Biochemical Genetics
https://doi.org/10.1007/s10528-024-11002-1

ORIGINAL ARTICLE

Whole Blood vs Serum-Derived Exosomes for Host and Pathogen-Specific Tuberculosis Biomarker Identification: RNA-Seq-Based Machine-Learning Approach

Dhammika Magana-Arachchi[1] · Dushantha Madegedara[2] · Upeka Bandara[1]

Received: 18 November 2023 / Accepted: 10 December 2024

Abstract

Mycobacterium tuberculosis (Mtb) remains a leading infectious disease responsible for millions of deaths. RNA sequencing is a rapidly growing technique and a powerful approach to understanding host and pathogen cross-talks via transcriptional responses. However, its application is limited due to the high costs involved.This study is a preliminary attempt to understand host–pathogen cross-talk during TB infection in different TB clinical cohorts using two biological fluids: Whole blood and serum exosomes (EXO). We conducted an RNA-sequencing machine-learning approach using 20 active TB (ATB), 11 latent TB (LTB), three healthy control (HC) whole blood datasets, and two ATB, LTB, and HC serum EXO datasets. During the study, host-derived differentially expressed genes (DEGs) were identified in both whole blood and EXOs, while EXOs were successful in identifying pathogen-derived DEGs only in LTB. The majority of the DEGs in whole blood were up-regulated between ATB and HC, and ATB and LTB, while down-regulated between LTB and HC, which was vice versa for the EXOs, indicating different mechanisms in response to different states of TB infection across the two different biological samples. The pathway analysis revealed that whole blood gene signatures were mainly involved in host immune responses, whereas exosomal gene signatures were involved in manipulating the host's cellular responses and supporting Mtb survival. Overall, identifying both host and pathogen-derived gene signatures in different biological samples for intracellular pathogens like Mtb is vital to decipher the complex interplay between the host and the pathogen, ultimately leading to more successful future interventions.

Keywords Biomarkers · Exosomes · Machine learning · RNA sequencing · Tuberculosis · Whole blood

Extended author information available on the last page of the article

Published online: 23 December 2024

Introduction

Tuberculosis (TB) remains one of the deadliest diseases caused by a single bacterial pathogen, *Mycobacterium tuberculosis* (Mtb), leading to more than 1.5 million deaths annually (World Health Organization 2022). With the recent understanding of TB infection and disease, researchers have demonstrated that TB exists in a continuous spectrum: Active TB (ATB), latent TB (LTB), incipient TB (with no evidence of active Mtb), and subclinical TB (with radiological abnormalities and microbiological evidence of active Mtb) (Migliori et al. 2021). After the TB bacteria enter the body, it can either be eliminated by activating the innate immune system or persists for decades in a dormant state in asymptomatic individuals with no clinical symptoms by triggering the adaptive immune response (Chandra et al. 2022). The latter has resulted in 2–3 billion latent TB individuals (23%), with a 5–10% lifetime risk of disease reactivation (World Health Organization 2022).

The existing TB diagnostic tests focus on detecting Mtb or specific host immune responses using relatively faster, simpler, and cost-effective methods, such as tuberculin skin test (TST), interferon-gamma release assays (IGRAs) (Loureiro et al. 2019; Barker et al. 2023) or nucleic acid amplification tests (NAATs), most commonly the polymerase chain reaction (PCR)-based tests (Chin et al. 2018; Nandlal et al. 2022). However, these routine diagnostic tests are not specific for predicting disease progression (Pai and Sotgiu 2016) or distinguishing active disease from latent infection (Carranza et al. 2020). Therefore, ongoing research to develop diagnostic tools for predicting TB disease progression is much demanded to address these challenges.

The choice of biological fluid for biomarker identification in the context of TB diagnosis is indeed crucial. Blood is frequently chosen as a biological fluid and recommended by the World Health Organization (WHO) for TB biomarker identification due to its systemic nature, ease of collection, and availability in a wide range of individuals, including pediatric and adult populations, and more over immunocompromised individuals (Wallis et al. 2010; Denkinger et al. 2015; Goletti et al. 2016). Even though blood is suitable for identifying host biomarkers associated with TB disease and infection, there may be better choices due to the low abundance of pathogen molecules in the blood, thus, requiring a larger volume of samples (Banada et al. 2013). Considering this, exosomes (EXO) have shown promise for pathogen and host biomarker identification in infectious diseases, especially for intracellular pathogens like Mtb (Lv et al. 2017; Alipoor et al. 2019; Lyu et al. 2019). Extracellular vesicles are an emerging area of interest for diagnostic biomarker identification, therapeutics, and clinical applications (Sen et al. 2023).

Biomarker identification for TB diagnosis through RNA sequencing (RNA-seq) is a promising approach that can provide valuable insights into host and pathogen transcriptomic profiling to decipher the molecular mechanisms underlying this complex host–pathogen interaction (Lv et al. 2017; Sambarey et al. 2017; Leong et al. 2018; Singhania et al. 2018; Estévez et al. 2020; Madamarandawala

et al. 2023). These studies have shown that transcriptomic profiling in different stages of TB disease and infection can vary significantly. Therefore, identifying specific gene expression patterns associated with different TB disease and infection stages can aid in developing better diagnostic tools, biomarkers for treatment response, and novel therapeutic strategies for TB.

However, the complexity and cost of the technique currently limit the clinical application of RNA-seq for TB diagnosis (Martínez-Pérez et al. 2022). Nevertheless, RNA-seq-based machine-learning approaches offer promising solutions to address these challenges and make TB diagnosis more efficient, accurate, and globally accessible (Singhania et al. 2018). Therefore, this study aims to assess the suitability of both whole blood and, serum-derived EXOs to identify the pathogen and host-derived biomarkers in different clinical groups of TB and healthy controls using publicly available RNA-seq datasets, especially the first study to include locally obtained active TB RNA-seq dataset from Sri Lanka for a combined machine-learning approach. The outcomes will contribute to choosing between whole blood, and serum-derived EXOs for host and pathogen-specific biomarker identification, and potential pathway analysis involved in each stage of TB disease, thereby aiding the development of new diagnostic tools and therapeutics.

Materials and Methods

Data Acquisition

All the selected transcriptomic datasets represent the pre-COVID period. The whole blood RNA-sequencing datasets of three different cohorts, Spain/Mozambique, India, and Sri Lanka, were retrieved from public repositories of EMBL-EBI Accession number: E-MTAB-7830, 10 ATB, and 11 LTB from Spain/Mozambique (https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-7830, accessed on 01 November 2022); Gene Expression Omnibus (GEO) Accession number: GSE122485, four subjects of ATB, and three HC from India (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122485, accessed on 07 November 2022), and NCBI Bio project Accession number: PRJNA720487, 10 ATB from Sri Lanka (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA720487, accessed on 01 November 2022). Serum-derived EXOs: The only serum-derived exosomal RNA-sequencing dataset was retrieved from the GEO Accession number: GSE94907, two ATB, LTB, and HCs from China (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE94907, accessed on 01 November 2022) (Table 1).

Inclusion and Exclusion Criteria

Overall, the ATB datasets consisted of newly diagnosed pulmonary TB patients who had not initiated anti-TB treatment or were within the first 5 days of treatment. The LTB patients were healthy individuals who had been exposed to microbiologically confirm pulmonary TB cases and had a positive result on either the tuberculin skin

Table 1 Transcriptomic data analyzed during the study

Biological sample	Dataset	Country	Literature	Sample type (Number of samples)	Accession number
Serum EXOs	GSE94907	China	Lv et al. 2017	HC (2)	GSM2491733
					GSM2491734
				LTB (2)	GSM2491735
					GSM2491736
				ATB (2)	GSM2491737
					GSM2491738
Whole blood	PRJNA720487	Sri Lanka	Madamarandawala et al. 2023	ATB (10)	SAMN18651959
					SAMN18651961
					SAMN18651952
					SAMN18651954
					SAMN18651956
					SAMN18651957
					SAMN18651960
					SAMN18651958
					SAMN18651953
					SAMN18651955

Table 1 (continued)

Biological sample	Dataset	Country	Literature	Sample type (Number of samples)	Accession number
	E-MTAB-7830	Spain and Mozambique	Estevez et al. 2020	LTB (11)	SAMEA5546154
					SAMEA5546155
					SAMEA5546157
					SAMEA5546159
					SAMEA5546166
					SAMEA5546167
					SAMEA5546176
					SAMEA5546177
					SAMEA5546182
					SAMEA5546188
					SAMEA5546186
				ATB (10)	SAMEA5546247
					SAMEA5546248
					SAMEA5546250
					SAMEA5546253
					SAMEA5546255
					SAMEA5546266
					SAMEA5546273
					SAMEA5546284
					SAMEA5546285
					SAMEA5546288

Table 1 (continued)

Biological sample	Dataset	Country	Literature	Sample type (Number of samples)	Accession number
	GSE122485	India	Sambarey et al. 2017	HC (3)	SAMN10415812
					SAMN10415811
					SAMN10415827
				ATB (4)	SAMN10415826
					SAMN10415821
					SAMN10415817
					SAMN10415814

test (TST) and/or the interferon-gamma release assay (IGRA), and normal chest X-ray or computed tomography (CT) scan.

To account for false positivity due to exposure to environmental mycobacteria and BCG vaccination, individuals were selected based on two-step TST/IGRA strategy: Negative results on both the TST and the QuantiFERON TB-Gold In-Tube enzyme-linked immunosorbent assay (ELISA) or IGRA, as well as the absence of chest radiograph abnormalities or pulmonary symptoms (applied to the Indian and Chinese study cohorts). The healthy, uninfected individuals had negative TST or IGRA tests, regular chest CT, and no clinical evidence of any diseases. The exclusion criteria for ATB patients included prior anti-TB treatment or any immunocompromising conditions. For LTB individuals, exclusion criteria included clinical symptoms or evidence of ATB and other non-tuberculosis respiratory infections. A detailed overview of inclusion and exclusion criteria based on each study is summarized in Supplementary Material 1.

Computing Resource and Software Operating Environment for RNA-Seq Analysis

All analyses were performed on a computer equipped with 24 GB of random access memory (RAM) and a 7 core processor. For programs requiring an Ubuntu operating system, Ubuntu 22.04 LTS was installed on the machine (Pertea et al. 2016).

Data Processing and Statistical Analysis

The downloaded sequences were mapped against the human reference genome (GRCh38) and the Mtb reference genome (NC00962) using HISAT2 software (version 2.2.1). The transcript assembly and quantification were performed using StringTie (version 1.2.2). The output files (.gtf files) were then used to generate two CSV files containing count matrices for genes and transcripts using a Python script for differential expression analysis using DESEQ2. These CSV files were imported into R (version 4.1.3) and analyzed for differential expression analysis using DESEQ2 (version 2.0.1). Each output file was then sorted based on the q-value < 0.05 or p-value < 0.01 and a Log2FoldChange ($> \pm 0$) for further analysis using PyCharm (version 2023.2.1).

For host-specific pathway enrichment analysis, the list of genes with a p-value < 0.01 from each comparison group was analyzed using the online Reactome pathway database (https://reactome.org). The output files were used to construct the bar charts for the top-20 pathways using PyCharm (version 2023.2.1). Pathogen-specific Gene ontology (GO) enrichment analysis was performed using the PANTHER overrepresentation test (Gene and Consortium 2000) with the list of Mtb genes identified between LTB and HC (q-value < 0.05 and Log2FoldChange $> \pm 0$). P-values were calculated using Fisher's exact test adjusted for false discovery rate. The GO analysis categorized gene functions into biological process (BP), molecular function (MF), cellular component (CC), and PANTHER pathways. Only significantly enriched gene functions (p-value $< 0{\cdot}05$) were reported. STRING (https://string-db.org/) was used to predict protein–protein interactions and visualize

Table 2 Total number of differentially expressed host genes and statistically significant DEGs of whole blood and EXOs

Biological sample	Combinations	Total number of DEGs identified	Number of host-specific genes	
			q-value < 0.05	p-value < 0.01
Serum-derived EXOs	ATB_HC	11139	None	406
	LTB_HC	14084	None	381
	ATB_LTB	17288	None	209
Whole blood	ATB_HC	9330	16	222
	LTB_HC	9046	2422	3044
	ATB_LTB	9232	1	454

network connectivity (Szklarczyk et al. 2019). The network was based on evidence from experiments, curated databases, or prediction of co-expression and gene fusions with medium confidence score (≥ 0·4).

Results

Host-Specific Differential Expression Analysis

Host-specific differential expression analysis was conducted using RNA-seq datasets obtained from whole blood and serum-derived EXOs of ATB, LTB, and HC individuals. Pairwise comparisons were performed between all clinical groups to identify host-specific biomarkers using whole blood and serum-derived EXO datasets, separately. In the analysis, none of the clinical groups showed differentially expressed genes (DEGs) with a q-value < 0.05 for the EXO data, p-value < 0.01, Log_2FoldChange > ± 0 was chosen. For whole blood, the pairwise comparison was made between q-value < 0.05 and p-value < 0.01, Log_2FoldChange > ± 0 (Table 2). According to the results, the number of identified DEGs in EXOs was higher compared to the whole blood. However, whole blood showed more statistically significant host-specific DEGs, particularly in the comparison between LTB and HC, where 3044 genes were identified (p-value < 0.01). In contrast, the comparison between ATB and HC in whole blood recorded the fewest statistically significant DEGs, with 222 genes identified (p-value < 0.01) (Table 2).

When comparing the proportions of up-regulated and down-regulated DEGs across clinical groups, whole blood accounted for a higher proportion overall. Notably, the comparison between LTB and HC in whole blood revealed a substantial proportion of down-regulated DEGs (28.6%) (Table 3).

Moreover, the gene expression analysis for each clinical group is displayed using the volcano plots, as shown in Fig. 1. The top 10 differentially expressed up-regulated and down-regulated genes are presented in Table 4. According to the results, both whole blood and EXOs have distinct set of up-regulated and down-regulated DEGs for each comparison group.

Table 3 Statistically significant up-regulated and down-regulated DEGs and proportions compared to the total identified host DEGs of each pairwise comparison (p-value < 0.01)

Clinical group combinations	Up-regulated: Down-regulated DEGs	Proportions compared to the total identified DEGs
ATB_HC (Whole blood)	190:32	0.020:0.003
ATB_HC (EXOs)	137:269	0.012:0.024
LTB_HC (Whole blood)	458:2586	0.051:0.286
LTB_HC (EXOs)	237:144	0.017:0.010
ATB_LTB (Whole blood)	142:312	0.015:0.034
ATB_LTB (EXOs)	130:79	0.007:0.004

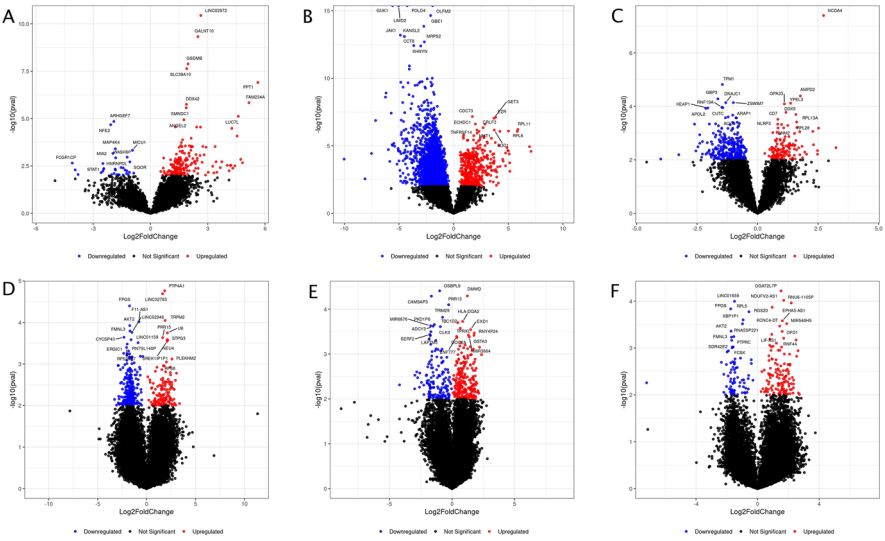

Fig. 1 Volcano plots showing the top 10 differentially up- and down-regulated host genes between **A** ATB and HC (Whole blood), **B** LTB and HC (Whole blood), **C** ATB and LTB (Whole blood), **D** ATB and HC (EXOs), **E** LTB and HC (EXOs), and **F** ATB and LTB (EXOs) (p-value < 0.01)

Host-Specific GO Enrichment Analysis

To study the different enrichment pathways involved with the significantly up-regulated or down-regulated host gene sets of all three categories in whole blood and EXOs, a host-specific pathway enrichment analysis was performed (Fig. 2). In here, between ATB and HC, the key up-regulated pathways in whole blood, involving with more than 100 genes, include "immune system, gene expression, metabolism of RNA, cytokine signaling in immune system, cellular responses to stress and stimuli, viral infection pathways and, adaptive immune system." The top-5 pathways involved in the down-regulated genes of EXOs are "transcriptional regulation by MECP2, GLI proteins bind promoters of hedgehog

Table 4 The top 10 differentially up- and down-regulated host genes in each pairwise comparison (p-value < 0.01)

Combinations	Up-regulated			Down-regulated		
	Genes	Log2FoldChange	p-value	Genes	Log2FoldChange	p-value
ATB_HC (Whole blood)	*LINC02972*	2.638079	3.64e-11	*ARHGEF7*	− 1.9203494	0.000014200
	GALNT10	2.486868	0.00e + 00	*NFE*	− 2.0841014	0.000021200
	GSDMB	1.969077	1.31e-08	*MAP4K4*	− 1.8408222	0.000393579
	SLC39A10	1.899683	2.33e-08	*MICU1*	− 0.9423223	0.000472138
	PPT1	5.629042	1.25e-07	*MIA2*	− 1.9694840	0.000688439
	FAM234A	5.162620	1.46e-06	*WASH6P*	− 1.2216312	0.001087896
	DDX42	1.885640	1.79e-06	*HNRNPDL*	− 1.8243478	0.001180342
	SMNDC1	1.871965	2.72e-06	*SQOR*	− 1.0813573	0.001933196
	LUC7L	4.600582	7.59e-06	*FCGR1CP*	− 4.0971384	0.002230691
	ANGEL2	1.750718	1.16e-05	*STAT1*	− 2.4938599	0.002365837
LTB_HC (Whole blood)	*CDC73*	1.716737	6.83e-08	*GUK1*	− 5.590494	0.00e + 00
	EZR	3.847007	7.77e-08	*POLD4*	− 4.251696	0.00e + 00
	GET3	3.686381	8.95e-08	*LIMD2*	− 5.039500	0.00e + 00
	ECHDC1	1.989226	2.44e-07	*OLFM2*	− 1.925163	0.00e + 00
	CRLF3	2.871493	2.48e-07	*GBE1*	− 2.104584	2.22e-15
	NMT1	2.697107	4.58e-07	*KANSL2*	− 2.727132	1.40e-14
	RPL11	5.881058	6.22e-07	*JAK1*	− 4.873800	6.31e-14
	TNFRSF14	2.124353	6.79e-07	*CCT8*	− 4.509240	8.05e-14
	SOD1	3.714139	6.82e-07	*MRPS2*	− 2.678784	2.01e-13
	RPL6	5.017380	8.09e-07	*KHNYN*	− 3.647608	3.76e-13

Table 4 (continued)

Combinations	Up-regulated			Down-regulated		
	Genes	Log2FoldChange	*p*-value	Genes	Log2FoldChange	*p*-value
ATB_LTB (Whole blood)	*NCOA4*	2.7280646	3.9900e-08	*TPM1*	− 1.455114	0.00001520
	AMPD2	1.7690452	4.0900e-05	*GBP3*	− 1.642097	0.00005240
	YPEL3	1.3644933	7.5900e-05	*ZSWIM7*	− 1.001890	0.00007240
	GPA33	1.1087074	8.0400e-05	*DNAJC1*	− 1.315815	0.00007260
	DDX5	1.1725643	1.6706e-04	*RNF19A*	− 1.474286	0.00010841
	RPL13A	1.5841461	2.0117e-04	*CUTC*	− 1.437116	0.00011195
	CD7	0.8382582	3.0423e-04	*KEAP1*	− 2.057370	0.00011495
	RPL28	1.6160524	3.8594e-04	*APOL2*	− 2.142757	0.00011799
	NLRP3	0.8336086	4.6414e-04	*ARAP1*	− 1.025691	0.00021318
	DNM2	0.9932769	4.7515e-04	*ACLY*	− 1.216808	0.00024734
ATB_HC (EXOs)	*PTP4A1*	1.847915	0.000017300	*FPGS*	− 1.7455318	0.000039900
	LINC02783	1.626193	0.000020300	*F11-AS1*	− 0.7867874	0.000095800
	TRPM2	1.890629	0.000089000	*AKT2*	− 1.7264732	0.000118994
	U6	2.124144	0.000177443	*FMNL3*	− 1.6971012	0.000160052
	PRR15	1.637857	0.000230735	*LINC02946*	− 1.4935614	0.000173049
	STPG3	2.101276	0.000260155	*CYCSP43*	− 2.3144285	0.000226750
	NEU4	2.157337	0.000266195	*LINC01159*	− 1.5894354	0.000260084
	SREK1IP1P1	2.079916	0.000285036	*RN7SL140P*	− 0.8738730	0.000306614
	PLEKHM2	2.586898	0.000758190	*ERGIC1*	− 1.9835315	0.000330224
	RPS8	2.025098	0.000892031	*RPS24P7*	− 2.0470518	0.000398680

Table 4 (continued)

Combinations	Up-regulated			Down-regulated		
	Genes	Log2FoldChange	*p*-value	Genes	Log2FoldChange	*p*-value
LTB_HC (EXOs)	*DMWD*	1.2167341	0.00005060	*OSBPL9*	− 0.9941647	0.00003900
	HLA-DQA2	0.8433164	0.00018789	*CAMSAP3*	− 1.6557196	0.00005130
	TBC1D2	0.4629914	0.00019973	*PRR13*	− 0.2734207	0.00007940
	TPRXL	0.6468851	0.00026604	*TRIM29*	− 0.7609004	0.00015112
	EXD1	1.4798408	0.00028551	*PKD1P6*	− 1.4150561	0.00022007
	RNY4P24	1.7757556	0.00034010	*MIR6876*	− 1.7142007	0.00023544
	DOC2A	1.3060375	0.00037840	*ADCY3*	− 1.5166258	0.00023933
	GSTA3	1.7510795	0.00038341	*CLK3*	− 0.9368490	0.00024523
	ZNF777	0.3618013	0.00040419	*LAPTM5*	− 1.6925257	0.00031742
	MIR3664	1.3736354	0.00041227	*SERF2*	− 1.7876184	0.00036809
ATB_LTB (EXOs)	*DGAT2L7P*	1.5245181	0.000060100	*LINC01659*	− 1.4951905	0.000100587
	NDUFV2-AS1	1.6978794	0.000095700	*FPGS*	− 1.7451342	0.000145699
	RNU6-1105P	2.1902738	0.000108796	*RPL5*	− 0.5585485	0.000167901
	RGS20	0.9486819	0.000135082	*XBP1P1*	− 0.9381111	0.000250013
	EPHA5-AS1	1.6088002	0.000261778	*RNA5SP221*	− 0.9810127	0.000303877
	MIR548H5	1.8958427	0.000301970	*AKT2*	− 1.7253005	0.000434696
	KCNC4-DT	1.4376205	0.000339367	*PTPRC*	− 1.5431894	0.000567897
	OFD1	1.6009472	0.000440694	*FMNL3*	− 1.6955506	0.000584822
	LIF-AS1	0.9203540	0.000510667	*SDR42E2*	− 1.6823817	0.000683951
	RNF44	1.6945005	0.000616399	*FCSK*	− 1.6243438	0.000945356

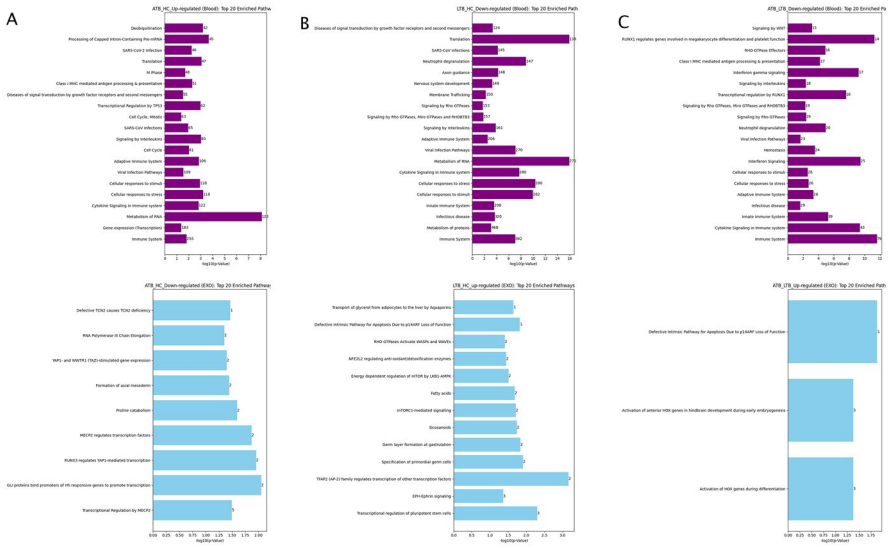

Fig. 2 Reactome pathway enrichment analysis of up-regulated and down-regulated host-specific genes differentiating whole blood and EXOs of **A** ATB vs HC, **B** LTB vs HC, and **C** ATB vs LTB

(Hh) responsive genes to promote transcription, RUNX3 regulation mediated by YAP1 transcription, MECP2 regulates transcription factor and proline catabolism" (Fig. 2A). Conversely, between LTB and HC, majority of the whole blood signatures which are up-regulated during the active disease are down-regulated in the latent stage (Fig. 2B). Between ATB and LTB, up-regulation of two important pathways is highlighted in EXOs, "activation of HOX genes during differentiation," and "defective intrinsic pathway for apoptosis due to p14ARF loss of function." Between ATB vs LTB, the main pathways involving interferon (IFN) gamma signaling, IFN signaling, cytokine signaling, neutrophil degranulation, cell differentiation, adaptive and innate immune system, WNT signaling, and RUNX1 transcriptional regulation are down-regulated in ATB compared to LTB (Fig. 2C).

Pathogen-Specific Differential Expression Analysis

According to the differential expression analysis for pathogen-specific biomarker identification, a total of 1377 statistically significant DEGs (q-value < 0.05) were identified in the EXO RNA-seq datasets between LTB and HC. Among them, 97 Mtb genes are up-regulated and 1280 down-regulated. Figure 3 shows the top 10 up-regulated and down-regulated Mtb genes. The top three up-regulated genes are *fmt* (tRNAfMet-formyl transferase), *PPE8* (PE family protein), and *PPE56* (PE family protein). The top three down-regulated genes are *rrs* (16S ribosomal RNA), *dnaA* (Chromosomal replication initiator protein), and *gyrB* (DNA gyrase subunit B) (Table 5). In addition, *rrs*, *rrl,* and *pks* genes were identified in comparisons between

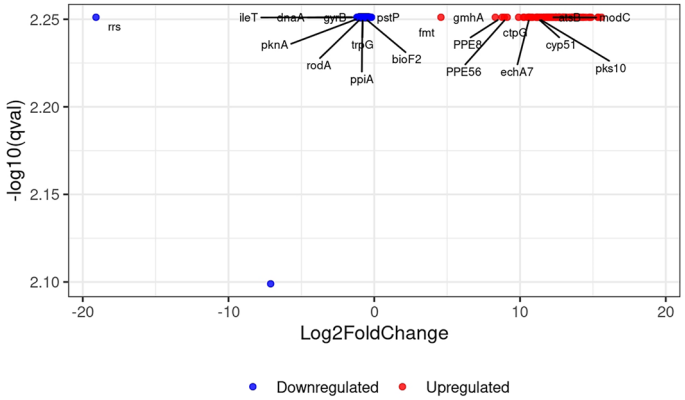

Fig. 3 Volcano plot showing top 10 differentially up-regulated and down-regulated pathogen-specific genes of serum-derived EXOs in the pairwise comparison between latent TB (LTB) and healthy individuals (HC) (q-value < 0.05)

Table 5 The top 10 differentially up- and down-regulated Mtb genes between LTB and HC (q-value < 0.05)

Up-regulated genes	Log2FoldChange	Down-regulated genes	Log2FoldChange
fmt	4.574438	*rrs*	− 19.0916090
PPE8	8.765720	*dnaA*	− 0.9833287
PPE56	9.119854	*gyrB*	− 0.9460011
ctpG	9.902101	*ileT*	− 1.0596897
gmhA	8.304650	*ppiA*	− 0.8084603
atsB	11.158690	*trpG*	− 0.7947428
pks10	11.140646	*pknA*	− 0.8765491
echA7	10.654208	*rodA*	− 0.8280315
cyp51	11.127032	*pstP*	− 0.7069693
modC	11.794122	*bioF2*	− 0.7192605

ATB and HC individuals; however, these results were not statistically significant (Supplementary material 2). Notably, no statistically significant pathogen-specific DEGs were observed in whole blood for any of the clinical groups.

Pathogen-Specific Gene Ontology (GO) Enrichment Analysis

The pathogen-specific GO enrichment analysis between LTB and HC showed six enrichment terms for biological processes. The top three enrichment terms were cellular process, biological process, and cellular metabolic process (false discovery rate [FDR] < 0.004). For molecular function (MF), 11 terms were identified, with the most enriched being ion binding, molecular function, carbohydrate derivative

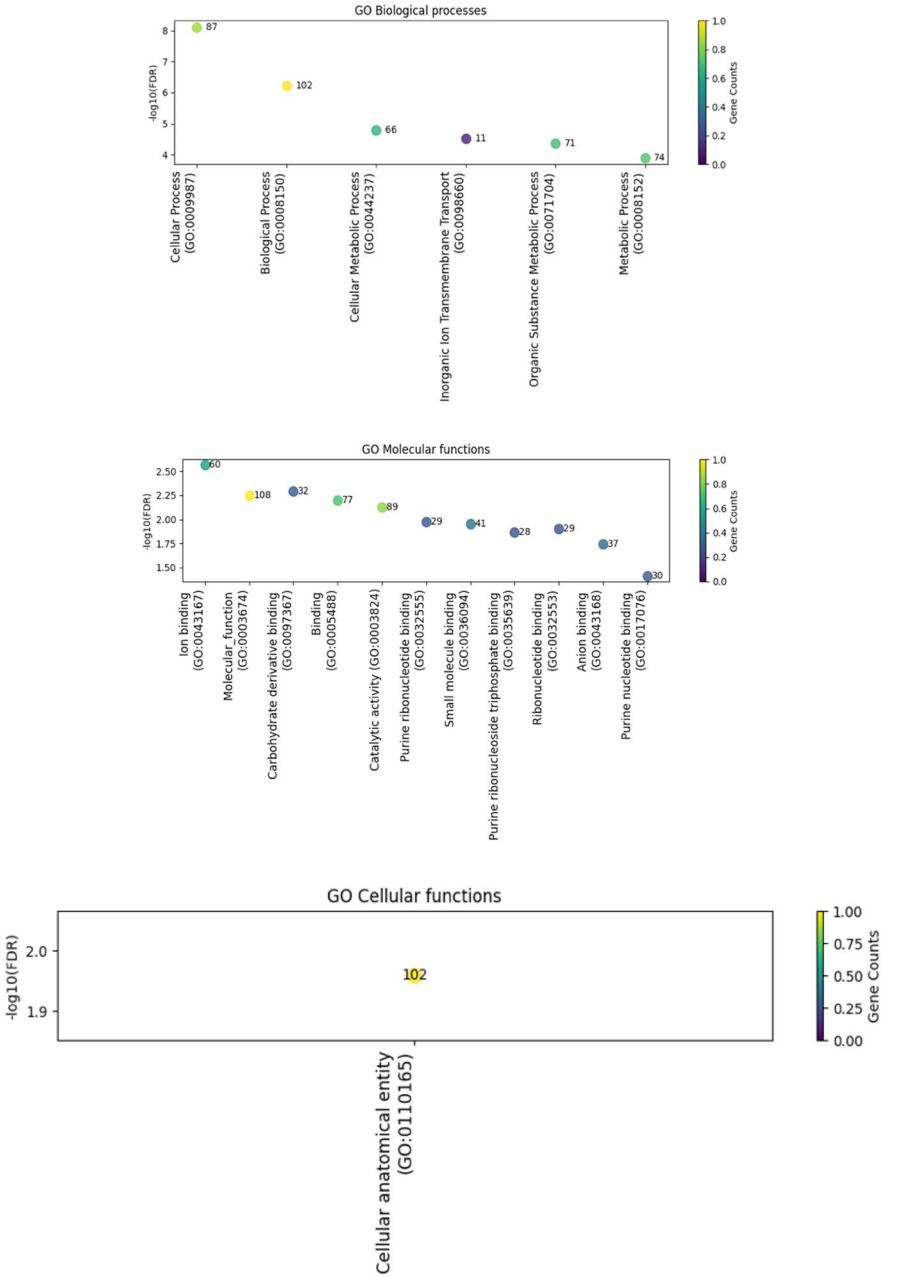

Fig. 4 Pathogen-specific GO enrichment analysis between latent TB (LTB) and healthy individuals (HC) serum-derived EXOs for biological process (BP), molecular function (MF), cellular component (CC) (p-value < 0.05)

binding, binding, and catalytic activity (FDR < 0.007). For cellular component (CC), only a single term was obtained: Cellular anatomical entity (FDR < 0.01) (Fig. 4). Moreover, PANTHER pathway analysis revealed a single pathway correlated with the respective proteins: Mycobacterial arginine biosynthesis (p-value < 0.05) (Table 6).

Protein–Protein Interaction (PPI) Network Analysis

The PPI network of up-regulated Mtb genes obtained for LTB vs HC using the STRING database revealed that the extracted interactome consisted of 97 nodes and 148 edges. The average node degree and average local clustering coefficient of the network were 3.05 and 0.394, respectively. The BP, MF, and CC GO enrichment analysis showed similar results to the PANTHER GO enrichment analysis (Fig. 5, Left). For the down-regulated Mtb genes, the network comprised 42 nodes and 25 edges, with an average node degree of 1.19 and an average local clustering coefficient of 0.352. None of the identified Mtb proteins were categorized under BP, MF, or CC GO enrichment terms. However, functional enrichment was observed for 9 genes classified as "secreted (KW-0964)" based on UniProt annotation. These genes *ompA*, *PPE57*, *esxW*, *VapC39*, *lipU*, *pstS1*, *PE13*, *VapC38*, and *VapC44* were categorized under CC enrichment within the term extracellular region (GO:0005576) (Fig. 5, Right, and Supplementary Material 2).

Discussion

This study primarily aims to compare the suitability of two biological samples, whole blood and serum-derived EXOs, for identifying host and pathogen-derived biomarkers across clinical groups (ATB vs HC, LTB vs HC, and ATB vs LTB). This also explores distinct mechanisms functioning in these two biological fluids in relation to host–pathogen interactions. Additionally, the study provides valuable insights into the advantages and limitations of each sample type for future TB biomarker identification.

A comparison of the DEGs revealed clear variation in gene up-regulation and down-regulation depending on TB disease stages and the biological fluids analyzed. Between ATB and HC, most host DEGs were up-regulated in whole blood but down-regulated in EXOs. These differences in gene expression patterns likely reflect the complex host–pathogen interactions during TB infection. Consistent with previous studies, the majority of up-regulated genes in whole blood were mainly involved in immune response and cellular regulation pathways, including "metabolism of RNA, cellular responses to stress, signaling by interleukins, adaptive immune system and cytokine signaling in immune system."

Interestingly, transcriptional regulation by methyl-CpG-binding protein 2 (MeCP2) and proline catabolism pathways were down-regulated in EXOs during the active stage. MeCP2 is a crucial protein that regulates gene expression through its

Table 6 PANTHER pathway analysis for the pathogen-specific DEGs of serum-derived EXOs in the pairwise comparison between latent TB (LTB) and healthy individuals (HC) (*p-value < 0.05)

PANTHER pathways	Identified genes	Fold Enrichment	P-value	FDR
Arginine biosynthesis (P02728)	3 (argD, arB, dapE)	11.44	0.0043*	0.411
Gonadotropin-releasing hormone receptor pathway (P06664)	1 (ctaC)	30.5	0.0625	1
Cholesterol biosynthesis (P00014)	1 (idsB)	30.5	0.0625	1
Vitamin D metabolism and pathway (P04396)	1 (fprB)	15.25	0.0923	1
Asparagine and aspartate biosynthesis (P02730)	1 (asnB)	7.62	0.149	1

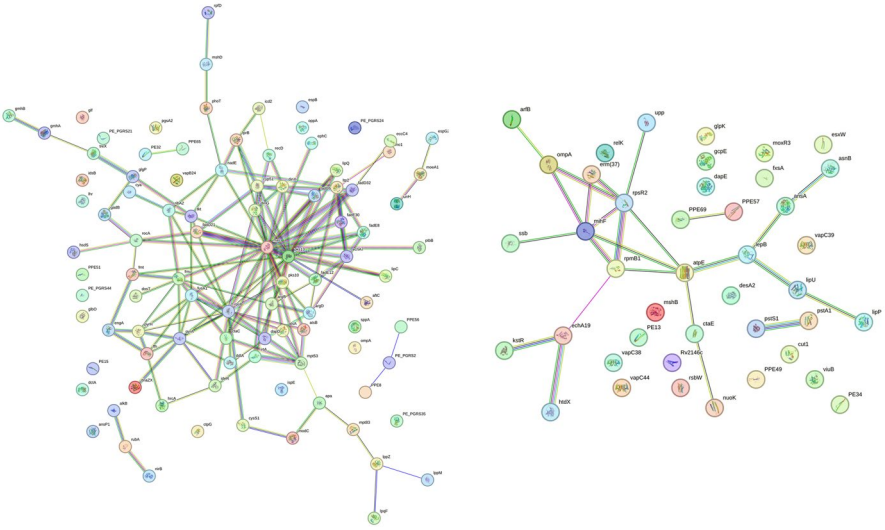

Fig. 5 PPI network obtained and visualized by STRING v12.0 for 97 up-regulated (left) and 42 down-regulated (right) Mtb genes in serum-derived EXOs between LTB and HC individuals

interaction with methylated DNA (Zalosnik et al. 2021). Given its role immune cells and cytokine production (Zalosnik et al. 2021), alterations in MeCP2 could influence immune function during TB infection. Previous studies have reported that Mtb infection can induce rapid alterations in DNA methylation, especially in leukocyte subsets, thereby compromising the host immune system's ability to respond effectively to mycobacteria (Chen et al. 2014; Frantz et al. 2019; DiNardo et al. 2020; Qin et al. 2021). Thus, MeCP2-linked DNA methylation may significantly impact the host's immune responsiveness to Mtb. However, further studies are required to elucidate the specific pathways and genes regulated by MeCP2 in immune activation. Another down-regulated pathway in EXOs was proline dehydrogenase 2 (*PRODH2*), also known as hydroxyproline dehydrogenase, which is specifically involved in hydroxyproline catabolism. Hydroxyproline is associated with hypoxia-inducible factor 1 alpha (HIF-1α), a critical transcription factor that regulates the body's response to low oxygen levels. Under hypoxic conditions, HIF-1α activation promotes cellular adaptation and survival by inducing the expression of genes that facilitate adaptation to low oxygen. However, down-regulation of this gene during the active stage under normoxic condition may enhance the bacterial killing during the early stage of TB infection (Phang 2023).

Our results show that, between LTB and HC, the majority of whole blood signatures that are up-regulated during the active disease are down-regulated in the latent stage. In contrast, EXOs show significant up-regulation of biological pathways associated with latency. The persistence of Mtb within the host for extended periods during dormancy, without triggering a robust immune response and resuscitation from multiple body sites, remains unclear. Previous studies revealed that both pluripotent hematopoietic stem cells (HpSCs) and mesenchymal stem cells (MSCs) provide

a niche for Mtb, allowing the bacterium to remain dormant for prolonged periods (Tornack et al. 2017; Mayito et al. 2019). Supporting the notion, our study observed the up-regulation of the *POU5F1* gene in EXOs during the latent stage, which regulates the transcriptional regulation of pluripotent stem cells pathway. These findings suggest that pluripotent stem cells may contribute to the long-term persistence of Mtb during dormancy and facilitates its dissemination through their migratory and differentiation capabilities. MSCs are also implicated in granuloma formation during latency by creating a favorable microenvironment for dormant Mtb bacilli and inhibiting the T-cell function, thereby supporting bacterial survival (Devi et al. 2023). The erythropoietin-producing hepatocellular carcinoma (Eph) receptors and their corresponding ephrin ligands play critical roles in cell signaling, migration, and tissue development (Darling and Lamb 2019). Our findings align with prior research demonstrating that Mtb can manipulate host pathways such as Eph signaling to support granuloma formation, enhancing its survival mechanism (Khounlothm et al. 2009).

Our pathway analysis revealed up-regulation of the TFAP2 transcription factor in EXOs. Members of the TFAP2 family modulate the expression of lipid droplet proteins and promote lipid droplet (LD) accumulation within cells (Scott et al. 2018). In agreement with that, Guirado et al. (2015) reported that greater LD accumulation in LTB granulomas compared to healthy individuals. Host fatty-acid pathway up-regulation in our study suggests that macrophages store neutral lipids as LDs, which can serve as an energy source for Mtb while simultaneously functioning as a host-defense mechanism against intracellular pathogens like Mtb (Guirado et al. 2015; Knight et al. 2018). LD accumulation may also enhance eicosanoid production-bioactive lipid mediators involved in the balance between pro- and anti-inflammatory immune response (Knight et al. 2018). This duality allows Mtb to manipulate eicosanoid biosynthesis for its survival and spread (Todorova et al. 2023). Thus, LDs represent a complex evolutionary interface between host defense and pathogen exploitation.

The up-regulation of the Mtb gene *fmt* (tRNAfMet-formyl transferase) in LTB patients is noteworthy due to *fmt*'s essential role in initiating protein synthesis, a critical process for bacterial growth and replication. While *fmt* is not universally essential across bacterial species, its deletion in Mtb has been linked to slower growth rates, which may still allow intracellular survival and disease progression (Vanunu et al. 2017). Our study supports this notion by showing *fmt* up-regulation in LTB patients, indicating that the mycobacteria are active, replicating, and surviving intracellularly. This up-regulation in LTB patients suggests that the bacteria are in a state that necessitates active protein synthesis, supporting their intracellular persistence and pathogenicity. Here, we identify, *PPE8* and *PPE56*, two PE/PPE family proteins up-regulated in LTB individuals. PE/PPE family proteins of Mtb are known to be involved in host immune modulation and immune evasion in favor of its survival within the host (D'Souza et al. 2023). Previous studies have also shown that *PPE8* and *PPE56* are involved with drug resistance of Mtb (Murcia et al. 2010; Gómez-González et al. 2023), therefore, worth studying the contribution to immune evasion and drug resistance in latently infected individuals.

Blood transcriptional profiling revealed that cytokine signaling, WNT signaling, and cell differentiation were down-regulated in ATB compared to LTB (Estévez et al. 2020). In contrast to Estevez et al., our study identified down-regulation of IFN gamma signaling and IFN signaling pathways in ATB. Typically, IFN-related gene signatures are elevated in incipient LTB cases with more viable bacteria (Burel et al. 2021). However, down-regulation of the WNT pathway in severe PTB cases may impair T-cell proliferation and activation, subsequently suppressing IFN pathways (Fan et al. 2015, 2017). These findings underscore the importance of considering disease stage when identifying biomarkers.

Our PPI network analysis for LTB vs HC indicates that, although Mtb remains dormant, it is metabolically active, performing low levels of cellular functions such as ion, nucleotide, and small molecule binding, ATP binding, and catalytic activity. The down-regulation of extracellular secretion of certain Mtb-derived molecules suggests that Mtb manipulates intracellular survival mechanisms during latency (López-Agudelo et al. 2022; Bo et al. 2023).

A limitation of our study is the relatively small sample size of exosomal RNA-sequencing datasets compared to whole blood. Future studies should prioritize EXO-related tuberculosis biomarker identification to enhance the validity of these findings. Despite this limitation, our results highlight the utility of whole blood and EXOs in TB biomarker identification, particularly for molecular diagnostics in challenging populations such as children and immunocompromised individuals. Our findings suggest that EXOs may be more effective in detecting viable mycobacteria in latently infected individuals, making them a promising tool for TB biomarker identification and triage in these cases.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10528-024-11002-1.

Author Contributions DM and WB contributed to the study conception and design. DM contributed to select the study population and sample collection. WB engaged in sample collection. DM and WB implemented the study. DM, WB and DM contributed in preparation of the manuscript. All authors have reviewed the manuscript and consent was given to publish.

Funding The research was funded by the National Institute of Fundamental Studies, Sri Lanka.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflicts of interest The authors declare no competing interests.

References

Alipoor SD, Tabarsi P, Varahram M et al (2019) Serum exosomal miRNAs are associated with active pulmonary tuberculosis. Dis Markers. https://doi.org/10.1155/2019/1907426

Banada PP, Koshy R, Alland D (2013) Detection of *Mycobacterium tuberculosis* in blood by use of the Xpert MTB/RIF assay. J Clin Microbiol 51:2317–2322. https://doi.org/10.1128/JCM.00332-13

Barker E, Moss J, Holmes H et al (2023) A cost-effectiveness evaluation of latent tuberculosis infection screening of a migrant population in Malaysia. Sci Rep 13:1–11. https://doi.org/10.1038/s41598-023-29648-z

Bo H, Moure UAE, Yang Y et al (2023) *Mycobacterium tuberculosis*-macrophage interaction: molecular updates. Front Cell Infect Microbiol 13:1–14. https://doi.org/10.3389/fcimb.2023.1062963

Burel JG, Singhania A, Dubelko P et al (2021) Distinct blood transcriptomic signature of treatment in latent tuberculosis infected individuals at risk of developing active disease. Tuberculosis 131:102127. https://doi.org/10.1016/j.tube.2021.102127

Carranza C, Pedraza-Sanchez S, de Oyarzabal-Mendez E, Torres M (2020) Diagnosis for latent tuberculosis infection: new alternatives. Front Immunol 11:1–13. https://doi.org/10.3389/fimmu.2020.02006

Chandra P, Grigsby SJ, Philips JA (2022) Immune evasion and provocation by *Mycobacterium tuberculosis*. Nat Rev Microbiol 20:750–766. https://doi.org/10.1038/s41579-022-00763-4

Chen YC, Hsiao CC, Chen CJ et al (2014) Aberrant Toll-like receptor 2 promoter methylation in blood cells from patients with pulmonary tuberculosis. J Infect 69:546–557. https://doi.org/10.1016/j.jinf.2014.08.014

Chin KL, Sarmiento ME, Norazmi MN, Acosta A (2018) DNA markers for tuberculosis diagnosis. Tuberculosis 113:139–152. https://doi.org/10.1016/j.tube.2018.09.008

D'Souza C, Kishore U, Tsolaki AG (2023) The PE-PPE family of *Mycobacterium tuberculosis*: proteins in disguise. Immunobiology 228:152321. https://doi.org/10.1016/j.imbio.2022.152321

Darling TK, Lamb TJ (2019) Emerging roles for Eph receptors and ephrin ligands in immunity. Front Immunol 10:1–15. https://doi.org/10.3389/fimmu.2019.01473

Denkinger CM, Kik SV, Cirillo DM et al (2015) Defining the needs for next generation assays for tuberculosis. J Infect Dis 211:S29–S38. https://doi.org/10.1093/infdis/jiu821

Devi A, Pahuja I, Singh SP et al (2023) Revisiting the role of mesenchymal stem cells in tuberculosis and other infectious diseases. Cell Mol Immunol 20:600–612. https://doi.org/10.1038/s41423-023-01028-7

DiNardo AR, Rajapakshe K, Nishiguchi T et al (2020) DNA hypermethylation during tuberculosis dampens host immune responsiveness. J Clin Invest 130:3113–3123. https://doi.org/10.1172/JCI134622

Estévez O, Anibarro L, Garet E et al (2020) An RNA-seq based machine learning approach identifies latent tuberculosis patients with an active tuberculosis profile. Front Immunol 11:1–12. https://doi.org/10.3389/fimmu.2020.01470

Fan L, Xiao H, Mai G et al (2015) Impaired M. tuberculosis antigen-specific IFN-γ response without IL-17 enhancement in patients with severe cavitary pulmonary tuberculosis. PLoS ONE 10:1–19. https://doi.org/10.1371/journal.pone.0127087

Fan L, Shen H, Huang H et al (2017) Impairment of Wnt/β-catenin signaling in blood cells of patients with severe cavitary pulmonary tuberculosis. PLoS ONE 12:1–14. https://doi.org/10.1371/journal.pone.0172549

Frantz FG, Castro RC, Fontanari C et al (2019) DNA Methylation impairs monocyte function in tuberculosis leading to disease progression. J Immunol 202:10

Gene T, Consortium O (2000) Gene ontology: tool for the. Gene Expr 25:25–29. https://doi.org/10.1038/75556

Goletti D, Petruccioli E, Joosten SA, Ottenhoff THM (2016) Tuberculosis biomarkers: from diagnosis to protection. Infect Dis Rep 8:24–32. https://doi.org/10.4081/idr.2016.6568

Gómez-González PJ, Grabowska AD, Tientcheu LD et al (2023) Functional genetic variation in pe/ppe genes contributes to diversity in *Mycobacterium tuberculosis* lineages and potential interactions with the human host. Front Microbiol 14:1–12. https://doi.org/10.3389/fmicb.2023.1244319

Guirado E, Mbawuike U, Keiser TL et al (2015) Characterization of host and microbial determinants in individuals with latent tuberculosis infection using a human granuloma model. Mbio. https://doi.org/10.1128/mBio.02537-14

Khounlothm M, Subbian S, Smith R III et al (2009) *Mycobacterium tuberculosis* interferes with the response to infection by inducing the host iEphA2 receptor. J Infect Dis 199:1797–1806. https://doi.org/10.1086/599096

Knight M, Braverman J, Asfaha K et al (2018) Lipid droplet formation in *Mycobacterium tuberculosis* infected macrophages requires IFN-γ/HIF-1α signaling and supports host defense. PLoS Pathog 14:1–26. https://doi.org/10.1371/journal.ppat.1006874

Leong S, Zhao Y, Joseph NM et al (2018) Existing blood transcriptional classifiers accurately discriminate active tuberculosis from latent infection in individuals from south India. Tuberculosis 109:41–51. https://doi.org/10.1016/j.tube.2018.01.002

López-Agudelo VA, Baena A, Barrera V et al (2022) Dual RNA sequencing of mycobacterium tuberculosis-infected human splenic macrophages reveals a strain-dependent host-pathogen response to infection. Int J Mol Sci. https://doi.org/10.3390/ijms23031803

Loureiro RB, Maciel ELN, Caetano R et al (2019) Cost-effectiveness of QuantiFERON TB-Gold In-Tube versus tuberculin skin test for diagnosis and treatment of Latent Tuberculosis Infection in primary health care workers in Brazil. PLoS ONE 14:1–24. https://doi.org/10.1371/journal.pone.0225197

Lv L, Li C, Zhang X et al (2017) RNA profiling analysis of the serum exosomes derived from patients with active and latent *Mycobacterium tuberculosis* infection. Front Microbiol 8:1–10. https://doi.org/10.3389/fmicb.2017.01051

Lyu L, Zhang X, Li C et al (2019) Small RNA profiles of serum exosomes derived from individuals with latent and active tuberculosis. Front Microbiol 10:1–10. https://doi.org/10.3389/fmicb.2019.01174

Madamarandawala P, Rajapakse S, Gunasena B et al (2023) A host blood transcriptional signature differentiates multi-drug/rifampin-resistant tuberculosis (MDR/RR-TB) from drug susceptible tuberculosis: a pilot study. Mol Biol Rep 50:3935–3943. https://doi.org/10.1007/s11033-023-08307-6

Martínez-Pérez A, Estévez O, González-Fernández Á (2022) Contribution and future of high-throughput transcriptomics in battling tuberculosis. Front Microbiol 13:1–16. https://doi.org/10.3389/fmicb.2022.835620

Mayito J, Andia I, Belay M et al (2019) Anatomic and cellular niches for *Mycobacterium tuberculosis* in latent tuberculosis infection. J Infect Dis 219:685–694. https://doi.org/10.1093/infdis/jiy579

Migliori GB, Ong CWM, Petrone L et al (2021) The definition of tuberculosis infection based on the spectrum of tuberculosis disease. Breathe 17:1–12. https://doi.org/10.1183/20734735.0079-2021

Murcia MI, Manotas M, Jiménez YJ et al (2010) First case of multidrug-resistant tuberculosis caused by a rare "Beijing-like" genotype of *Mycobacterium tuberculosis* in Bogotá, Colombia. Infect Genet Evol 10:678–681. https://doi.org/10.1016/j.meegid.2010.03.010

Nandlal L, Perumal R, Naidoo K (2022) Rapid molecular assays for the diagnosis of drug-resistant tuberculosis. Infect Drug Resist 15:4971–4984. https://doi.org/10.2147/IDR.S381643

Pai M, Sotgiu G (2016) Diagnostics for latent TB infection: Incremental, not transformative progress. Eur Respir J 47:704–706. https://doi.org/10.1183/13993003.01910-2015

Pertea M, Kim D, Pertea GM et al (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11:1650–1667. https://doi.org/10.1038/nprot.2016.095

Phang JM (2023) The regulatory mechanisms of proline and hydroxyproline metabolism: recent advances in perspective. Front Oncol 12:1–6. https://doi.org/10.3389/fonc.2022.1118675

Qin W, Scicluna BP, van der Poll T (2021) The role of host cell DNA methylation in the immune response to bacterial infection. Front Immunol 12:1–17. https://doi.org/10.3389/fimmu.2021.696280

Sambarey A, Devaprasad A, Mohan A et al (2017) Unbiased identification of blood-based biomarkers for pulmonary tuberculosis by modeling and mining molecular interaction networks. EBioMedicine 15:112–126. https://doi.org/10.1016/j.ebiom.2016.12.009

Scott CC, Vossio S, Rougemont J, Gruenberg J (2018) TFAP2 transcription factors are regulators of lipid droplet biogenesis. Elife 7:1–24. https://doi.org/10.7554/eLife.36330

Sen S, Xavier J, Kumar N et al (2023) Exosomes as natural nanocarrier-based drug delivery system: recent insights and future perspectives. 3 Biotech. https://doi.org/10.1007/s13205-023-03521-2

Singhania A, Verma R, Graham CM et al (2018) A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. Nat Commun. https://doi.org/10.1038/s41467-018-04579-w

Szklarczyk D, Gable AL, Lyon D et al (2019) STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47:D607–D613. https://doi.org/10.1093/nar/gky1131

Todorova Y, Emilova R, Milanov V et al (2023) Eicosanoid and cytokine levels differentiate between stages of MTB infection. Folia Med (Plovdiv) 65:399–406. https://doi.org/10.3897/folmed.65.e80599

Tornack J, Reece ST, Bauer WM et al (2017) Human and mouse hematopoietic stem cells are a depot for dormant *Mycobacterium tuberculosis*. PLoS ONE 12:1–18. https://doi.org/10.1371/journal.pone.0169119

Vanunu M, Lang Z, Barkan D (2017) The gene fmt, encoding tRNAfMet-formyl transferase, is essential for normal growth of *M. bovis*, but not for viability. Sci Rep 7:1–8. https://doi.org/10.1038/s41598-017-15618-9

Wallis RS, Pai M, Menzies D et al (2010) Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice. Lancet 375:1920–1937. https://doi.org/10.1016/S0140-6736(10)60359-5

World Health Organization (2022) Global Tuberculosis Report

Zalosnik MI, Fabio MC, Bertoldi ML et al (2021) MeCP2 deficiency exacerbates the neuroinflammatory setting and autoreactive response during an autoimmune challenge. Sci Rep 11:1–15. https://doi.org/10.1038/s41598-021-90517-8

Authors and Affiliations

Dhammika Magana-Arachchi[1] · Dushantha Madegedara[2] · Upeka Bandara[1]

✉ Dhammika Magana-Arachchi
dhammika.ma@nifs.ac.lk

Dushantha Madegedara
dmadegedara@yahoo.com

Upeka Bandara
upekaay@gmail.com

[1] National Institute of Fundamental Studies, Kandy, Sri Lanka

[2] Respiratory Disease Treatment Unit, General Teaching Hospital, Kandy, Sri Lanka