered to identify DETs at the above-stated threshold in both regions. We performed region-specific statistical power analysis using the RNASeqPower R package. Power was estimated as a function of absolute $\log_2$fold-change ($|\log_2 FC|$), using the actual mean sequencing depths (Uganda: 29.93 million reads/sample; Southern Africa: 29.76 million), and dispersion values (biological coefficient of variation, BCV) derived from our DESeq2 fits, stratified by lower quartile (Q1), median, and upper quartile (Q3) values (Supplementary Table 3) with the significance threshold set at $\alpha = 0.05$. Power in the Uganda cohort (n = 59) exceeded 0.95 for detecting large transcript changes ($|\log_2 FC| \geq 2.0$) and remained ≥ 0.8 for moderate effects ($|\log_2 FC| \approx 1.0 – 1.5$) under typical dispersion levels. The Southern Africa cohort (n = 38) demonstrated reduced sensitivity at comparable effect sizes, although power remained ≥0.8 for $|\log_2 FC| \geq 2.0$ under lower dispersion. For small effects ($|\log_2 FC| \approx 0.5$), both cohorts were underpowered, particularly Southern Africa (Supplementary Figure 5). Although there is a risk of false negatives for modest transcript-level changes, especially in the Southern Africa subset, we appropriately controlled for false-positive rates using the Benjamini-Hochberg FDR correction within the DESeq2 analysis pipeline.

Results

Patient characteristics

The study included RNA sequences from 97 HIV-infected children from Uganda (n = 59) and Southern Africa (Botswana/Eswatini; n = 38), with a median age of 12 years (*Q1*–*Q3*: 7–14) and a 56% female representation. Active tuberculosis (ATB) was diagnosed in 60% of participants overall (Table 1). Further breakdown of the distribution of the different NIH TB diagnostic categories by country is shown in Supplementary Table 1. The age and sex distributions were similar between Uganda and Southern Africa, as shown in Table 1. Within Southern Africa, Botswana and Eswatini exhibited no significant differences in age (Wilcoxon p=0.063) or sex distribution (Fisher's exact p = 0.511, Supplementary Table 7). Uganda had a marginally higher proportion of ATB compared to Southern Africa (61% vs. 58%; Table 1). Clinical laboratory parameters such as CD4 count, total white blood cell count, and HIV viral load were

inconsistently recorded across study sites in Uganda, Eswatini, and Botswana, precluding their systematic inclusion in the analyses.

Data quality control

MA plots demonstrated stabilization of variance following VST normalization for both Uganda and Southern Africa (Supplementary Figure 5). Unsupervised PCA plots before and after ComBat-Seq correction showed no evident clustering by condition (ATB vs. NoTB) in either region, nor was there obvious clustering by batch (Supplementary Figures 3 and 4). Within Southern Africa, a country-stratified PCA showed overlap between Botswana and Eswatini with no country-specific clustering (Supplementary Figure 8). We demonstrated effective batch correction for both Uganda and Southern Africa by comparing variance in two principal components – PC1 and PC2. Uganda exhibited a pronounced batch effect, with batch explaining 20.7% of the variance (PERMANOVA: $R^2 = 0.207$, $F = 14.84$, $p = 0.001$). However, this batch effect was effectively removed by ComBat-Seq ($R^2 = 0.019$, $F = 1.10$, $p = 0.339$), without introducing differences in dispersion. The condition (ATB vs. NoTB) showed only a weak, non-significant trend before correction ($R^2 = 0.053$, $p = 0.070$), which disappeared post-correction ($R^2 = 0.016$, $p = 0.336$). The Southern Africa cohort showed no significant batch effect even before correction ($R^2 = 0.051$, $p = 0.217$), and ComBat-Seq further reduced the variance attributable to batch ($\Delta R^2 \approx -0.029$). Batch-associated dispersion remained non-significant both pre- and post-correction. Regional effects and condition effects were similarly non-significant across analyses. (Supplementary Tables 4 and 5). Quality control metrics derived from fastp analysis, such as the percentage of sequencing reads passing quality filters, base quality scores at or above Q30, and duplication rates, were consistent across all collection sites in Uganda, Botswana, and Eswatini. The Kruskal-Wallis test revealed no significant difference in the proportion of reads passing filters among the cohorts ($p = 0.139$). This result suggests the absence of site-specific technical bias in sequencing library quality (Supplementary Table 6 and Supplementary Figure 7).

Distinct upregulated transcripts in HIV-TB coinfection across geographic regions.

We compared transcript abundance in each country by evaluating the [~29 million] transcripts per sample meeting our criterion for sufficient expression (>10 reads in >3 samples); this represented 54.9% and 45.8% of the total transcripts from our East African (Ugandan) and Southern African (Botswana and Eswatini) samples, respectively (Figure 1a). Of this number,

566 transcripts were differentially expressed (|log2FC| > 2, adj. p < 0.05) in Uganda and 211 transcripts in Southern Africa, using the same criteria. There were 360 upregulated and 206 downregulated, compared with 166 upregulated and 45 downregulated in Southern Africa (Figure 1b). Multidimensional Scaling (MDS) analysis using upregulated transcripts revealed distinct clusters of Active Tuberculosis (ATB) and No Tuberculosis (NoTB) in both regions (Figure 1c), indicating a transcript-level separation between coinfected and mono-infected children.

There was, however, a low overlap between the upregulated DETs in Uganda and those in Southern Africa. Only nine transcripts were commonly upregulated in HIV-TB coinfected children between the two regions. Of these, only six had a baseMean expression greater than 50 in either region (high-confidence transcripts). These included *LTF*, *MMP8*, *OLFM4*, *PRTN3*, *METTL7B,* and *DDX39B* (Table 2). The baseMean is used to normalize for sequencing depth and is defined as the mean of normalized counts of all samples. The minimal overlap and distinct transcripts in each region were further evidenced by the clear demarcation in the variance-stabilizing transformed (VST) counts heatmap (Figure 2).

In addition to the shared differentially expressed transcripts, five shared genes (*HLA-B*, *PCBP2*, *ITGB2*, *HLA-F*, and *BAG6*) exhibited region-specific isoforms (Table 3), highlighting splicing variability that is masked by gene-level analysis. Moreover, some genes had multiple upregulated transcripts, particularly *PCBP2* and *BAG6*. These were based on high-confidence transcripts.

Significant overlap in immune pathways between regions despite minimal concordance in differentially expressed transcripts

Pathway analysis of upregulated transcripts in HIV-TB coinfection revealed significant enrichment of immune pathways in both regions (Figure 3). Notably, the top three pathways in both regions were identical: *Immune System* (Uganda FDR = 1.54e-13; SA FDR = 1.80e-12), *Innate Immune System* (Uganda FDR = 7.14e-12; SA FDR = 5.70e-11), and *Neutrophil Degranulation* (Uganda FDR = 4.92e-15; SA FDR = 1.34e-10). Despite minimal overlap in differentially expressed transcripts (DETs), eight pathways were shared between Southern Africa (out of 15 total enriched pathways) and Uganda (out of 14 total enriched pathways), accounting for 53% and 57% of their respective pathway sets (Figure 4a). The shared pathways, in addition to immune and neutrophil function, also included pathways related to infectious diseases, such as SARS-CoV infection (Figure 4b). Note that these samples were collected before the SARS-CoV-2 pandemic of 2019.

The seven pathways unique to Southern Africa included: immunoregulatory interactions between a Lymphoid and a non-lymphoid cell, Adaptive Immune System, Antigen Presentation: folding, assembly, and peptide loading of class I MHC, ER-phagosome pathway, Signaling by Interleukins, DAP12 interactions, and SARS-CoV-2 activates/modulates innate and adaptive immune responses. In contrast, Uganda exhibited six unique pathways: Antimicrobial peptides, SARS-CoV-2 Infection, Interferon Signaling, Alpha-defensins, Disease, and Interferon alpha/beta signaling (Supplementary Figure 2).

Shared transcripts are nearly exclusively involved in neutrophil degranulation

Notably, using the Reactome database in STRING, four of the six high-confidence shared transcripts were found to be linked to neutrophil degranulation. Protein-protein interaction networks among shared transcripts revealed strong connectivity with a network of 14 nodes and 69 edges (Figure 5). *LTF* and *MMP8* formed a hub interacting with *PRTN3* and *OLFM4*, suggesting coordinated regulation in neutrophil-mediated immunity. Moreover, 7 out of 10 proteins interacting with the input proteins are involved in neutrophil degranulation, namely: *TCN1, DEFA4, CEACAM8, AZU1, PGLYRP1, ELANE,* and *CD177*. Subcellular localization analysis of the shared transcripts confirmed their association with granules – tertiary, specific, and secretory granule lumen (Supplementary Figure 1).

Discussion

In this study, we aimed to identify and characterize differentially expressed gene isoforms/transcripts in children coinfected with HIV and TB across two Sub-Saharan regions: East Africa (Uganda) and Southern Africa (Botswana and Eswatini). The study revealed region-specific differentially expressed transcripts (DETs), evidenced by limited overlap, particularly among the upregulated transcripts in the two regions. However, there was marked concordance in the pathways enriched in both regions, with immune system, innate immune system, and neutrophil degranulation as the top three terms in both regions. The marked regional divergence in DETs may reflect evolutionary pressures from distinct MTB lineages [32] and HIV subtypes (A/D vs. C) [33], which may differentially impact the host immune system. Despite this, the conserved enrichment of neutrophil degranulation transcripts highlights a shared immunologic signature of HIV-TB co-infection in children. Because this analysis was conducted on bulk whole blood, such signatures could reflect either differences in cell-type abundance (e.g., neutrophilia) or transcriptional regulation within neutrophils. While these findings do not directly imply mechanistic involvement, they may serve as a basis for future studies exploring potential diagnostic or prognostic utility. Accordingly, our objective was not to disentangle these sources but to characterize a reproducible whole-blood signature that could serve as a practical biomarker in high-burden pediatric settings.

HIV and MTB have been independently shown to alter the transcriptional profile of the host cells through changes in alternative splicing, leading to the expression of different gene isoforms. For example, HIV, through Tat1 protein, alters alternative splicing in $CD4^+$ cells [15] while MTB can cause global perturbations in alternative splicing, particularly in macrophages [13]. Because gene isoforms arising from alternative splicing may have distinct properties that affect function,

stability, and localization, transcript-level analysis provides a more granular view of host responses to TB-HIV coinfection than conventional gene-level approaches. In the context of HIV-TB coinfection, DETs can therefore be interpreted as transcriptional signatures shaped by alternative splicing and other regulatory processes. Such transcript-level signatures may offer valuable insights into the diversity of host responses across settings and serve as a foundation for future studies.

We observed over 2.5 times more differentially expressed transcripts in East Africa than in Southern Africa (566 vs 211). This contrast likely reflects the influence of multiple factors, including host and pathogen genetic diversity in the two regions. Sub-Saharan Africa exhibits exceptional genetic diversity, including substantial intra-region variation, likely shaped by prolonged evolutionary pressures such as diverse pathogen exposure. This variation may contribute to differences in host genetic mechanisms, including gene regulation and pathogen recognition systems (e.g., HLA heterogeneity). Differences in *Mycobacterium tuberculosis* lineages are well documented across the African continent [32]. MTB lineage 4 is common in all three countries – Uganda, Eswatini, and Botswana [34–36], though with differing sub-lineages. Uganda has the Uganda sub-lineage as the predominant one, while Botswana has sub-lineages LAM, S, T, and X [34]. Additionally, Uganda and Eswatini have a considerable presence of Lineage 3 [37] and Lineage 2 [36], respectively. Similarly, HIV subtypes vary: Uganda has subtypes A and D, and recombinants of the two, while Eswatini and Botswana predominantly have subtype C [33]. Besides HIV and MTB, regional differences in exposure to endemic pathogens, viral co-infections, and environmental and sociodemographic factors (e.g., nutrition) may further influence host transcriptional profiles. For example, we have previously shown that HIV infected children in Uganda had a higher diversity of viral coinfections compared to those in Botswana [38]. These differences may operate through several mechanisms, including epigenetic modifications or altered immune responses.

Among the shared DETs across the regions, transcripts from four genes (*LTF, MMP8*, *OLFM4*, and *PRTN3)* mapped to neutrophil degranulation. Because our analysis used bulk whole blood, these DETs could reflect either increased neutrophil abundance or altered transcription within neutrophils. Regardless of the underlying source, their consistent detection underscores neutrophil-associated signatures as a defining feature of HIV-TB co-infection in children. These genes have been shown to have a dual role: antimicrobial defense (via LTF/PRTN3) and immunopathology (via MMP8-driven tissue damage). Lactoferrin (*LTF*) is an iron-binding glycoprotein found in mucosal secretions and neutrophils [39]. It has wide antimicrobial activity

against bacteria, fungi, and viruses, including HIV-1 [40]. It has also been proposed as an adjunctive TB therapy [39] and, in addition, to enhance the efficacy of the Bacillus Calmette-Guérin (BCG) vaccine [41].

On the other hand, proteinase 3 (*PRTN3*) is a serine protease with broad antimicrobial activity [42,43]. The other shared transcript from matrix metalloprotease 8 (*MMP8*) has been extensively characterized and is known to drive collagen degradation in pulmonary TB, contributing to cavity formation. Its secretion is upregulated in TB via AMP-activated protein kinase (AMPK) signaling in neutrophils, correlating with disease severity [44]. *MMP8* inhibition (e.g., via doxycycline [44]) warrants further investigation as a host-directed therapy to mitigate lung damage in TB-HIV coinfection. Moreover, neutrophils have been previously proposed as a host-directed target for TB therapy [45,46] or suggested as such [47]. The fourth transcript in the neutrophil degranulation pathway, *OLMF4*, is expressed in two neutrophil subpopulations, blood and extravascular tissue, and is associated with NETs [48]. NETs are involved in neutrophil phagocytic activity. Taken together, these DETs represent a reproducible neutrophil-related transcriptional signature, consistent with previous reports from adult HIV-TB and TB cohorts [49,50]. A question remains as to whether our results could be influenced by the fact that most studies informing the databases are from adults.

We also observed isoform-specific variation in five shared genes (*ITGB2*, *PCBP2*, *BAG6*, *HLA-F*, and *HLA-B*). The isoform difference would have been masked in conventional differential gene expression (DGE) analysis. It is well established that different gene isoforms can exhibit distinct regulation, localization, and function, among other differences [14,51]. One of these genes, *ITGB2* (Integrin beta 2), has been previously linked to TB susceptibility. Among these, *ITGB2* is notable, as gene-level studies have previously linked it to TB susceptibility [52]. In our data, transcript-level differences in ITGB2 and other immune-related genes suggest that isoform-resolved analysis adds value by uncovering transcriptional signatures that may refine understanding of host responses in HIV-TB co-infection. While the functional consequences of isoform variation were not examined in this study, these findings provide a rationale for future studies that integrate isoform-specific annotation and validation.

The study, however, faced several limitations, including variable availability of clinical laboratory parameters, such as $CD4^+$ count, total white blood cell count, and HIV viral load, which limited their systematic inclusion in the analyses. Secondly, the inability to rule out latent TB infection (LTBI) in the control group and the collapsing of the different TB phenotypes (confirmed, probable, and possible) into active TB may have diluted the observed transcriptomic differences. We also did not account for HIV and *M. tuberculosis* strain types in either region, yet pathogen genetics is critical for host-pathogen interactions. However, the effect due to pathogen heterogeneity may have been minimized or controlled for by conducting analyses by region. We also note that our study did not account for fine-scale population genetic structure within the Ugandan and Southern African cohorts. While we mitigated broad ancestral stratification by analyzing each region separately, we lacked genotype data and could not include ancestry principal components as covariates in our models. To address this limitation, we quantified demographic diversity between and within regions. Age and sex distributions were similar between Uganda and Southern Africa (Table 1) and between Botswana and Eswatini (Supplementary Table 7). These similarities help minimize between-group and within-group demographic imbalance. Unmeasured host genetic heterogeneity may therefore act as a confounding variable, particularly for identifying transcripts with small or modest effect sizes. Future studies incorporating genotyping data would be valuable for controlling this potential source of variation and confirming our findings.

Conclusions

Our transcript-resolution analysis of HIV-TB coinfected children from geographically distinct African populations reveals a conserved upregulation of immune pathways despite significant regional heterogeneity in individual transcripts. The neutrophil degranulation pathway was strikingly enriched among the small set of transcripts shared between regions. While this signature could reflect either transcriptional changes within neutrophils or an increase in neutrophil abundance, a limitation of bulk RNAseq, its consistent emergence across diverse populations highlights its potential value as a conserved biomarker signature for pediatric HIV-TB coinfection. Furthermore, the discovery of region-specific isoforms for immune genes, such as ITGB2 and HLA-B, which are masked in conventional gene-level analysis, underscores the

value of transcript-resolution studies in uncovering nuanced and population-specific host responses. This work is fundamentally exploratory, generating hypotheses for future validation and verification.

Recommendations

Our findings warrant further investigation to validate the conserved neutrophil degranulation transcript signature in independent pediatric cohorts to confirm its value as a biomarker for HIV-TB co-infection in children while controlling for pathogen genetic diversity and other factors such as nutrition.

DATA AVAILABILITY

All source data underlying the graphs and charts presented in the main text and Supplementary Information are available in Figshare under the title “CAfGEN RNA-seq count matrices and supporting processed data for ‘Conserved neutrophil degranulation transcripts in HIV-TB coinfected children across East and Southern Africa’” (https://doi.org/10.6084/m9.figshare.30551264). The RNA-seq datasets generated and analyzed during the current study are being deposited under the project name “CAfGEN RNA-seq Data” into the European Genome-phenome Archive (https://www.ebi.ac.uk/ega/) in accordance with the H3Africa Consortium consensus agreement. Data deposition is being coordinated through H3ABioNet/AfriGen-D, the authorized H3Africa data submission center, in accordance with the NIH H3Africa Data Access and Release Policy. The datasets will become publicly discoverable on the EGA once accession identifiers are issued. Access will be controlled and granted only to bona fide researchers whose proposed use aligns with the participants’ informed consent, subject to approval by the H3Africa Data and Biospecimen Access Committee (DBAC). Requests for access should be submitted via the H3Africa data portal (https://www.h3abionet.org/resources/h3africa-archive). Until the accession number is available, qualified researchers may contact the corresponding author for information on the submission status or to initiate the access request process.

CODE AVAILABILITY

Scripts used for data analysis are available at https://github.com/katagirya [53].

Author Contributions

EK, BM, SK, MW, GM, AK, BN, MM, GA, IK, SWM, MLJ, CB, GM#, and NH: conceptualization, writing – review and editing, funding acquisition, investigation, and project administration. EK, SK, SM, MW, MA, JM, AM, BN, GR, BM, LW, and KM: data curation. EK, SK, SM, and NH: formal analysis. GM, NH, SK, MLJ, DPK, CB, DK, GA: funding acquisition. EK, SK, GK, GR, BN, MM, CB, GM, and NH: methodology. SK, GK, MD: project administration. GM, NH, MLJ, and EMW: supervision. EK, SK, SM, GM, MA, FAR, DK, and NH: validation and writing – original draft. The CAfGEN Consortium coordinated participant recruitment, sample processing, and sequencing, and contributed to data curation and quality control. All authors contributed to the article and approved the submitted version.
GM = Graeme Mardon and GM# = Gerald Mboowa

Competing interests

Gerald Mboowa is an Editorial Board Member for Communications Medicine but was not involved in the editorial review or peer review, nor in the decision to publish this article.

Acknowledgments

We would like to acknowledge Nasinghe Emmanuel, Gaseene Sebetso, Thembela Mavuso, Bheki Ntshangase, Buhle Dlamini, Abhilash Sathyamoorthi, Bathusi Mathuba, Yves Mafulu, Gerald Mboowa, Harriet Nakayiza, Edgar Kigozi, Fred Katabazi, Keboletse Mokete, Lesego Ketumile, Kennedy Sichone, Keofentse Mathuba, LeToya Balebetse, Muambi Muyaya, Nancy Zwane, Nicholas Muriithi, Sibongile Mumanga, Thabo Diphoko, Thobile Jele, and Thato Regonamanye. Finally, we acknowledge the families and children who participated in the study.

The project described was supported by Award Number U54AI110398, administered by the National Institute of Allergy and Infectious Diseases (NIAID), the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD), and the National Human Genome Research Institute (NHGRI), as part of the NIH Common Fund H3Africa Initiative. Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number U01HD114479. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References:

1. Gelaw, Y. A., Williams, G., Soares Magalhães, R. J., Gilks, C. F. & Assefa, Y. HIV Prevalence Among Tuberculosis Patients in Sub-Saharan Africa: A Systematic Review and Meta-analysis. *AIDS Behav.* **23**, 1561–1575 (2019).

2. Kwan, C. & Ernst, J. D. HIV and tuberculosis: A deadly human syndemic. *Clin. Microbiol. Rev.* **24**, 351–376 (2011).

3. UNAIDS. *Global HIV & AIDS Statistics — 2024 Fact Sheet*. (2024).

4. World Health Organization. *Global Tuberculosis Report*. (2024).

5. Venturini, E. *et al.* Tuberculosis and HIV co-infection in children. *BMC Infect. Dis.* **14**, 1–10 (2014).

6. McNerney, R. *et al.* Tuberculosis Diagnostics and Biomarkers: Needs, Challenges, Recent Advances, and Opportunities. *J. Infect. Dis.* **205**, S147–S158 (2012).

7. Sivakumaran, D. *et al.* Host blood-based biosignatures for subclinical TB and incipient TB: A prospective study of adult TB household contacts in Southern India. *Front. Immunol.* **13**, 1051963 (2022).

8. Xu, H., Blair, R. V., Veazey, R. S. & Wang, X. Immunopathogenesis in HIV-associated pediatric tuberculosis. *Pediatr. Res.* **91**, 21 (2021).

9. Azad, A. K., Lloyd, C., Sadee, W. & Schlesinger, L. S. Challenges of Immune Response Diversity in the Human Population Concerning New Tuberculosis Diagnostics, Therapies, and Vaccines. *Front. Cell. Infect. Microbiol.* **10**, 519431 (2020).

10. Tepekule, B. *et al.* Transcriptional profile of Mycobacterium tuberculosis infection in people living with HIV. *iScience* **27**, (2024).

11. Duffy, F. J., Thompson, E. G., Scriba, T. J. & Zak, D. E. Multinomial modelling of TB/HIV co-infection yields a robust predictive signature and generates hypotheses about the HIV+TB+ disease state. *PLoS One* **14**, (2019).

12. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).

13. Kalam, H., Singh, K., Chauhan, K., Fontana, M. F. & Kumar, D. Alternate splicing of transcripts upon Mycobacterium tuberculosis infection impacts the expression of functional protein domains. *IUBMB Life* vol. 70 845–854 (2018).

14. Kalam, H., Fontana, M. F. & Kumar, D. Alternate splicing of transcripts shape macrophage response to Mycobacterium tuberculosis infection. *PLoS Pathog.* **13**, e1006236 (2017).

15. Byun, S., Han, S., Zheng, Y., Planelles, V. & Lee, Y. The landscape of alternative splicing in HIV-1 infected CD4 T-cells. *BMC Med. Genomics* **13**, 1–10 (2020).

16. Imbeault, M., Giguère, K., Ouellet, M. & Tremblay, M. J. Exon Level Transcriptomic Profiling of HIV-1-Infected CD4+ T Cells Reveals Virus-Induced Genes and Host Environment Favorable for Viral Replication. *PLoS Pathog.* **8**, e1002861 (2012).

17. Simmons, J. D. *et al.* Differentially expressed transcript isoforms associate with resistance to tuberculin skin test and interferon gamma release assay conversion. *PLoS One* **18**, (2023).

18. Gupta, R. K. *et al.* Concise whole blood transcriptional signatures for incipient tuberculosis: a systematic review and patient-level pooled meta-analysis. *Lancet Respir. Med.* **8**, 395–406 (2020).

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

20. Mboowa, G. *et al.* Open Peer Review The Collaborative African Genomics Network (CAfGEN): Applying Genomic technologies to probe host factors important to the progression of HIV and HIV-tuberculosis infection in sub-Saharan Africa [version 1; referees: 1 approved]. *AAS Open Res.* **1**, 3 (2018).

21. Zar, H. J., Workman, L. J., Little, F. & Nicol, M. P. Diagnosis of Pulmonary Tuberculosis in Children: Assessment of the 2012 National Institutes of Health Expert Consensus Criteria. *Clin. Infect. Dis.* **61**, S173–S178 (2015).

22. Graham, S. M. *et al.* Clinical Case Definitions for Classification of Intrathoracic Tuberculosis in Children: An Update. *Clin. Infect. Dis.* **61**, S179–S187 (2015).

23. Piccini, P., Chiappini, E., Tortoli, E., de Martino, M. & Galli, L. Clinical peculiarities of tuberculosis. *BMC Infect. Dis.* **14**, 1–12 (2014).

24. WHO. *Guidance for national tuberculosis programmes on the management of tuberculosis in children Second edition*. (2014).

25. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).

26. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

27. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

28. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

29. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: Batch effect adjustment for RNA-seq count data. *NAR Genomics Bioinforma.* **2**, (2020).

30. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).

31. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

32. Chihota, V. N. *et al.* Geospatial distribution of Mycobacterium tuberculosis genotypes in Africa. *PLoS One* **13**, (2018).

33. Nduva, G. M., Nazziwa, J., Hassan, A. S., Sanders, E. J. & Esbjörnsson, J. The role of phylogenetics in discerning hiv-1 mixing among vulnerable populations and geographic regions in sub-saharan africa: A systematic review. *Viruses* vol. 13 1174 (2021).

34. Mogashoa, T. *et al.* Genetic diversity of Mycobacterium tuberculosis strains circulating in Botswana. *PLoS One* **14**, e0216306 (2019).

35. Micheni, L. N., Kassaza, K., Kinyi, H., Ntulume, I. & Bazira, J. Diversity of Mycobacterium tuberculosis Complex Lineages Associated with Pulmonary Tuberculosis in Southwestern, Uganda. *Tuberc. Res. Treat.* **2021**, 1–6 (2021).

36. Dlamini, T. C., Mkhize, B. T., Sydney, C., Maningi, N. E. & Malinga, L. A. Molecular investigations of Mycobacterium tuberculosis genotypes among baseline and follow-up strains circulating in four regions of Eswatini. *BMC Infect. Dis.* **23**, 1–13 (2023).

37. Wampande, E. M. *et al.* Genetic variability and consequence of Mycobacterium tuberculosis lineage 3 in Kampala-Uganda. *PLoS One* **14**, e0221644 (2019).

38. Mwesigwa, S. *et al.* Unmapped exome reads implicate a role for Anelloviridae in childhood HIV-1 long-term non-progression. *npj Genomic Med.* **6**, 1–9 (2021).

39. Actor, J. K. Lactoferrin: A modulator for immunity against tuberculosis related granulomatous pathology. *Mediators of Inflammation* vol. 2015 409596 (2015).

40. Zupin, L. *et al.* Association Between LTF Polymorphism and Risk of HIV-1 Transmission Among Zambian Seropositive Mothers. *Curr. HIV Res.* **16**, 52–57 (2017).

41. Hwang, S. A., Arora, R., Kruzel, M. L. & Actor, J. K. Lactoferrin enhances efficacy of the BCG vaccine: comparison between two inbred mice strains (C57BL/6 and BALB/c). *Tuberculosis* **89**, S49–S54 (2009).

42. Pham, C. T. N. Neutrophil serine proteases: Specific regulators of inflammation. *Nature Reviews Immunology* vol. 6 541–550 (2006).

43. Reeves, E. P. *et al.* Killing activity of neutrophils is mediated through activation of proteases by K+ flux. *Nature* **416**, 291–297 (2002).

44. Ong, C. W. M. *et al.* Neutrophil-Derived MMP-8 Drives AMPK-Dependent Matrix Destruction in Human Pulmonary Tuberculosis. *PLoS Pathog.* **11**, (2015).

45. Dallenga, T. *et al.* Targeting neutrophils for host-directed therapy to treat tuberculosis. *International Journal of Medical Microbiology* vol. 308 142–147 (2018).

46. Remot, A., Doz, E. & Winter, N. Neutrophils and close relatives in the hypoxic environment of the tuberculous granuloma: New avenues for host-directed therapies?

Front. Immunol. **10**, 439028 (2019).

47. Scott, N. R. *et al.* S100A8/A9 regulates CD11b expression and neutrophil recruitment during chronic tuberculosis. *J. Clin. Invest.* **130**, 3098–3112 (2020).

48. Welin, A. *et al.* The Human Neutrophil Subsets Defined by the Presence or Absence of OLFM4 Both Transmigrate into Tissue In Vivo and Give Rise to Distinct NETs In Vitro. *PLoS One* **8**, e69575 (2013).

49. Kulkarni, V. *et al.* A Two-Gene Signature for Tuberculosis Diagnosis in Persons With Advanced HIV. *Front. Immunol.* **12**, 436 (2021).

50. Meier, S. *et al.* Neutrophil degranulation, NETosis and platelet degranulation pathway genes are co-induced in whole blood up to six months before tuberculosis diagnosis. *PLoS One* **17**, e0278295 (2022).

51. Liu, Y. *et al.* Impact of Alternative Splicing on the Human Proteome. *Cell Rep.* **20**, 1229–1241 (2017).

52. Liu, H. *et al.* Correlation between ITGB2 expression and clinical characterization of glioma and the prognostic significance of its methylation in low-grade glioma(LGG). *Front. Endocrinol. (Lausanne).* **13**, 1106120 (2023).

53. Eric, K. CAfGEN DET analysis (v1.0.0). (2025) doi:https://doi.org/10.5281/zenodo.17521612.

Tables

Table 1: Patient Characteristics by Region

	Age (*Q1, Q3*)	Females	ATB	NoTB
Uganda	10 (7,14)	32 (54%)	36 (61%)	23 (39%)
Southern Africa	12 (5,15)	22 (58%)	22 (58%)	16 (42%)
Combined	12 (7,14)	54 (56%)	58 (60%)	39 (40%)

Note: Q1 = First Quartile, Q3 = Third Quartile; ATB = Active Tuberculosis, NoTB = No Tuberculosis.

Table 2: Shared upregulated transcripts in Uganda and Southern Africa regardless of base mean.

	Transcript	Gene	Uganda BM (log2FC)	SA BM (log2FC)
1	ENST00000231751.4	*LTF*	2149.75 (5.04)	2878.68 (4.02)
2	ENST00000236826.3	*MMP8*	1009.98 (5.29)	1127.59 (5.60)
3	ENST00000219022.2	*OLFM4*	228.43 (4.89)	629.02 (4.56)
4	ENST00000234347.5	*PRTN3*	95.35 (4.88)	223.33 (5.56)
5	ENST00000394252.3	*METTL7B*	103.66 (4.70)	143.65 (3.82)
6	ENST00000473685.1	*DDX39B*	450.18 (19.73)	96.45 (23.02)
7	ENST00000177648.9	*ALPK1*	34.61 (29.37)	146.6 (9.06)
8	ENST00000369733.3	*COL17A1*	37.09 (4.44)	56.58 (4.50)
9	ENST00000251582.7	*ADAMTS2*	38.72 (8.26)	15.63 (8.39)

Note: BM = baseMean, SA = Southern Africa

Table 3: Region-specific transcripts in genes common between the two regions.

	Gene	Uganda Tx	Uganda BM (log2FC)	SA Tx	SA BM (log2FC)
1	*HLA-B*	ENST00000437265	1492.6 (30)	ENST00000492573	669.79 (18.2)
2	*PCBP2*	ENST00000439930	94.46 (14.16)	ENST00000552296	146.40 (20.4)
				ENST00000552819	81.31 (19.1)
3	*ITGB2*	ENST00000397854	85.83 (15.6)	ENST00000397857	59.83 (25.1)
4	*HLA-F*	ENST00000489502	248.53 (30)	ENST00000334668	143.71 (17.7)
5	*BAG6*	ENST00000443182	60.95 (28.1)	ENST00000383446	132.27 (14.5)
				ENST00000442479	160.29 (10.1)

Tx = Transcript, SA = Southern Africa, BM = baseMean

COLLABORATIVE AFRICAN GENOMICS NETWORK (CAfGEN)

Eric Katagirya [1], Busisiwe Mlotshwa [2], Samuel Kyobe [1], Savannah Mwesigwa [1], Gaone Retshabile [2], Lesedi Williams [2], Marion Amujal [1], Gerald Mboowa [1], David P. Kateete [1], Misaki Wayengera [1], Sununguko Wata Mpoloka [2], Angella N. Mirembe [3], Ishmael Kasvosve [2], Koketso Morapedi [2], Makhosazana Dlamini [6], Betty Nsangi [3], Grace P. Kisitu [3], Adeodata R. Kekitiinwa [3], Gabriel Anabwani [4], Moses L. Joloba [1], Eddie Mujjwiga Wampande [1], Chester W. Brown [6], Graeme Mardon [7], Neil A. Hanchard [8], Mogomotsi Matshaba [4,7], Masego Tsimako- Johnstone [2], and Keofentse Mathuba [2]

[1] College of Health Sciences, Makerere University, Kampala, Uganda, [2] University of Botswana, Gaborone, Botswana, [3] Baylor College of Medicine Children's Foundation Uganda (Baylor Uganda), Kampala, Uganda, [4] Botswana-Baylor Children's Clinical Centre of Excellence, Gaborone, Botswana, [5] Baylor College of Medicine Children's Foundation, Eswatini, Mbabane, Eswatini, [6] University of Tennessee Health Science Center, Le Bonheur Children's Hospital, Memphis, TN, USA, [7] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA, [8] National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA.

Figure legends

Figure 1: Transcript filtering, differential expression, and MDS separation in pediatric HIV–TB coinfection across regions. a, Proportion of transcripts retained after applying the expression threshold (≥10 reads in ≥3 samples). **b,** Volcano plots of differentially expressed transcripts (DETs; |log2FC| > 2, adjusted $P < 0.05$) for Uganda and Southern Africa. **(c)** Multidimensional scaling (MDS) based on upregulated transcripts shows separation between children with active tuberculosis (ATB) and those without tuberculosis (NoTB) in both regions.

Figure 2: **Heat map of upregulated transcripts across regions**. Variance-stabilized (VST) normalized counts for upregulated transcripts are shown by sample (columns) and transcript (rows) for Southern Africa and Uganda. Rows highlighted in red indicate transcripts that are upregulated in both regions, representing shared differentially expressed transcripts (DET). Color intensity reflects relative expression.

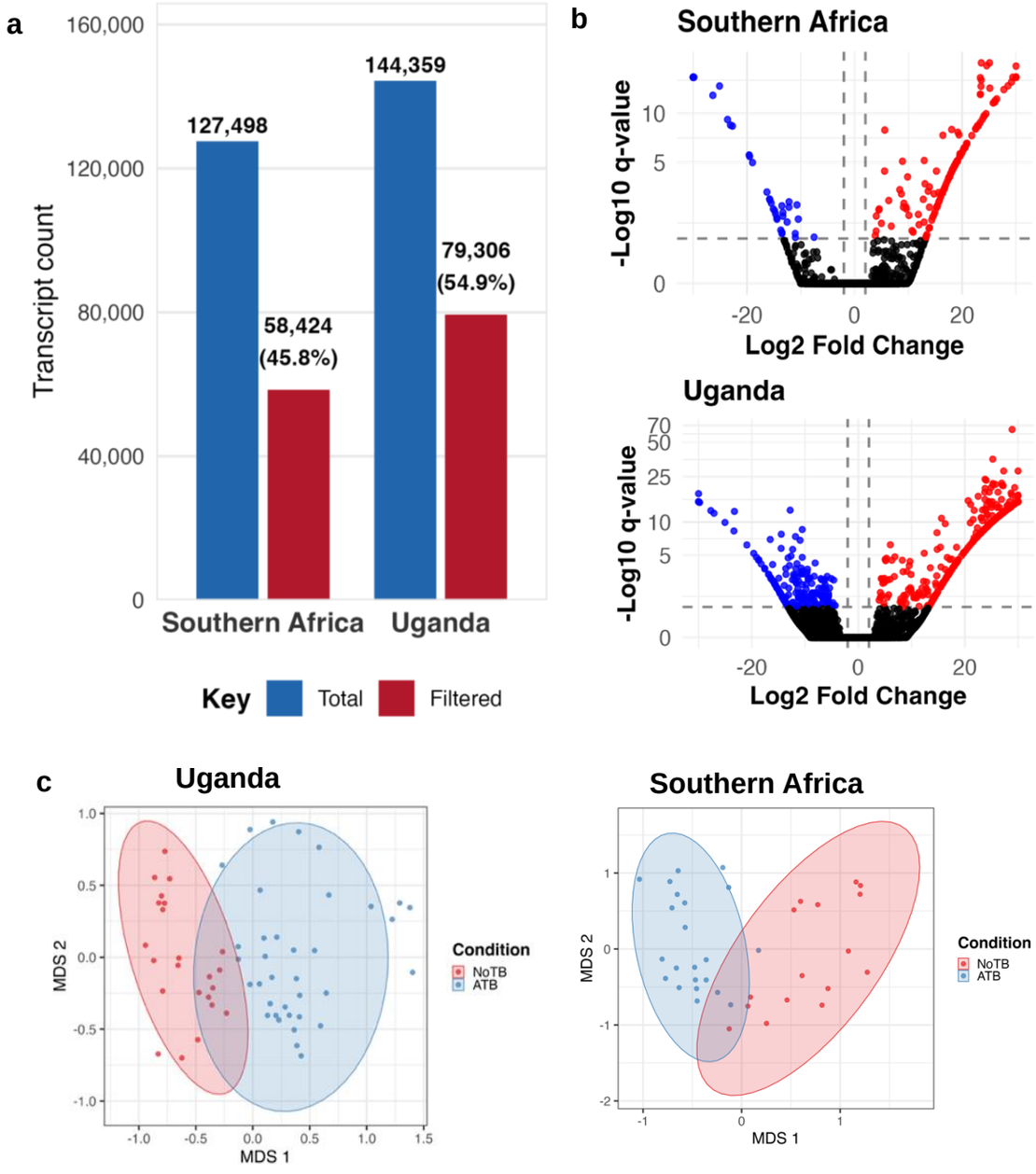
Figure 3: Pathway chord diagrams for upregulated transcripts. a, Southern Africa. **b,** Uganda. Chord links connect upregulated genes to enriched Reactome pathways, illustrating membership of genes (left sectors) in immune pathways (right sectors), including neutrophil degranulation, innate immunity, and adaptive immune processes, among others.

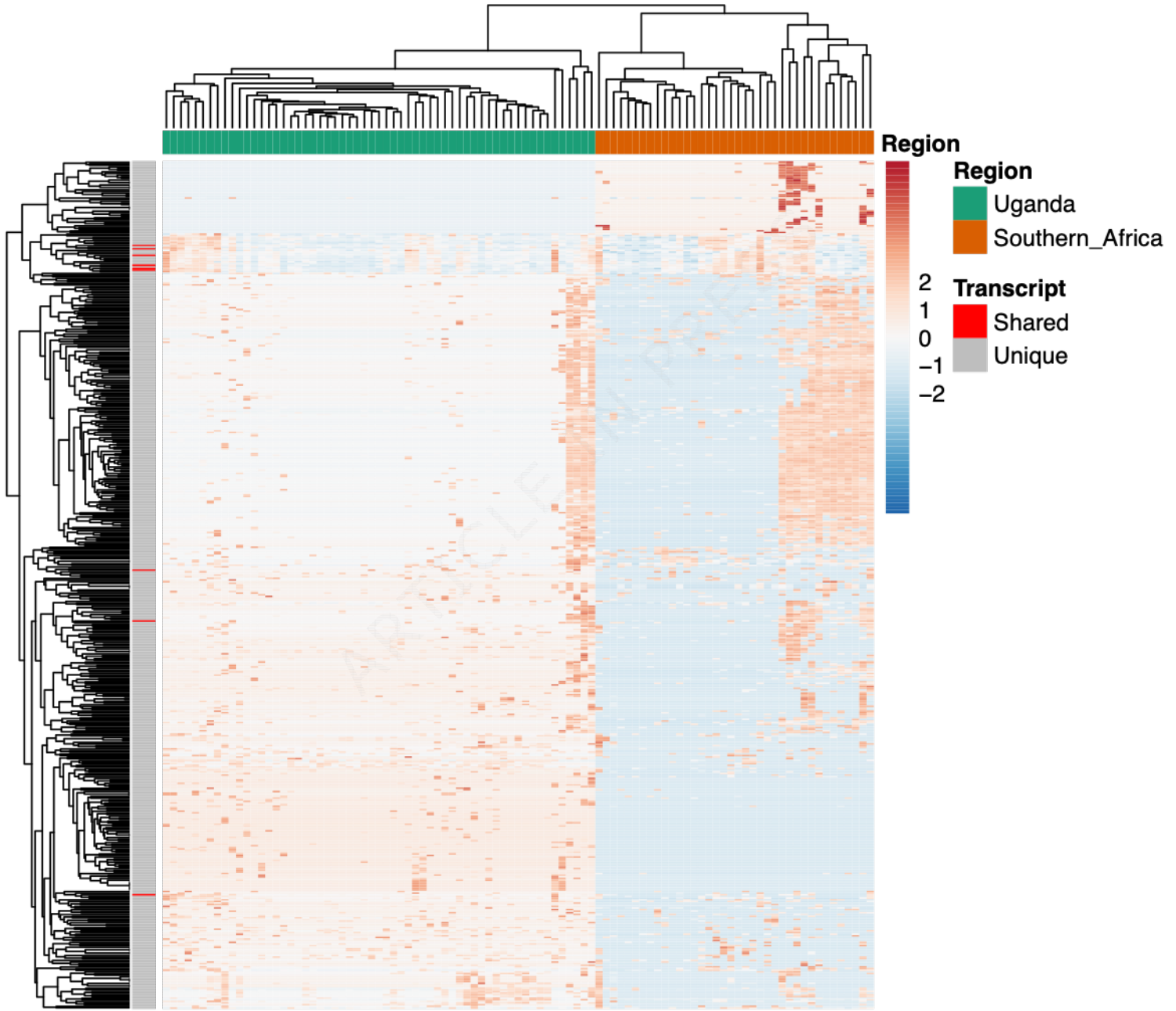
Figure 4: Shared Reactome pathways across regions. a, Venn diagram showing the overlap of enriched Reactome pathways between Uganda and Southern Africa. **b,** Bar–bubble–line plot for shared pathways: bars show GeneRatio (number of differentially expressed transcripts with baseMean > 50 divided by total genes in the pathway) by region; bubble size encodes the number of genes; dashed line indicates –log10(FDR) on the secondary y-axis.

Figure 5: Protein-protein interaction (PPI) network centered on PRTN3, MMP8, OLFM4, and LTF. Network generated with STRING; input genes are highlighted, with additional interactors shown and edge thickness proportional to interaction confidence. The network contains 14 nodes and 69 edges, with LTF and MMP8 acting as central hubs.

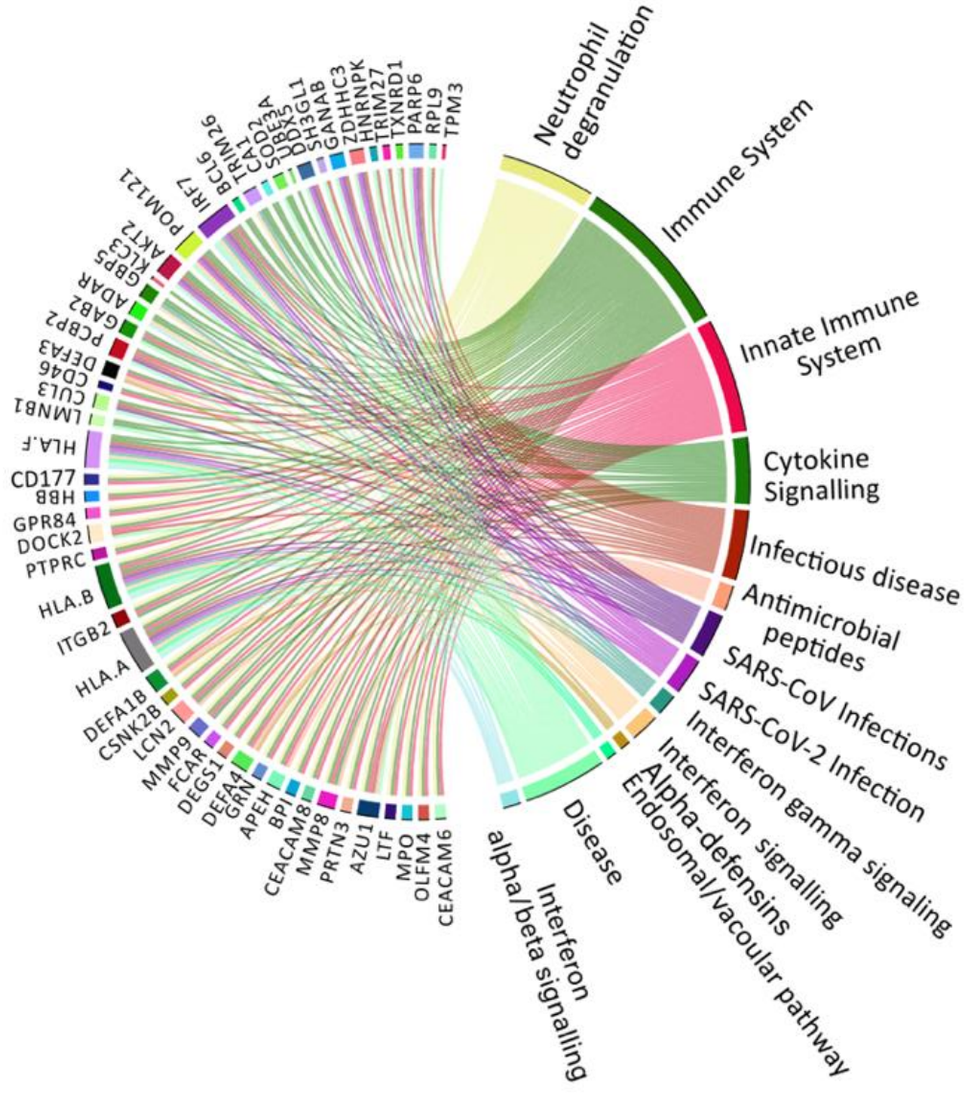
Editor's Summary:
Katagirya et al. conduct differential transcript expression analysis using bulk whole blood RNAseq from HIV-infected children with and without tuberculosis in Uganda, Botswana, and Eswatini. They identify distinct transcripts upregulated in tuberculosis infection by region, but find most transcripts lie within conserved pathways, most notably neutrophil degranulation.

Peer review information:
Communications Medicine thanks Gerrit Woltmann and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

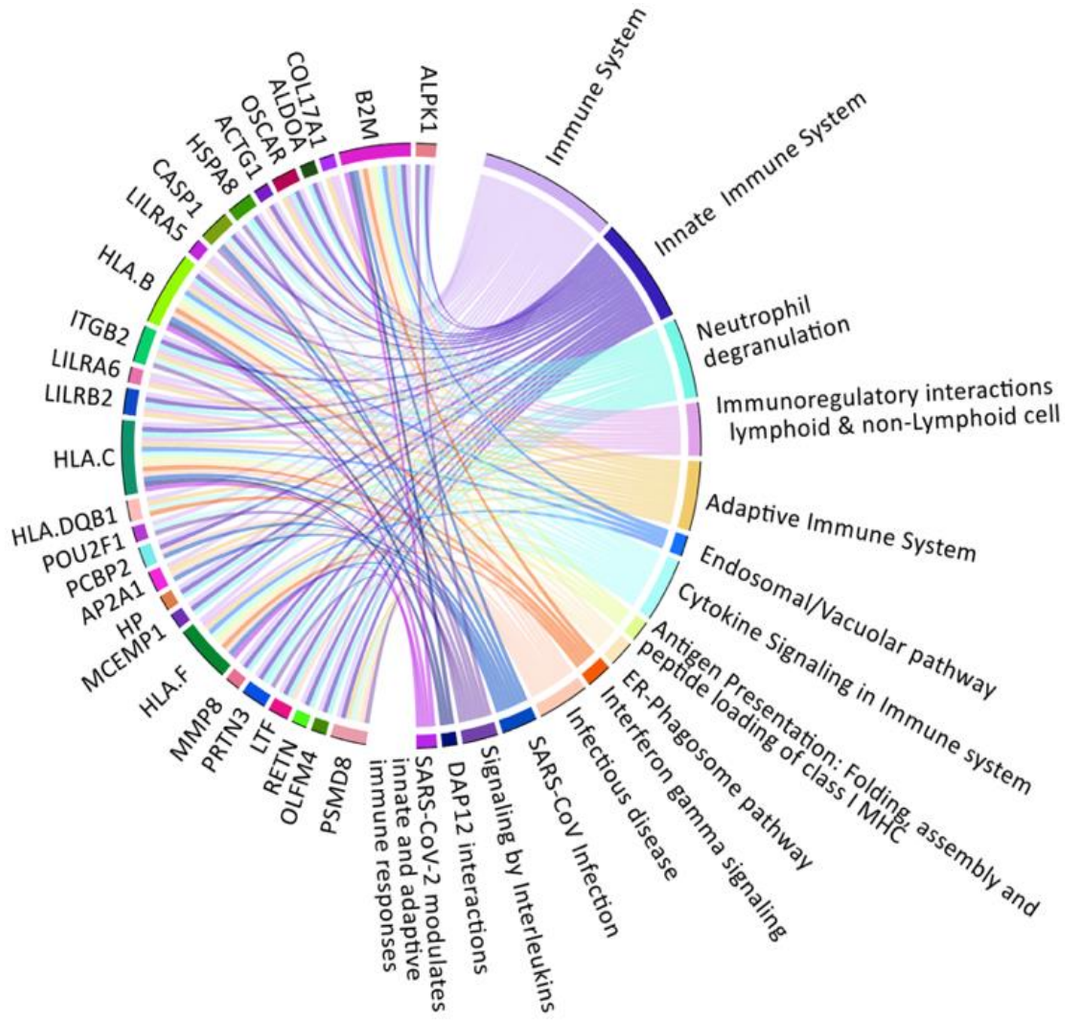

a
160,000
120,000
80,000
40,000
0
Transcript count
127,498
58,424
(45.8%)
144,359
79,306
(54.9%)
Southern Africa
Uganda
Key
Total
Filtered
b
Southern Africa
-Log10 q-value
10
5
0
-20
0
20
Log2 Fold Change
Uganda
-Log10 q-value
70
50
25
10
5
0
-20
0
20
Log2 Fold Change
c
Uganda
MDS 2
1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
1.0
1.5
MDS 1
Condition
NoTB
ATB
Southern Africa
MDS 2
1
0
-1
-2
-1
0
1
MDS 1
Condition
NoTB
ATB


Region
Region
Uganda
Southern_Africa
Transcript
Shared
Unique
2
1
0
−1
−2

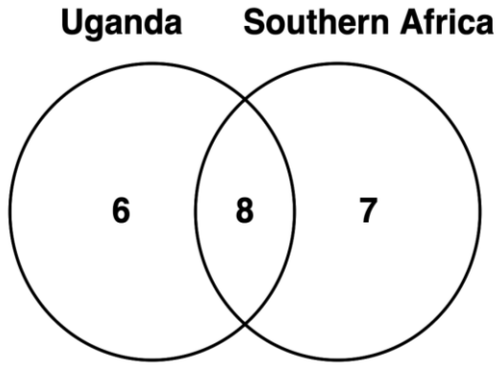
a **Uganda**



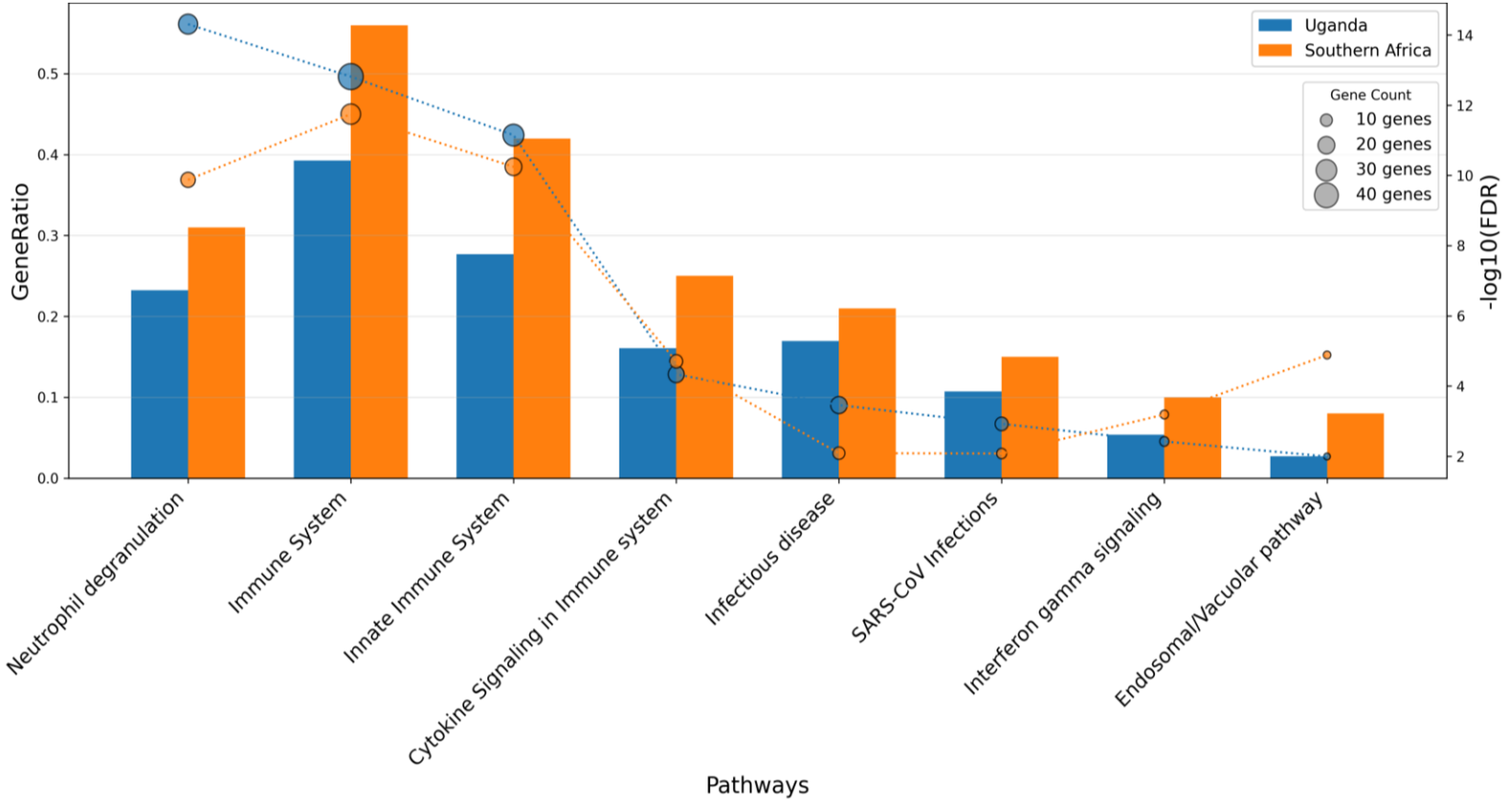
b **Southern Africa**

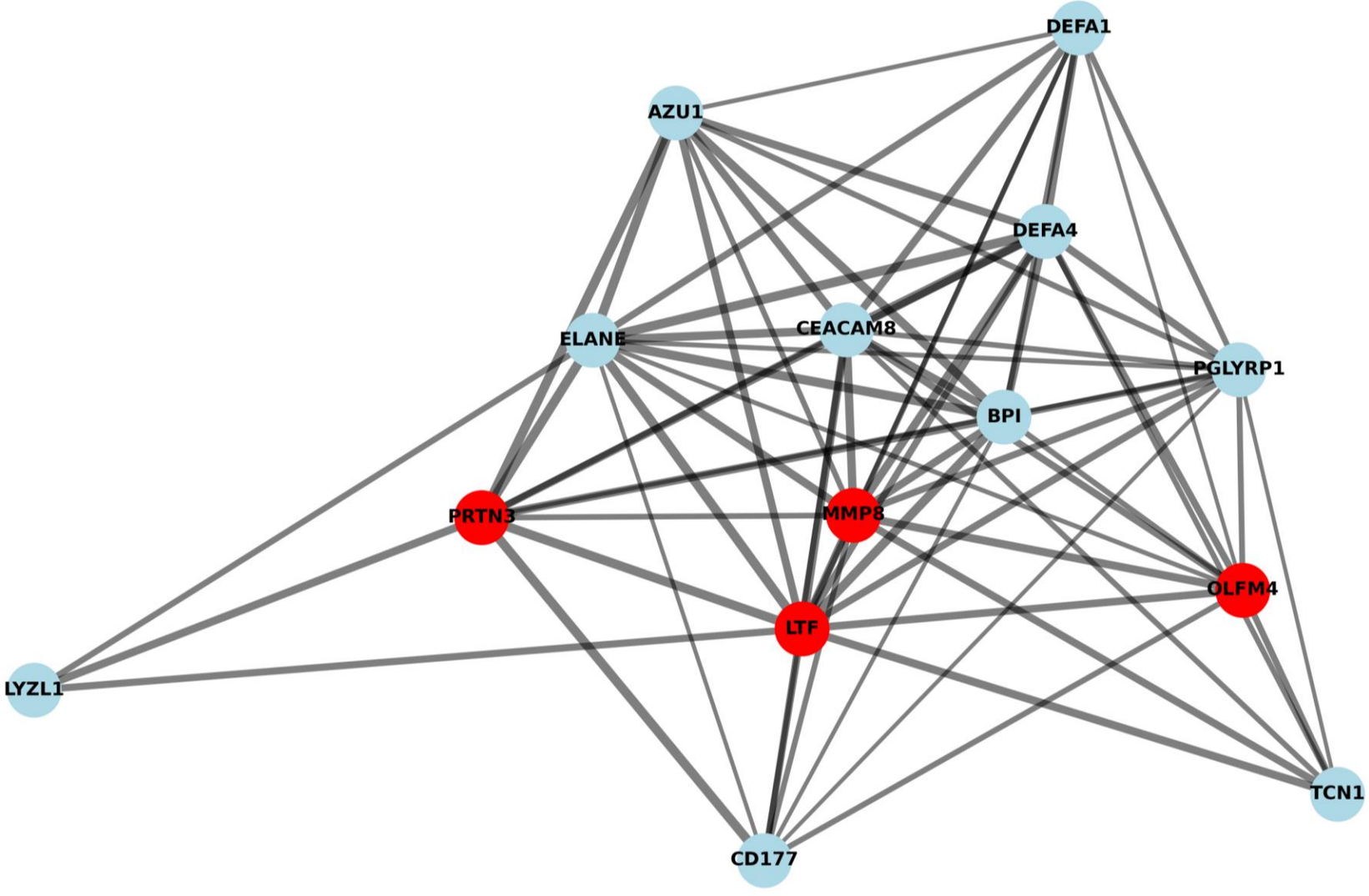


a


Uganda
Southern Africa
6
8
7

b


Uganda
Southern Africa
Gene Count
10 genes
20 genes
30 genes
40 genes
GeneRatio
-log10(FDR)
0.0
0.1
0.2
0.3
0.4
0.5
2
4
6
8
10
12
14
Neutrophil degranulation
Immune System
Innate Immune System
Cytokine Signaling in Immune system
Infectious disease
SARS-CoV Infections
Interferon gamma signaling
Endosomal/Vacuolar pathway
Pathways


DEFA1
AZU1
DEFA4
CEACAM8
ELANE
PGLYRP1
BPI
PRTN3
MMP8
OLFM4
LTF
LYZL1
TCN1
CD177