World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Transcriptomic profiling of host-pathogen interactions during early *Mycobacterium tuberculosis* infection stages

Chiamaka Francisca Igweonu [1,*] and Babatunde Owolabi [2]

[1] *Department of Biological Sciences, Western Illinois University, USA.*
[2] *Lagos State Ministry of Health (Grant Management Unit), Lagos, Nigeria.*

World Journal of Advanced Research and Reviews, 2024, 22(01), 2054–2073

Publication history: Received on 24 February 2024; revised on 26 April 2024; accepted on 29 April 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.22.1.1055

Abstract

Understanding the early molecular interactions between *Mycobacterium tuberculosis* (M.tb) and host immune cells is essential to identifying novel diagnostic markers and therapeutic targets for tuberculosis (TB). Despite advances in TB research, the global burden of the disease remains substantial, with millions of new infections annually, many in resource-constrained settings. The pathogen's ability to persist within host macrophages during initial infection stages underlies its pathogenicity and resistance to clearance. This study aims to elucidate transcriptomic changes in both host and pathogen during early stages of M.tb infection using high-throughput RNA sequencing (RNA-Seq). Human peripheral blood-derived macrophages were infected with virulent M.tb H37Rv, and dual RNA-Seq was performed at multiple early time points (2h, 6h, and 24h post-infection). Bioinformatic analysis revealed dynamic expression changes in host innate immune genes, including early upregulation of TNF, IL-6, and type I interferon pathways. Concurrently, M.tb exhibited transcriptional adaptation to intracellular stress, with induction of genes involved in dormancy survival, lipid metabolism, and ESX-1 secretion system. Gene ontology and KEGG pathway analyses indicated key host responses such as autophagy regulation, inflammasome activation, and metabolic reprogramming. Notably, a subset of non-coding RNAs and long intergenic transcripts were differentially expressed, suggesting novel regulatory mechanisms. These findings provide a comprehensive view of the early transcriptional landscape of M.tb-host interactions and offer candidate targets for host-directed therapies. Furthermore, transcriptomic biomarkers identified here hold promise for the development of early diagnostic tools to distinguish latent from active TB. Integration of transcriptomics with proteomics and epigenomics could enhance mechanistic understanding of TB pathogenesis.

Keywords: *Mycobacterium Tuberculosis*; Transcriptomics; Host-Pathogen Interaction; Early Infection; Macrophages; RNA Sequencing

1. Introduction

1.1. Global Burden of Tuberculosis and Research Gaps in Early Infection

Tuberculosis (TB) remains one of the leading infectious disease killers globally, with an estimated 10.6 million new cases and 1.6 million deaths reported in 2021 alone (1). Despite decades of public health efforts, TB continues to pose significant challenges due to complex transmission dynamics, the rise of multidrug-resistant strains, and latent infections that complicate disease surveillance (2). High-burden countries, especially in sub-Saharan Africa and Southeast Asia, account for the majority of global cases, often exacerbated by co-infections such as HIV (3).

While substantial progress has been made in diagnosing and treating active TB, early infection stages remain poorly understood. Most individuals infected with *Mycobacterium tuberculosis* (M. tuberculosis) do not develop active disease

* Corresponding author: Chiamaka Francisca Igweonu.

immediately but enter a latent state, which can persist for years before reactivation (4). The immunological and molecular events that govern this transition from exposure to latent or active disease are not fully delineated, representing a significant research gap (5).

Current diagnostics primarily detect established infection or active disease, offering little insight into the initial host-pathogen interactions that occur shortly after exposure (6). As a result, opportunities for early intervention are missed, particularly in high-risk populations such as household contacts or immunocompromised individuals. Furthermore, early-stage biomarkers capable of distinguishing between latent infection, incipient TB, and progressing disease are critically lacking (7).

Understanding the early stages of TB infection at the molecular level could revolutionize disease management by facilitating preclinical diagnosis and targeted preventive therapy. This requires integrative approaches that combine immunology, microbiology, and systems biology to capture the subtle host and pathogen changes occurring during early infection (8). Addressing these gaps is essential for developing novel diagnostics, vaccines, and treatment strategies that can interrupt TB transmission before clinical disease manifests (9).

1.2. Importance of Understanding Host-Pathogen Dynamics in Early Stages

The early stages of *M. tuberculosis* infection are marked by a complex interplay between the host immune system and the invading pathogen. Upon aerosol exposure, alveolar macrophages are among the first responders to internalize bacilli, initiating a cascade of immune responses that may either control or permit pathogen survival (10). This interaction sets the stage for long-term outcomes, including latent infection, sterilizing immunity, or progression to active disease (11).

A critical gap in TB research lies in the insufficient understanding of these early immune events. While the granuloma has been extensively studied in established disease, less is known about cellular and molecular responses that occur prior to its formation (12). Identifying key immunological checkpoints and pathogen adaptations during this window is essential for informing vaccine design and immune-modulating therapies (13).

Moreover, host responses in early infection can be highly variable depending on genetic background, nutritional status, co-infections, and microbiome composition (14). Transcriptomic profiling offers a high-resolution view into these variations by revealing patterns of gene expression that reflect real-time immune processes (15). Early transcriptomic signatures may provide predictive insights into disease trajectory and distinguish individuals at high risk of progression from those with controlled infection.

In-depth knowledge of host-pathogen dynamics during early infection can thus enable stratified approaches to TB prevention and care. It also opens avenues for precision medicine in TB control by tailoring interventions based on individual immune responses and infection risk profiles (16).

1.3. Scope and Objectives of Transcriptomic Profiling in TB

Transcriptomic profiling, which measures genome-wide gene expression, has emerged as a powerful tool to explore host responses to *M. tuberculosis* infection. This approach enables researchers to capture dynamic biological changes across various stages of infection, from initial exposure through to disease progression and resolution (17). By analyzing RNA expression patterns, transcriptomics reveals molecular signatures that would be invisible to traditional immunological assays.

The primary objective of transcriptomic profiling in TB is to identify biomarkers and immune pathways involved in early infection that could distinguish latent, incipient, and progressing disease (18). Such biomarkers are urgently needed to develop prognostic tools that can guide preemptive therapy, especially in individuals with subclinical or asymptomatic infection. Additionally, transcriptomic data can help decode mechanisms of immune evasion employed by *M. tuberculosis*, offering targets for host-directed therapies (19).

This study specifically focuses on transcriptomic analyses of peripheral blood and lung tissue samples collected during early infection. Through high-throughput RNA sequencing and bioinformatic modeling, the goal is to map gene expression trajectories that define protective versus pathological immune responses (20). A secondary objective is to compare expression profiles between high-risk groups, such as HIV-positive individuals, and immunocompetent hosts, to identify vulnerabilities and guide context-specific interventions.

Ultimately, the scope of this investigation extends to both clinical and translational research. By bridging the gap between bench and bedside, transcriptomic profiling holds the potential to transform TB diagnostics, inform vaccine development, and enhance global efforts to reduce the disease burden (21).

2. Pathophysiology of early *M. Tuberculosis* infection

2.1. Entry into the Host and Alveolar Macrophage Interaction

The infection cycle of *Mycobacterium tuberculosis* (M. tuberculosis) begins upon inhalation of aerosolized droplets carrying viable bacilli. These droplets penetrate deep into the lungs and reach the alveolar spaces, where the bacteria first encounter alveolar macrophages—the sentinel phagocytes of the respiratory tract (5). The bacilli are engulfed via phagocytosis, a process mediated by surface receptors such as the mannose receptor, complement receptor 3 (CR3), and DC-SIGN, allowing the bacteria to enter the macrophage in a membrane-bound phagosome (6).

Unlike many other intracellular pathogens, *M. tuberculosis* has evolved strategies to survive within this hostile environment. It arrests phagosome maturation by interfering with vesicle acidification and preventing phagolysosome fusion, thereby creating a protected intracellular niche (7). The pathogen can also exploit host lipids and iron for nutrition, supporting its replication within macrophages without triggering strong inflammatory responses early on.

The interaction between *M. tuberculosis* and alveolar macrophages sets the stage for either containment or dissemination. Infected macrophages produce chemokines such as MCP-1 and IL-8, which recruit additional immune cells, including monocytes, neutrophils, and dendritic cells, to the site of infection (8). These cells participate in the early containment response, though excessive recruitment may lead to tissue damage and aid bacterial spread.

Moreover, the intracellular persistence of *M. tuberculosis* in macrophages facilitates antigen presentation via both MHC class I and II pathways, triggering adaptive immune responses (9). However, this process can be delayed, allowing the pathogen a head start in establishing infection. Therefore, the early macrophage interaction is a critical determinant of disease outcome, influencing whether the infection remains latent or progresses (10).

2.2. Immune Evasion and Granuloma Initiation

M. tuberculosis has developed sophisticated strategies to evade immune clearance, particularly during the early stages of infection. One key mechanism is the modulation of host cytokine signaling to dampen pro-inflammatory responses. The pathogen can induce the production of IL-10 and TGF-β, which inhibit macrophage activation and reduce the bactericidal effects of reactive oxygen and nitrogen species (11). Simultaneously, it interferes with Toll-like receptor (TLR) signaling pathways, impairing the innate immune system's ability to mount an effective defense (12).

Despite these evasion strategies, the immune system responds by organizing immune cells into granulomas—structured aggregates of infected macrophages, epithelioid cells, multinucleated giant cells, and surrounding lymphocytes. Granuloma formation begins within two to three weeks post-infection and serves to contain the pathogen and prevent systemic dissemination (13). However, *M. tuberculosis* can exploit the granuloma environment as a niche for long-term survival. Within this microenvironment, oxygen tension and nutrient availability are limited, leading to metabolic adaptation of the bacilli into a dormant state (14).

Recent studies have shown that granulomas are dynamic rather than static, with continuous cellular turnover and bacterial activity occurring within their core. The balance between host containment and bacterial persistence is finely regulated, and disruption of this balance—through immune suppression or other factors—can lead to reactivation (15). Understanding how *M. tuberculosis* manipulates granuloma architecture and function is essential for developing therapies that target latent reservoirs and prevent disease progression.

2.3. Transition from Latency to Active Infection

Latency is a hallmark of *M. tuberculosis* infection, characterized by the persistence of viable bacteria within the host without clinical symptoms or radiographic evidence of disease. This latent state affects approximately one-quarter of the global population and represents a substantial reservoir for future reactivation (16). Reactivation occurs when host immune surveillance fails, typically due to immunosuppressive conditions such as HIV infection, malnutrition, or aging (17).

At the cellular level, the transition from latency to active TB involves a shift in the host-pathogen equilibrium. Dormant bacilli reinitiate replication, often triggered by changes in the microenvironment within granulomas, including hypoxia

resolution, altered lipid metabolism, and decreased cytokine production (18). Simultaneously, the immune system may experience functional exhaustion, with reduced T-cell proliferation and diminished macrophage responsiveness contributing to disease progression.

Molecular studies have shown that reactivation is accompanied by altered transcriptional profiles in both host and pathogen. Host gene expression shifts toward a pro-inflammatory state, marked by increased levels of IFN-γ, IL-1β, and TNF-α, which are associated with tissue damage and cavity formation (19). In parallel, *M. tuberculosis* upregulates genes involved in cell wall remodeling, virulence, and energy metabolism, facilitating its escape from dormancy (20).

Identifying the molecular triggers and regulatory networks involved in reactivation is a critical goal in TB research. It offers potential for developing predictive biomarkers and targeted interventions aimed at preventing disease resurgence in latently infected individuals. Effective control of TB will ultimately depend on our ability to detect and block reactivation before clinical disease emerges (21).

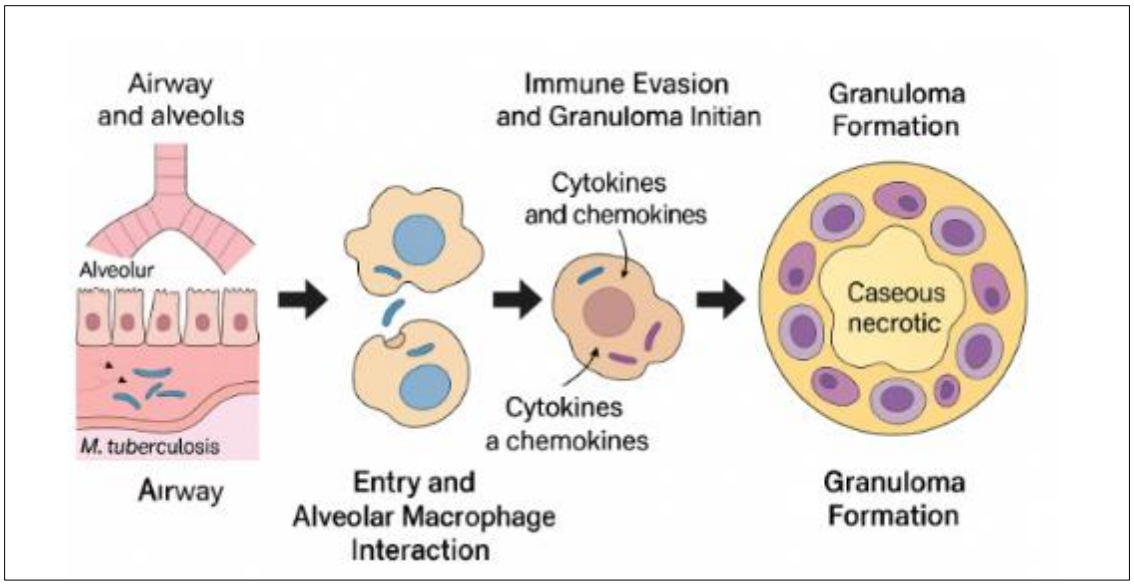

Figure 1 Schematic diagram showing early-stage host-pathogen interaction and granuloma formation

3. Transcriptomic approaches in infection biology

3.1. RNA-seq, Microarrays, and Single-Cell Transcriptomics Overview

Transcriptomic technologies have revolutionized the study of host-pathogen interactions in tuberculosis (TB), enabling comprehensive profiling of gene expression during infection. Among these, RNA sequencing (RNA-seq) has emerged as the gold standard due to its high sensitivity, dynamic range, and ability to detect novel transcripts (11). RNA-seq provides an unbiased view of the transcriptome by sequencing cDNA derived from total or polyadenylated RNA, allowing quantification of both host and pathogen transcripts in co-infection models.

In contrast, microarrays, though historically important, rely on predefined probes and are limited in detecting unknown or low-abundance transcripts. Nonetheless, microarrays have contributed significantly to early TB research by revealing differential gene expression patterns associated with latent and active disease (12). However, they lack the resolution and adaptability of next-generation sequencing platforms, particularly in capturing the complexity of dual transcriptomes in host-pathogen systems.

Single-cell RNA sequencing (scRNA-seq) is a more recent advancement that provides gene expression data at the individual cell level, offering insights into cellular heterogeneity and microenvironmental context during TB infection (13). This is particularly valuable for studying lung granulomas, where diverse cell types interact dynamically with *M. tuberculosis*. scRNA-seq allows the dissection of cell-specific responses and identification of rare immune cell subsets that may influence disease outcomes (14).

Each method offers distinct advantages depending on the experimental objective. RNA-seq remains the most widely used technique for dual transcriptomics due to its scalability and ability to detect both host and bacterial RNAs simultaneously (15). When integrated with scRNA-seq or spatial transcriptomics, these technologies can uncover the complex molecular dialogue between host and pathogen with unprecedented resolution, supporting the discovery of biomarkers, therapeutic targets, and mechanisms of immune evasion (16). Advancing transcriptomic methods will be critical for mapping the temporal and spatial landscapes of TB infection and guiding precision medicine approaches.

3.2. Sample Preparation and Isolation of Host vs. Pathogen RNA

Effective transcriptomic analysis of host-pathogen interactions in tuberculosis begins with meticulous sample preparation. Tissues such as lung biopsies, bronchoalveolar lavage (BAL), and peripheral blood mononuclear cells (PBMCs) are commonly used for host RNA profiling, while infected macrophage cultures or sputum samples enable simultaneous capture of *M. tuberculosis* RNA (17). Given the low abundance of bacterial RNA relative to host RNA, protocols must be optimized to enrich for pathogen transcripts without compromising host data integrity.

Total RNA is extracted using reagents such as TRIzol or specialized kits that preserve RNA integrity while minimizing contamination with genomic DNA (18). For dual RNA-seq applications, rRNA depletion is favored over poly-A selection to retain both bacterial and eukaryotic transcripts, as most bacterial RNAs lack polyadenylation (19). Commercial kits now provide targeted rRNA depletion for human and *M. tuberculosis* rRNAs, improving the efficiency of dual transcriptomics and maximizing sequencing depth for relevant transcripts.

To reduce background noise and enhance pathogen signal, physical and enzymatic methods such as cell lysis, differential centrifugation, and host-cell RNA subtraction may be employed (20). Laser-capture microdissection (LCM) and flow cytometry-based sorting can also be applied to isolate infected host cells or granuloma regions of interest, enabling spatial transcriptomic analysis (21).

RNA integrity is assessed via capillary electrophoresis, ensuring RNA integrity number (RIN) scores above 7 for downstream sequencing. Isolated RNA is then quantified, quality-checked, and stored at −80°C to preserve transcript stability. Proper sample handling and RNA enrichment strategies are essential for capturing the full transcriptional interplay between host and pathogen, allowing for accurate and reproducible transcriptomic analyses (22).

3.3. Bioinformatic Pipelines for Dual RNA-seq

The bioinformatic analysis of dual RNA-seq data in tuberculosis studies involves multiple computational steps to distinguish and interpret transcripts originating from both host and *M. tuberculosis*. The first challenge is the accurate alignment of reads to two separate reference genomes. This is typically achieved through sequential or parallel alignment using tools like HISAT2 or STAR for the human genome, and BWA or Bowtie2 for the bacterial genome (23). Reads are filtered and classified based on alignment quality to prevent cross-mapping and ensure precise taxonomic assignment (24).

Once aligned, gene expression is quantified using featureCounts or HTSeq, generating count matrices for both host and pathogen transcripts. These data are normalized using methods such as DESeq2 or edgeR to account for differences in sequencing depth and gene length (25). Batch effects and sample heterogeneity are controlled through surrogate variable analysis (SVA) or principal component analysis (PCA) prior to differential expression analysis (26).

Functional enrichment tools like DAVID, GSEA, and KEGG are used to interpret host gene expression changes in biological context, highlighting pathways related to immune activation, apoptosis, and metabolic reprogramming. On the bacterial side, pathway analysis can reveal stress responses, virulence gene activation, and metabolic adaptations during intracellular survival (27). Transcriptome-wide association studies (TWAS) can further integrate genotype and expression data to identify regulatory variants influencing susceptibility or resistance to TB (28).

Advanced pipelines now incorporate deconvolution algorithms to resolve cellular heterogeneity and isolate cell-type-specific responses, particularly useful in lung tissue or PBMC samples. Integration with single-cell datasets can enhance resolution and support cross-validation of findings (29). Machine learning models are also increasingly applied to identify transcriptomic biomarkers predictive of disease stage or treatment outcome.

The dual RNA-seq approach thus offers a systems-level view of host-pathogen interactions, supporting precision TB research through accurate annotation, quantitative modeling, and mechanistic insight (30).

Table 1 Comparative Summary of Transcriptomic Technologies for TB Research

Technology	Sensitivity	Resolution	Cost (per sample)	Strengths	Limitations
Microarray	Moderate – limited to predefined probes	Bulk – average across mixed cell populations	Low to moderate (~$100–$200)	Cost-effective for large cohorts; established pipelines	Cannot detect novel transcripts; low dynamic range
Bulk RNA-seq	High – detects low-abundance transcripts	Bulk – population-level gene expression	Moderate (~$200–$500)	Unbiased; suitable for host-pathogen dual RNA-seq	Masks cell-type-specific expression; limited spatial context
Single-cell RNA-seq (scRNA-seq)	Very high – captures transcriptomes of individual cells	Single-cell – resolves cellular heterogeneity	High (~$800–$1,500)	Dissects immune cell subsets; identifies rare populations	High cost; complex data analysis
Spatial Transcriptomics	High – combines histological and gene expression data	Spatial – maps expression within tissue context	Very high (~$1,500–$3,000)	Preserves tissue architecture; localizes host-pathogen zones	Expensive; lower transcriptome depth compared to scRNA-seq
Dual RNA-seq (host + pathogen)	High – simultaneously detects host and bacterial RNAs	Bulk or single-cell depending on setup	Moderate to high (~$500–$1,200)	Enables direct host-pathogen interaction profiling	Pathogen RNA often low abundance; needs deep sequencing

4. Host transcriptomic responses in early *M. Tuberculosis* infection

4.1. Innate Immune Activation: TLRs, Cytokines, and Chemokines

Upon encountering *Mycobacterium tuberculosis* (M. tuberculosis), the host innate immune system is activated through pattern recognition receptors (PRRs), most notably the Toll-like receptors (TLRs), which detect conserved mycobacterial components such as lipoproteins, lipomannan, and peptidoglycan (14). TLR2, often in heterodimeric combination with TLR1 or TLR6, plays a central role in initiating the early inflammatory response. Activation of TLRs leads to downstream signaling cascades involving MyD88, NF-κB, and MAP kinases, ultimately driving the transcription of pro-inflammatory cytokines (15).

Key cytokines produced include tumor necrosis factor-alpha (TNF-α), interleukin-1 beta (IL-1β), and interleukin-6 (IL-6), all of which contribute to the recruitment and activation of additional immune cells. TNF-α, in particular, is critical for granuloma maintenance and containment of the bacilli, while IL-1β promotes macrophage activation and pyroptotic cell death as a defense strategy (16). The IL-6 response serves as both a pro-inflammatory and anti-inflammatory signal, helping to shape the outcome of infection through STAT3 signaling (17).

Chemokines such as CXCL10 (IP-10), CCL2 (MCP-1), and CCL5 (RANTES) are also strongly induced, facilitating the recruitment of monocytes, neutrophils, and lymphocytes to the site of infection (18). These chemokines help orchestrate the formation of granulomas by concentrating immune cells around infected macrophages. However, excessive chemokine production can contribute to tissue pathology, particularly in progressive TB (19).

Interestingly, *M. tuberculosis* can modulate TLR expression and cytokine production to create an immune environment favorable for its persistence. For example, some strains downregulate TLR4 signaling or upregulate IL-10, an immunosuppressive cytokine, to dampen the immune response (20). These evasive tactics delay adaptive immunity, prolong bacterial survival, and promote chronic infection.

Overall, the interplay between TLR-mediated sensing and cytokine-chemokine networks defines the early immune landscape of TB infection. Understanding these mechanisms is essential for developing interventions that strengthen innate defenses without exacerbating immunopathology (21).

4.2. Modulation of Apoptosis, Autophagy, and Metabolic Pathways

M. tuberculosis manipulates host cellular processes such as apoptosis, autophagy, and metabolism to favor its intracellular survival. Apoptosis, a form of programmed cell death, can limit infection by preventing bacterial replication and promoting antigen presentation. However, *M. tuberculosis* inhibits apoptosis in infected macrophages by upregulating anti-apoptotic proteins such as Bcl-2 and by activating PI3K/Akt pathways that delay caspase activation (22). By blocking apoptosis, the pathogen avoids destruction and continues to replicate within host cells.

Conversely, in some contexts, *M. tuberculosis* induces necrotic or pyroptotic forms of cell death, which facilitate bacterial dissemination through tissue damage and uncontrolled inflammation (23). This dual modulation of host cell fate allows the pathogen to balance intracellular survival with transmission potential.

Autophagy, a catabolic pathway that targets intracellular pathogens for lysosomal degradation, is another key host defense mechanism. In response to infection, host cells upregulate autophagy-related genes such as ATG5 and LC3 to sequester and destroy intracellular *M. tuberculosis* (24). However, the bacterium actively suppresses autophagic flux through virulence factors like Eis and PknG, which interfere with phagosome maturation and autophagosome-lysosome fusion (25). Additionally, *M. tuberculosis* can subvert mTOR signaling to inhibit autophagy initiation, thereby evading immune clearance.

At the metabolic level, infected macrophages undergo reprogramming toward aerobic glycolysis, a phenomenon known as the Warburg effect, to support energy demands during inflammation (26). This shift enhances production of reactive oxygen species (ROS) and nitric oxide (NO), both of which have antimicrobial effects. However, *M. tuberculosis* adapts by upregulating stress response genes and antioxidant defenses, enabling it to persist in this hostile environment (27).

The pathogen also manipulates host lipid metabolism, inducing foamy macrophage formation rich in cholesterol and triglycerides that serve as nutrient reservoirs. This lipid accumulation, driven by PPARγ activation, supports long-term bacterial survival during latency (28).

Collectively, these interactions highlight how *M. tuberculosis* intricately modulates host apoptosis, autophagy, and metabolism to evade immune destruction and create a favorable intracellular niche (29).

4.3. Host Cell Reprogramming and Immune Escape Signals

Host cell reprogramming during *M. tuberculosis* infection extends beyond localized immune responses and involves widespread transcriptional and epigenetic changes that affect both innate and adaptive immunity. One of the central features of this reprogramming is the induction of immune-regulatory molecules such as PD-L1, IL-10, and indoleamine 2,3-dioxygenase (IDO), which collectively suppress T cell activity and facilitate immune escape (30). These factors are upregulated in infected macrophages and dendritic cells, inhibiting the proliferation and effector function of CD4+ and CD8+ T cells, thereby weakening the host's ability to mount an effective adaptive response (31).

Epigenetic modifications play a pivotal role in host-pathogen interactions, influencing the accessibility of key immune genes. *M. tuberculosis* induces histone modifications such as H3K27 trimethylation to silence inflammatory genes in macrophages, thereby modulating chromatin structure to favor a permissive environment for bacterial persistence (32). DNA methylation of host gene promoters, including those encoding pro-inflammatory cytokines, has also been observed during latent and chronic stages of infection (33). These heritable yet reversible changes ensure prolonged modulation of host responses even after bacterial clearance.

Transcriptomic profiling has revealed that *M. tuberculosis* drives a distinct host gene expression program characterized by upregulation of type I interferon pathways, suppression of antigen presentation genes, and skewing toward a regulatory macrophage phenotype (M2) (34). The M2 phenotype, marked by high expression of arginase-1 and IL-10, promotes tissue repair but also compromises bacterial killing. Moreover, infected macrophages express lower levels of costimulatory molecules like CD80 and CD86, impairing effective T-cell priming (35).

Another escape strategy involves extracellular vehicles (EVs) secreted by infected cells. These vesicles carry bacterial antigens, RNA, and immune-modulatory molecules that alter the function of recipient immune cells at distant sites (36).

EV-mediated communication can either promote immune tolerance or trigger systemic inflammation, depending on the context and cargo composition.

Together, these strategies underscore the ability of *M. tuberculosis* to reprogram host cells at multiple levels, including transcriptional, metabolic, and epigenetic layers. This reprogramming ensures immune evasion, persistence, and the establishment of latent or chronic infection, presenting significant challenges for TB treatment and vaccine development (37).

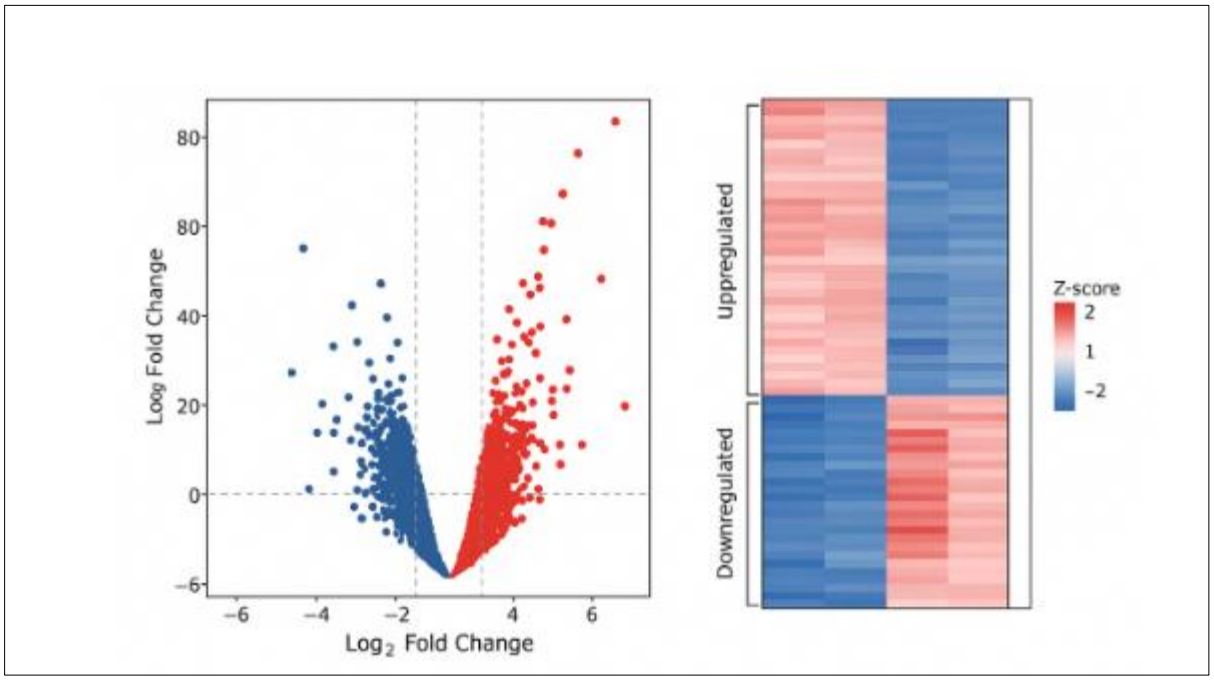

Figure 2 Volcano plot and heatmap of upregulated and downregulated host genes at 48 hours post-infection

5. Pathogen gene expression during early intracellular survival

5.1. Activation of Dormancy Regulons and Stress Response Pathways

One of the key survival strategies employed by *Mycobacterium tuberculosis* (M. tuberculosis) during latent infection is the activation of dormancy regulons and stress response pathways. These gene networks enable the bacterium to endure hostile conditions such as hypoxia, nutrient deprivation, acidic pH, oxidative stress, and nitric oxide exposure within granulomas (18). Central to this response is the DosR (dormancy survival regulator) regulon, a transcriptional program comprising over 50 genes that are upregulated in response to oxygen limitation and nitric oxide signaling (19).

The DosR regulon includes genes such as hspX (encoding α-crystallin), acg, and Rv2623, which contribute to growth arrest and metabolic adaptation under anaerobic conditions. These genes facilitate a non-replicative persistent (NRP) state in which *M. tuberculosis* reduces its metabolic activity and becomes tolerant to antibiotics, particularly those targeting active cell division (20). Additionally, DosS and DosT, the two histidine kinase sensors upstream of DosR, allow the bacterium to monitor environmental cues and rapidly adjust gene expression profiles in response to changing conditions (21).

In parallel, general stress response genes are activated to mitigate damage and support intracellular survival. These include oxidative stress regulators such as sigH and ahpC, which counteract reactive oxygen species, as well as chaperones like groEL and dnaK that refold misfolded proteins (22). Metabolic regulators including isocitrate lyase (icl) and phosphoenolpyruvate carboxykinase (pckA) are also upregulated, enabling the bacterium to shift toward lipid-based metabolism during nutrient deprivation (23).

This dormancy program provides *M. tuberculosis* with resilience and adaptability, allowing it to persist in host tissues for years without triggering immune clearance or clinical symptoms. Understanding the transcriptional regulation of dormancy is crucial for identifying drug targets capable of eliminating latent bacilli and reducing reactivation risk (24).

5.2. ESX Secretion Systems and Cell Wall Remodeling Genes

The ESX (early secretory antigenic target-6 system) secretion systems are critical to *M. tuberculosis* pathogenesis, allowing the bacterium to transport effector proteins across its complex cell envelope. Of the five ESX systems (ESX-1 to ESX-5), ESX-1 is the most extensively studied and is essential for virulence, especially during early infection and immune evasion (25). ESX-1 mediates the secretion of proteins such as ESAT-6 and CFP-10, which disrupt phagosomal membranes, facilitate cytosolic access, and promote antigen presentation modulation (26).

These effector proteins also trigger host cell death pathways, such as necrosis and pyroptosis, leading to tissue damage and enhanced bacterial dissemination (27). ESX-1 components are regulated by transcriptional factors including EspR and WhiB6, which respond to environmental stress and host-derived signals (28). Mutations or deletions in the ESX-1 locus, as seen in the vaccine strain *Mycobacterium bovis* BCG, result in significant attenuation of virulence, underlining its importance in pathogenesis (29).

Beyond protein secretion, *M. tuberculosis* dynamically remodels its cell wall in response to host-derived pressures. Genes involved in the synthesis and modification of mycolic acids, arabinogalactan, and peptidoglycan—such as pks13, embC, and murA—are differentially expressed under stress conditions (30). These modifications increase cell wall thickness and permeability barrier, enhancing resistance to antimicrobial peptides and reactive oxygen species (31).

Furthermore, the bacterium produces surface lipids like phthiocerol dimycocerosates (PDIMs) and phenolic glycolipids (PGLs), which modulate host immune responses and contribute to immune evasion (32). Their biosynthesis is controlled by regulators including PhoP and SigE, which are also involved in coordinating broader virulence responses.

Combined, the ESX secretion systems and cell wall remodeling genes are vital for *M. tuberculosis* adaptation, immune modulation, and persistence. Targeting these pathways could lead to novel therapeutics that impair bacterial survival and enhance host immune clearance (33).

5.3. Antigenic Variation and Immune Modulation by *M. tuberculosis*

M. tuberculosis employs antigenic variation and immune modulation to persist in the host by avoiding detection and destruction. While its genome is relatively conserved compared to other pathogens, *M. tuberculosis* expresses a diverse array of antigenic proteins in a phase-variable manner, especially under immune pressure (34). The PE/PPE family of proteins, which constitutes nearly 10% of the genome, is notable for its antigenic diversity and capacity to modulate host immune responses through altered T-cell recognition (35).

These proteins are often secreted through ESX-5, facilitating antigenic variation on the bacterial surface and interfering with host antigen processing pathways. This variability reduces the effectiveness of antigen-specific T-cell responses and may contribute to vaccine failure, as seen with limited efficacy of BCG in adults (36).

In addition to antigenic variation, *M. tuberculosis* actively modulates host immune responses through the secretion of immunomodulatory lipids, nucleic acids, and proteins. For example, the LAM (lipoarabinomannan) molecule can inhibit dendritic cell maturation and suppress the expression of costimulatory molecules necessary for T-cell activation (37). Similarly, secreted proteins such as Rv1988 and PtpA interfere with host cell signaling pathways, blocking phagosome-lysosome fusion and antigen presentation (38).

M. tuberculosis also skews the cytokine environment by inducing regulatory cytokines like IL-10 and TGF-β, thereby suppressing Th1-mediated immunity and allowing persistent infection (39). These immune evasion mechanisms collectively hinder effective clearance, promote latency, and contribute to long-term pathogen survival.

Understanding antigenic variation and immune modulation is essential for the design of next-generation TB vaccines and immunotherapies. Targeting conserved antigens and blocking immunosuppressive signals could improve host immunity and reduce disease burden (40).

Table 2 Selected *M. tuberculosis* Genes Differentially Expressed During Early Infection Stages and Their Roles

Gene	Expression Pattern	Function / Role	Associated Pathway
esxA (ESAT-6)	Upregulated	Disrupts phagosomal membrane; aids in cytosolic escape	ESX-1 secretion system
esxB (CFP-10)	Upregulated	Binds ESAT-6; facilitates secretion and immune modulation	ESX-1 secretion system
hspX (acr)	Upregulated	Stress response protein; mediates dormancy and survival	DosR regulon / hypoxia response
Rv2031c	Upregulated	Part of dormancy regulon; expressed under low oxygen	DosR regulon
sigH	Induced under oxidative stress	Sigma factor regulating antioxidant and stress response genes	General stress response
mbtB	Upregulated	Involved in siderophore (mycobactin) synthesis for iron uptake	Iron acquisition / metabolic adaptation
pknG	Constitutively expressed	Inhibits phagosome–lysosome fusion; promotes intracellular survival	Intracellular survival / signaling
fadD26	Upregulated	Involved in lipid metabolism and virulence lipid synthesis	Cell wall lipid synthesis
Rv2623	Upregulated	Universal stress protein; regulates bacterial growth rate	Dormancy and persistence
eis	Induced early	Enhances resistance to host-mediated stress; inhibits autophagy	Immune evasion / antibiotic resistance

6. Dual RNA-SEQ and cross-talk analysis

6.1. Experimental Models for Dual RNA-seq in TB (in vitro, in vivo)

Dual RNA sequencing (dual RNA-seq) enables simultaneous profiling of host and pathogen transcriptomes, offering unprecedented insights into their dynamic interactions during tuberculosis (TB) infection. To generate meaningful data, both in vitro and in vivo experimental models are employed, each with distinct advantages and limitations (23).

In vitro models typically involve the infection of human or murine macrophage cell lines (e.g., THP-1 or RAW264.7) or primary monocyte-derived macrophages with *Mycobacterium tuberculosis* at defined multiplicity of infection (MOI). These models allow controlled manipulation of variables such as oxygen tension, nutrient availability, and drug exposure (24). Time-course sampling of infected cultures provides high-resolution insights into temporal transcriptional changes. However, these models lack the tissue complexity and immunological context of a whole organism (25).

In vivo models, especially murine infection systems, are more physiologically relevant and are commonly used to study transcriptional responses in lung tissue or granulomas. Genetically modified mice (e.g., IFN-γ knockout) can help dissect specific immune pathways. Additionally, non-human primates (NHPs) offer closer immunological and pathological resemblance to human TB, enabling longitudinal studies of latent and active infection stages (26). Despite their strengths, in vivo models present challenges in pathogen RNA recovery due to the host transcriptome's dominance and the low abundance of bacterial RNA.

To overcome these challenges, cell-sorting techniques and laser-capture microdissection are used to enrich infected cells or granuloma regions before RNA extraction (27). Emerging organoid and ex vivo lung tissue models also offer promising alternatives, bridging the gap between reductionist in vitro setups and complex in vivo systems. Thus, the choice of model directly influences the resolution, interpretability, and translational potential of dual RNA-seq studies in TB (28).

6.2. Analytical Frameworks for Identifying Interactome Modules

Identifying meaningful interactions between host and pathogen transcriptional profiles requires robust analytical frameworks capable of integrating high-dimensional dual RNA-seq datasets. The first step typically involves quality filtering, mapping reads to both host and pathogen genomes, and generating separate count matrices for each organism using tools such as STAR for the host and Bowtie2 or BWA for *M. tuberculosis* (29). Following alignment, normalization methods such as TPM (transcripts per million) or variance stabilizing transformation (VST) are applied to correct for sequencing depth and RNA composition (30).

To identify interactome modules, co-expression network analysis is employed. Weighted Gene Co-expression Network Analysis (WGCNA) constructs gene modules based on expression similarity, which can then be correlated with clinical phenotypes or experimental conditions (31). Modules are examined for enrichment of Gene Ontology (GO) terms or KEGG pathways to infer biological function. On the pathogen side, bacterial modules are often smaller but can reveal clusters associated with virulence, dormancy, or stress response pathways (32).

Cross-species interaction networks are constructed by correlating host and pathogen gene modules using biclustering, canonical correlation analysis (CCA), or joint non-negative matrix factorization (jNMF). These methods identify transcriptional overlaps or dependencies that may indicate molecular crosstalk (33). For instance, host immune activation genes may correlate negatively with bacterial stress response genes, suggesting an adaptive pathogen reaction to immune pressure.

Machine learning algorithms such as random forests and support vector machines (SVMs) are also used to classify infection states or predict key regulatory genes from dual RNA-seq data. Feature selection methods rank genes based on importance scores, identifying transcriptional hubs that may control host-pathogen interactions (34).

Integrated datasets can further be visualized using platforms like Cytoscape, allowing dynamic modeling of transcriptomic networks. Host-pathogen interactome maps constructed from dual RNA-seq provide valuable insight into systems-level interactions and identify novel targets for host-directed therapies or antimicrobial agents (35). The precision of these analytical frameworks is critical for translating dual RNA-seq data into actionable biological knowledge.

6.3. Examples of Host-Pathogen Interaction Clusters (transcriptional overlap)

Dual RNA-seq studies in TB have uncovered specific clusters of host and pathogen genes whose transcriptional profiles are interdependent. One example involves the host type I interferon response and the *M. tuberculosis* DosR regulon. Elevated expression of IFN-stimulated genes (ISGs), such as IFIT1, OASL, and MX1, in infected macrophages has been found to coincide with upregulation of bacterial dormancy genes including hspX and Rv2031c, suggesting a bacterial adaptation to interferon-induced stress (36).

Another prominent cluster includes the host pro-inflammatory cytokine genes IL1B, TNF, and CXCL10, which show a positive correlation with *M. tuberculosis* genes involved in lipid metabolism and iron acquisition, such as mbtB and fadD26. These bacterial genes help the pathogen access host-derived nutrients under immune pressure, indicating metabolic cross-talk as a form of immune evasion (37).

Studies have also identified clusters linking host autophagy genes (e.g., ATG5, BECN1) with *M. tuberculosis* secretion system genes (e.g., esxA, esxB), suggesting that bacterial effector proteins directly modulate host degradation pathways to avoid clearance (38). This transcriptional overlap implies coordinated regulatory mechanisms that sustain intracellular persistence.

In granulomatous tissue samples, dual RNA-seq has shown co-activation of host angiogenesis markers (VEGFA, ANGPT2) and bacterial stress response genes, reflecting the hypoxic microenvironment and its influence on both host and pathogen gene expression (39). These clusters exemplify how transcriptional patterns reflect mutual adaptation during chronic infection.

Such findings validate the utility of dual RNA-seq in uncovering synchronous gene expression modules that define host-pathogen interaction landscapes. These interaction clusters offer a mechanistic basis for designing diagnostics and interventions that disrupt critical pathways sustaining TB infection (40).

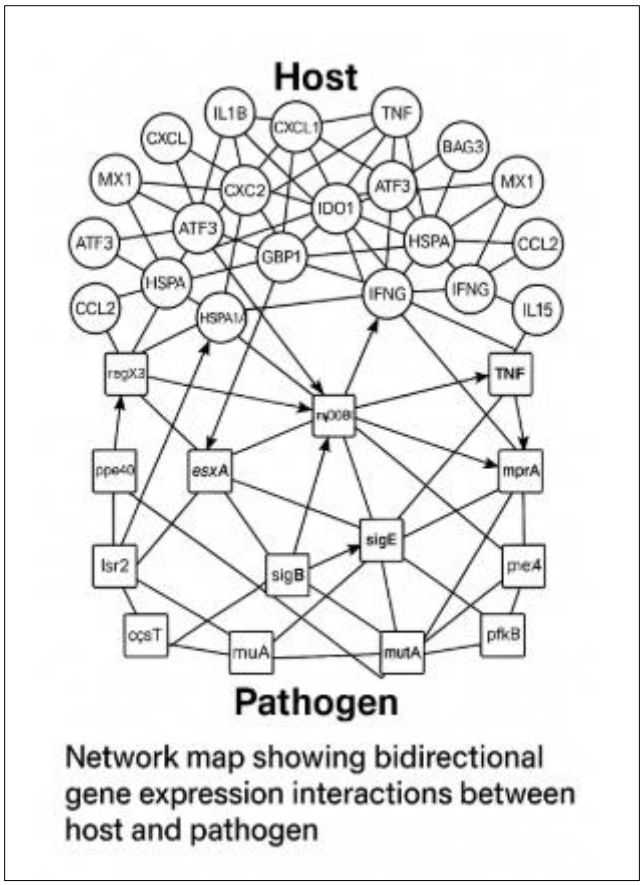

Figure 3 Network map showing bidirectional gene expression interactions between host and pathogen

7. Case studies and comparative analyses

7.1. Human Lung Epithelial Cell Models vs. Murine Macrophages

Experimental model selection significantly influences transcriptomic insights in tuberculosis (TB) research. Human lung epithelial cell lines, such as A549 and Calu-3, are increasingly used to study early-stage host responses to *Mycobacterium tuberculosis* (M. tuberculosis), particularly in the context of mucosal immunity and pathogen entry (27). These epithelial models offer advantages in mimicking the first point of contact between inhaled bacilli and the respiratory tract, allowing the investigation of barrier integrity, cytokine secretion, and cell-cell signaling.

Lung epithelial cells respond to infection by producing chemokines such as CXCL8 and CCL20, which recruit immune cells to the infection site, and by expressing pattern recognition receptors like TLR2 and NOD2 (28). However, they lack robust phagocytic capacity and certain innate immune functionalities present in macrophages, limiting their utility in studying intracellular survival and bacterial clearance (29).

Conversely, murine macrophage models, including RAW264.7 and bone marrow-derived macrophages (BMDMs), remain widely used due to their well-characterized immune phenotypes and ease of genetic manipulation. They efficiently phagocytose *M. tuberculosis*, generate reactive oxygen and nitrogen species, and undergo regulated cell death in response to infection (30). These models are ideal for examining intracellular signaling pathways and host defense mechanisms.

Nonetheless, species differences in immune gene expression and signaling pathways between mice and humans limit the translational relevance of murine macrophages. For instance, certain cytokines and TLR responses differ substantially, leading to divergent transcriptomic profiles (31). Therefore, the integration of human epithelial models with macrophage systems offers a more comprehensive view of early TB pathogenesis, capturing both barrier and innate immune responses. Selecting complementary models allows researchers to dissect compartment-specific transcriptional responses and enhance the biological relevance of dual RNA-seq analyses (32).

7.2. Virulent vs. Attenuated *M. tuberculosis* Strains Transcriptomic Contrast

Comparative transcriptomic profiling of virulent and attenuated *Mycobacterium tuberculosis* strains reveals distinct bacterial adaptation mechanisms and host immune responses. Virulent strains such as H37Rv and Beijing lineage strains elicit strong inflammatory reactions and exhibit enhanced survival within host cells, while attenuated strains like H37Ra and *M. bovis* BCG typically induce reduced pathogenicity and altered immune signatures (33).

At the bacterial transcriptome level, virulent strains upregulate genes involved in lipid metabolism (e.g., fadE24, Rv2945c), iron acquisition (mbtB, irtA), and ESX-1 secretion system components (espA, esxA), facilitating intracellular adaptation and immune modulation (34). These virulence-associated genes enhance phagosomal escape, inhibit autophagy, and interfere with antigen presentation, contributing to bacterial persistence (35).

In contrast, attenuated strains demonstrate reduced expression of ESX-1-associated genes and lower induction of the DosR regulon under hypoxic or nitric oxide stress conditions. This downregulation correlates with impaired ability to establish latency-like states and less robust transcriptional responses to host-imposed stress (36). Furthermore, attenuated strains show diminished regulation of regulatory proteins such as SigH and WhiB3, limiting their capacity to coordinate complex survival programs (37).

From the host perspective, infection with virulent strains leads to the upregulation of pro-inflammatory cytokines (IL1B, TNF, IFNG) and chemokines (CCL2, CXCL10), accompanied by interferon-stimulated gene signatures, especially during early infection stages (38). This strong response reflects recognition of virulence factors and immune-stimulatory components. Meanwhile, attenuated strains elicit milder transcriptomic shifts, characterized by lower expression of inflammatory genes and reduced activation of NF-κB and STAT1 pathways (39).

Notably, dual RNA-seq studies have shown that macrophages infected with virulent strains exhibit higher expression of autophagy suppressors (e.g., mTOR) and anti-apoptotic regulators (BCL2), indicating the pathogen's capacity to actively subvert host defense (40). These transcriptional contrasts underscore the significance of virulence-associated genes in shaping both pathogen survival strategies and host immune dynamics.

Such comparative analyses are essential for identifying molecular determinants of pathogenicity and for informing the design of improved vaccines and attenuated strains that elicit protective, rather than suppressive, host responses (41).

7.3. Time-Course Differential Expression Study (6h, 24h, 48h)

Time-course transcriptomic studies allow researchers to capture the dynamic evolution of host and pathogen gene expression across different stages of infection. A typical timeline includes early (6 hours), intermediate (24 hours), and late (48 hours) time points post-infection, providing a kinetic perspective on transcriptional reprogramming during *Mycobacterium tuberculosis* infection (42).

At 6 hours post-infection, the host response is dominated by the upregulation of innate immune genes such as TLR2, NOD2, IL1B, and TNF. These early signals initiate pathogen recognition and promote the recruitment of immune cells through chemokines like CXCL8 and CCL2 (43). On the bacterial side, early transcriptional changes involve ESX-1-associated genes (esxA, esxB), enabling phagosomal disruption and immune interference.

By 24 hours, host macrophages begin transitioning toward a more regulated immune response, characterized by increased expression of IL10, SOCS3, and autophagy-related genes (BECN1, ATG5) (44). Bacterial gene expression shifts to include dormancy regulon components (e.g., hspX, Rv2623) and stress response genes (sigH, ahpC), indicating adaptation to host-induced pressures.

At 48 hours, the host exhibits feature of immune modulation and polarization. M2 markers such as ARG1 and IL4RA may emerge, while pro-inflammatory responses subside in favor of granuloma-associated processes (45). Concurrently, *M. tuberculosis* increases the expression of lipid metabolism genes (fadD26, tgs1), preparing for long-term survival under nutrient-limited conditions.

These temporal expression patterns reveal how both host and pathogen progressively recalibrate their strategies. Early activation is followed by immune modulation and persistence signals, highlighting key windows for therapeutic intervention. Time-course dual RNA-seq thus provides a mechanistic basis for understanding infection progression and identifying stage-specific biomarkers or drug targets (46).

Table 3 Temporal Transcriptomic Markers of Infection Stage Progression

Time Point	Key Transcriptomic Markers	Function / Role	Implication for Stage
6 hours	TLR2, NOD2, IL1B, TNF, CXCL8	Early innate immune activation, inflammation, neutrophil recruitment	Indicates recognition and initial immune response
24 hours	IL10, SOCS3, ATG5, BECN1, IFIT1	Immune regulation, autophagy initiation, type I IFN response	Marks transition to immune modulation and intracellular stress
48 hours	ARG1, IL4RA, VEGFA, MMP9, CXCL10	M2 polarization, angiogenesis, matrix remodeling, chronic inflammation	Reflects granuloma development and immune adaptation
Latent TB	GBP5, IFI44L, BATF2, Rv2031c (bacterial)	Host immune surveillance, dormancy regulation in pathogen	Characteristic of controlled infection and dormancy maintenance
Active TB	MX1, OASL, IFNG, IL1RN, HLA-DRB1	Heightened inflammation, antigen presentation, interferon signaling	Indicates progression to symptomatic disease

8. Application of transcriptomic signatures in diagnostics and therapeutics

8.1. Biomarkers for Early TB Detection and Disease Stratification

Identifying reliable biomarkers for early tuberculosis (TB) detection and disease stratification remains a global health priority. Transcriptomic profiling has emerged as a powerful tool for uncovering host gene expression signatures that distinguish between latent TB infection (LTBI), incipient TB, and active disease. Numerous studies have identified gene panels with diagnostic potential, particularly those involving interferon-stimulated genes (ISGs) (32). For example, the upregulation of genes such as IFI44L, GBP5, and BATF2 has been associated with progression from latent to active TB and offers predictive potential (33).

Several transcriptional signatures have been validated across independent cohorts and populations. One notable example is the 16-gene RISK6 signature, which can predict progression to active TB up to 12 months before clinical diagnosis (34). Similarly, three- and four-gene signatures have demonstrated high sensitivity and specificity in distinguishing active TB from other respiratory diseases or latent infection in HIV-coinfected individuals (35). These biomarkers may support early therapeutic interventions and help prioritize individuals for preventive therapy.

Transcriptomic biomarkers also aid in disease stratification by reflecting the extent and nature of immune activation. Gene expression patterns involving type I interferons, inflammation, and neutrophil activation have been correlated with more severe disease phenotypes and greater bacterial burden (36). Stratification based on these markers can inform risk-based treatment strategies, optimizing clinical outcomes and resource allocation.

Moreover, dual RNA-seq data allow for parallel pathogen-derived transcript quantification. Pathogen RNA load and expression of virulence or dormancy-related genes may serve as correlates of infection stage or treatment response (37). Thus, transcriptomic biomarkers, both host- and pathogen-derived, have potential as minimally invasive, blood-based tools for TB diagnosis, prognostication, and monitoring. Their integration into clinical workflows, however, will require standardized platforms, cross-population validation, and regulatory endorsement (38).

8.2. Predictive Models for Drug Response and Resistance Using Transcriptomic Data

Transcriptomic data are increasingly being applied to predict drug response and resistance in TB, supporting the personalization of treatment regimens and surveillance of resistance emergence. Differential gene expression profiles of both the host and *Mycobacterium tuberculosis* during treatment can offer insights into therapeutic efficacy, host tolerance, and the potential for resistance development (39). By monitoring changes in transcriptomic signatures over time, researchers can assess whether treatment is suppressing the bacterial transcriptional response or inadvertently selecting for resistant subpopulations.

On the bacterial side, RNA-seq has identified genes whose upregulation correlates with drug exposure. For instance, increased expression of the efflux pump genes Rv1258c and Rv1410c has been linked to isoniazid and rifampicin

tolerance, respectively (40). Similarly, transcriptional activation of stress response genes such as sigE and dnaK is frequently observed in response to antimicrobials and may predict subclinical resistance or tolerance phenotypes (41).

Host transcriptomic responses can also serve as indirect indicators of treatment success. A decline in inflammation-related genes (e.g., IL1B, CXCL10) and restoration of homeostatic gene expression patterns have been associated with treatment efficacy (42). Machine learning models trained on these temporal transcriptomic patterns can classify responders and non-responders with high accuracy, enabling dynamic treatment adjustments.

Moreover, predictive models that incorporate transcriptomic features have been developed to identify potential resistance mutations even before phenotypic resistance is apparent. Combining expression-based classifiers with genotypic data enhances the sensitivity and specificity of resistance prediction (43).

These approaches pave the way for transcriptome-guided TB therapy, wherein real-time transcriptional monitoring could tailor drug combinations, reduce treatment duration, and mitigate resistance evolution. Implementing these predictive tools in clinical practice, however, will require scalable platforms, integrated bioinformatics, and extensive validation (44).

8.3. Integration with Multi-Omics and Machine Learning Frameworks

To fully realize the potential of transcriptomics in TB research and clinical application, integration with multi-omics and machine learning frameworks is essential. Multi-omics approaches combine transcriptomics with genomics, proteomics, metabolomics, and epigenomics to provide a systems-level understanding of host-pathogen interactions (45). These data layers reveal complementary insights—for example, while transcriptomics captures gene activity, proteomics and metabolomics provide functional context in terms of protein abundance and metabolic flux.

Machine learning algorithms, including random forests, support vector machines (SVM), and deep learning models, are increasingly employed to integrate and analyze high-dimensional omics datasets. These tools can identify complex patterns and construct predictive models for diagnosis, treatment response, or disease progression (46). Feature selection techniques help pinpoint the most informative biomarkers, improving interpretability and reducing computational burden.

Integration platforms such as Multi-Omics Factor Analysis (MOFA) and Similarity Network Fusion (SNF) are used to harmonize disparate data types, while visualization tools like t-SNE and UMAP enable intuitive exploration of multi-omics landscapes. These frameworks enhance biological discovery and accelerate translation to precision medicine.

Ultimately, the convergence of transcriptomics, multi-omics, and AI-based analytics will unlock new biomarkers, drug targets, and clinical tools for TB control in diverse global settings (47).

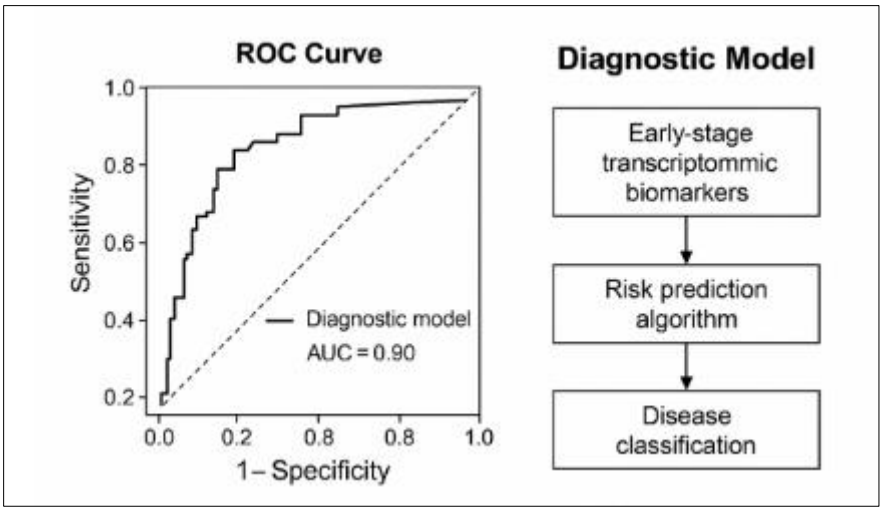

Figure 4 Diagnostic model incorporating early-stage transcriptomic biomarkers with ROC curve performance

9. Challenges and future perspectives

9.1. Limitations of Sample Quality, Heterogeneity, and Model Systems

Despite the transformative potential of transcriptomics in tuberculosis (TB) research, several limitations constrain its applicability and interpretation. One primary challenge is the variability in sample quality. RNA is inherently labile, and degradation due to improper handling, prolonged storage, or suboptimal extraction protocols can compromise data integrity, leading to biased or incomplete transcriptomic profiles (36). This issue is especially pertinent in field settings where biospecimen collection conditions may be inconsistent.

Another limitation arises from biological heterogeneity. Clinical TB samples—especially those derived from sputum, blood, or granulomatous tissue—contain diverse cell populations, including immune cells at various activation states, epithelial cells, and extracellular matrix components (37). This heterogeneity can obscure cell-specific transcriptional signals and complicate downstream analyses. Bulk RNA-seq, in particular, averages signals across mixed populations, potentially masking meaningful expression changes in rare or functionally distinct cells.

Moreover, the choice of experimental model systems introduces limitations in data extrapolation. In vitro models often lack the complexity of in vivo host-pathogen interactions, while murine models, although widely used, do not fully recapitulate the immunopathological features of human TB (38). Differences in gene regulation, immune cell composition, and lung architecture between species can lead to discordant findings. As a result, translating transcriptomic insights from experimental settings to human disease remains a challenge.

Overcoming these limitations requires standardized protocols for sample processing, stratified analyses to address heterogeneity, and validation across human-relevant models. Acknowledging and addressing these issues is essential for maximizing the reliability and translational value of TB transcriptomics (39).

9.2. Opportunities for Spatial Transcriptomics and AI-based Meta-analysis

Emerging technologies such as spatial transcriptomics and artificial intelligence (AI)-driven meta-analysis offer powerful opportunities to overcome current limitations and deepen our understanding of TB pathogenesis. Spatial transcriptomics allows gene expression to be mapped within intact tissue sections, preserving the spatial architecture of granulomas and other infection niches. This technology enables the identification of transcriptional gradients and cell-cell interactions that are critical for understanding localized immune responses and bacterial persistence (40).

In TB, spatial transcriptomics can reveal how distinct regions within granulomas—such as necrotic cores, peripheral immune zones, and fibrotic capsules—differ in gene expression and cellular composition (41). Such analyses may identify microenvironments that favor bacterial survival or immune suppression, offering targets for localized therapeutic intervention. Additionally, this approach can dissect the spatial co-expression of host and pathogen genes, providing insights into functional host-pathogen interfaces that were previously invisible in bulk transcriptomic analyses.

Parallel to these advances, AI-based meta-analysis can synthesize findings across heterogeneous datasets to identify robust, reproducible signatures of disease. Machine learning algorithms can integrate transcriptomic data from different cohorts, time points, tissues, and experimental platforms, controlling for confounding variables such as batch effects or demographic diversity (42). Meta-learning frameworks enable the generation of consensus biomarkers that perform consistently across populations and facilitate the discovery of novel regulatory networks involved in TB pathogenesis and resolution.

Moreover, AI models can predict clinical outcomes such as treatment response, relapse risk, or progression from latent to active disease using multi-modal data inputs. These predictive tools support precision medicine strategies, enabling more effective and individualized TB management (43). Cloud-based platforms and federated learning methods can also facilitate global data sharing while preserving patient privacy, further enhancing the impact of AI in TB research.

By harnessing spatial transcriptomics and AI-powered analytics, researchers can transcend current methodological constraints and generate actionable insights into the spatiotemporal dynamics of TB infection and immunity (44).

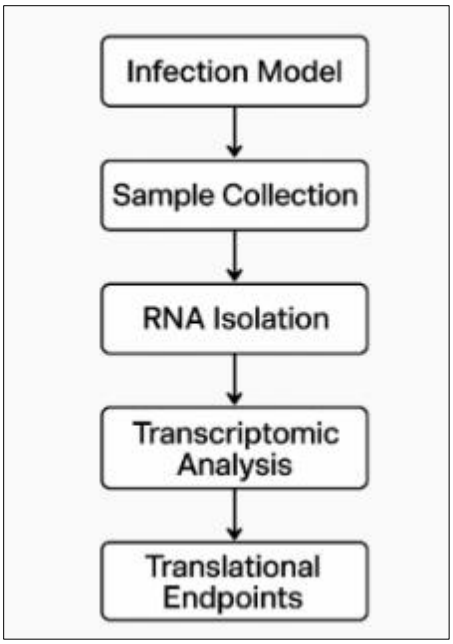

Figure 5 Flowchart summarizing pipeline from infection model to transcriptomic insight and translational endpoint

10. Conclusion

10.1. Recapitulation of Key Transcriptomic Insights in Early TB Infection

Transcriptomic studies have significantly advanced our understanding of host-pathogen interactions during the early stages of *Mycobacterium tuberculosis* (M. tuberculosis) infection. These insights have uncovered the complex molecular events that shape the trajectory of tuberculosis (TB) from initial exposure to latency or active disease. Early infection is characterized by rapid activation of innate immune signaling pathways, particularly those mediated by Toll-like receptors, type I and II interferons, and pro-inflammatory cytokines. Gene expression signatures within the first 6 to 24 hours post-infection reflect an intense host effort to contain the pathogen, with upregulation of IL1B, TNF, and CXCL10, accompanied by recruitment of monocytes and macrophages to the site of infection.

Simultaneously, M. tuberculosis activates stress response pathways, including the DosR dormancy regulon, and remodels its transcriptome to evade host defenses. The pathogen upregulates genes related to cell wall integrity, secretion systems (e.g., ESX-1), and lipid metabolism to ensure intracellular survival. Dual RNA-seq has enabled parallel profiling of both host and pathogen transcripts, revealing interdependent transcriptional programs and molecular crosstalk.

Temporal transcriptomic patterns have also shown how the immune system transitions from an inflammatory to a regulatory state, marked by increased expression of IL10, ARG1, and autophagy-related genes by 48 hours post-infection. These changes suggest a host attempt to mitigate tissue damage while simultaneously allowing pathogen persistence. Collectively, transcriptomic data provide a high-resolution, time-resolved view of early TB infection, identifying critical windows for intervention, key immune pathways that dictate disease outcomes, and molecular signatures that may serve as diagnostic or prognostic biomarkers in both latent and progressive forms of TB.

10.2. Research Implications for Host-Directed Therapies and Vaccine Design

The wealth of transcriptomic insights from early TB infection has profound implications for the development of host-directed therapies (HDTs) and next-generation vaccines. By dissecting host gene expression patterns that favor pathogen containment or disease progression, researchers can identify immune pathways that are amenable to therapeutic modulation. For example, augmenting protective responses such as autophagy, inflammasome activation, or specific cytokine signaling (e.g., IL-1β or IFN-γ) could enhance bacterial clearance without exacerbating tissue damage. Conversely, dampening excessive inflammation through selective targeting of pathways like type I interferon signaling may reduce immunopathology while preserving antimicrobial efficacy.

HDTs informed by transcriptomic data could complement conventional antibiotic therapy, particularly in multidrug-resistant TB, by restoring effective immune function and shortening treatment duration. These approaches also offer opportunities to repurpose existing immunomodulatory drugs—such as mTOR inhibitors or checkpoint blockade agents—based on their ability to shift transcriptional programs in favor of host defense.

Transcriptomic profiling also supports rational vaccine design by identifying antigens and host responses associated with protective immunity. Signatures of early containment, such as robust Th1-type responses, macrophage activation, and limited expression of immune suppressive genes, provide benchmarks for evaluating vaccine efficacy. Vaccine candidates that mimic these protective transcriptional profiles may offer improved protection over existing options like BCG.

Furthermore, transcriptomic biomarkers can guide vaccine trial stratification by identifying individuals with high-risk profiles for progression, allowing more precise assessment of efficacy. Overall, transcriptomic research offers a systems-level framework to develop tailored, multi-pronged strategies that not only prevent TB but also enhance therapeutic success in infected individuals.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] World Health Organization. Global tuberculosis report 2023. Geneva: WHO; 2023.

[2] Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, et al. Tuberculosis. Nat Rev Dis Primers. 2016; 2:16076. https://doi.org/10.1038/nrdp.2016.76

[3] Esmail H, Barry CE 3rd, Young DB, Wilkinson RJ. The ongoing challenge of latent tuberculosis. Philos Trans R Soc Lond B Biol Sci. 2014;369(1645):20130437. https://doi.org/10.1098/rstb.2013.0437

[4] Barry CE, Boshoff HIM, Dartois V, Dick T, Ehrt S, Flynn J, et al. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nat Rev Microbiol. 2009;7(12):845–55. https://doi.org/10.1038/nrmicro2236

[5] Jo EK, Yang CS, Choi CH, Harding CV. Intracellular signalling cascades regulating innate immune responses to Mycobacteria: branching out from Toll-like receptors. Cell Microbiol. 2007;9(5):1087–98. https://doi.org/10.1111/j.1462-5822.2007.00909.x

[6] Manca C, Reed MB, Freeman S, Mathema B, Kreiswirth B, Barry CE, et al. Differential monocyte activation underlies strain-specific Mycobacterium tuberculosis pathogenesis. Infect Immun. 2004;72(9):5511–4. https://doi.org/10.1128/IAI.72.9.5511-5514.2004

[7] Kumar NP, Moideen K, Nancy A, Viswanathan V, Shruthi BS, Sivakumar S, et al. Plasma chemokines are early biomarkers of active pulmonary tuberculosis. Sci Rep. 2019; 9:18290. https://doi.org/10.1038/s41598-019-54433-z

[8] Teles RM, Graeber TG, Krutzik SR, Montoya D, Schenk M, Lee DJ, et al. Type I interferon suppresses type II interferon-triggered human anti-mycobacterial responses. Science. 2013;339(6126):1448–53. https://doi.org/10.1126/science.1233665

[9] Mehra S, Pahar B, Dutta NK, Conerly CN, Philippi-Falkenstein K, Alvarez X, et al. Transcriptional reprogramming in nonhuman primate lungs infected with Mycobacterium tuberculosis. J Infect Dis. 2010;202(11):1677–89. https://doi.org/10.1086/657321

[10] Kiran D, Podell BK, Chambers M, Basaraba RJ. Host-directed therapy targeting the Mycobacterium tuberculosis granuloma: a review. Semin Immunopathol. 2016;38(2):167–83. https://doi.org/10.1007/s00281-015-0521-0

[11] Manca C, Tsenova L, Freeman S, Barczak AK, Tovey M, Murray PJ, et al. Hypervirulent Mycobacterium tuberculosis W/Beijing strains upregulate type I interferons and increase pathogenesis. J Exp Med. 2005;202(11):1551–60. https://doi.org/10.1084/jem.20051119

[12] Berry MPR, Graham CM, McNab FW, Xu Z, Bloch SAA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature. 2010;466(7309):973–7. https://doi.org/10.1038/nature09247

[13] Singhania A, Verma R, Graham CM, Lee J, Tran T, Richardson M, et al. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. Nat Commun. 2018; 9:2308. https://doi.org/10.1038/s41467-018-04579-w

[14] Moreira-Teixeira L, Sousa J, McNab FW, Torrado E, Cardoso F, Machado H, et al. Type I IFN exacerbates disease in tuberculosis-susceptible mice by inducing neutrophil-mediated lung inflammation and NETosis. Nat Commun. 2020; 11:5566. https://doi.org/10.1038/s41467-020-19484-4

[15] Luo Y, Riedel I, Dos Santos JC, Huth A, Rojas LA, Härtlova A. Transcriptional regulation of innate immunity and cell death pathways by type I interferons in tuberculosis. Front Cell Infect Microbiol. 2022; 12:840973. https://doi.org/10.3389/fcimb.2022.840973

[16] Rao PK, Missiakas D, Schneewind O. Glycolipids and Mycobacterium tuberculosis pathogenesis. Pathog Dis. 2021;79(8): ftab056. https://doi.org/10.1093/femspd/ftab056

[17] Arnett E, Weaver AM, Woodyard KC, Achee NL, Kolivras KN, Larkin TE, et al. Dual RNA-seq reveals lung epithelial and Mycobacterium tuberculosis transcriptome interactions. PLoS Pathog. 2022;18(5): e1010377. https://doi.org/10.1371/journal.ppat.1010377

[18] Rustad TR, Sherrid AM, Minch KJ, Sherman DR. Hypoxia: a window into Mycobacterium tuberculosis latency. Cell Microbiol. 2009;11(8):1151–9. https://doi.org/10.1111/j.1462-5822.2009.01325.x

[19] Leistikow RL, Morton RA, Bartek IL, Frimpong I, Wagner K, Voskuil MI. The Mycobacterium tuberculosis DosR regulon assists in metabolic homeostasis and enables rapid recovery from nonrespiring dormancy. J Bacteriol. 2010;192(6):1662–70. https://doi.org/10.1128/JB.00926-09

[20] Shukla P, Crook A, Pelosi AJ, Tiberi S, Koh GC. Host-directed therapies for tuberculosis: can we harness the host immune system to combat TB? Int J Infect Dis. 2022; 122:353–61. https://doi.org/10.1016/j.ijid.2022.05.055

[21] Moguche AO, Shafiani S, Clemons C, Larson RP, Dinh C, Higdon LE, et al. ICOS and Bcl6-dependent pathways maintain a CD4 T cell population with memory-like properties during tuberculosis. J Exp Med. 2015;212(5):715–28. https://doi.org/10.1084/jem.20142110

[22] Sia JK, Georgieva M, Rengarajan J. Innate immune defenses in human tuberculosis: an overview of the interactions between Mycobacterium tuberculosis and innate immune cells. J Immunol Res. 2015; 2015:747543. https://doi.org/10.1155/2015/747543

[23] Penn-Nicholson A, Mbandi SK, Thompson EG, Mendelsohn SC, Suliman S, Erasmus M, et al. RISK6, a 6-gene transcriptomic signature of TB disease risk, diagnosis, and treatment response. Sci Transl Med. 2019;11(505): eaaw2617. https://doi.org/10.1126/scitranslmed.aaw2617

[24] Scriba TJ, Penn-Nicholson A, Shankar S, Hraha T, Thompson EG, Sterling D, et al. Sequential inflammatory processes define human progression from M. tuberculosis infection to tuberculosis disease. PLoS Pathog. 2017;13(11): e1006687. https://doi.org/10.1371/journal.ppat.1006687

[25] Zhu J, Zhao Y, Wang X, Song Q, Jiang L, Li Y, et al. Dynamic transcriptomics of Mycobacterium tuberculosis-infected macrophages reveals pathogen strategies for immune evasion and persistence. Infect Immun. 2020;88(2): e00493-19. https://doi.org/10.1128/IAI.00493-19

[26] Pisu D, Huang L, Grenier JK, Russell DG. Dual RNA-seq of M. tuberculosis-infected macrophages in vivo reveals ontologically distinct host-pathogen interactions. Cell Rep. 2020;30(2):335–350.e4. https://doi.org/10.1016/j.celrep.2019.12.028

[27] Liu Q, Pan Y, Xu S, Gao Y, Wang X, Liu J, et al. Transcriptomic signatures of different Mycobacterium tuberculosis lineages reveal distinct virulence and host immune response profiles. PLoS Pathog. 2021;17(5): e1009604. https://doi.org/10.1371/journal.ppat.1009604

[28] Avraham R, Haseley N, Fan A, Bloom-Ackermann Z, Livny J, Hung DT. A highly multiplexed and sensitive RNA-seq protocol for simultaneous analysis of host and pathogen transcriptomes. Nat Protoc. 2016;11(8):1477–91. https://doi.org/10.1038/nprot.2016.084

[29] Temitope Abiodun Ogunkoya. Transforming hospital-acquired infection control through interdisciplinary, evidence-based nursing bundles in U.S. acute care. Int J Eng Technol Res Manag [Internet]. 2022 Dec ;6(12). Available from: https://doi.org/10.5281/zenodo.15533974

[30] Traag VA, Doorenweerd C, Boone CE, Eggenhuisen TM, van der Valk J, Boehme CC, et al. Dual RNA sequencing of blood from TB patients reveals an interferon signature related to disease severity. Front Immunol. 2021; 12:625084. https://doi.org/10.3389/fimmu.2021.625084

[31] Nair S, Raghunandanan S, Bishai WR, Ramakrishnan L, George B. Transcriptomic profiling of tuberculosis progression and treatment response: a systematic review. Front Genet. 2021; 12:657964. https://doi.org/10.3389/fgene.2021.657964

[32] Stanley SA, Cox JS. Host–pathogen interactions during Mycobacterium tuberculosis infections. Pathogenesis of Mycobacterium tuberculosis and its interaction with the host organism. 2013 Jul 24:211-41.

[33] Pieters J. Mycobacterium tuberculosis and the macrophage: maintaining a balance. Cell host & microbe. 2008 Jun 12;3(6):399-407.

[34] Agyemang, Cindy. 2024. "Variations in the Impact of Racial Attitudes on State-Level Policy Diffusion." APSA Preprints. doi: 10.33774/apsa-2024-jfd2n.

[35] Chandra P, Grigsby SJ, Philips JA. Immune evasion and provocation by Mycobacterium tuberculosis. Nature Reviews Microbiology. 2022 Dec;20(12):750-66.

[36] Jayachandran R, BoseDasgupta S, Pieters J. Surviving the macrophage: tools and tricks employed by Mycobacterium tuberculosis. InPathogenesis of Mycobacterium tuberculosis and its Interaction with the Host Organism 2012 Nov 16 (pp. 189-209). Berlin, Heidelberg: Springer Berlin Heidelberg.

[37] Ekundayo F, Ikumapayi OJ. Leadership practices in overseeing data engineers developing compliant, high-performance REST APIs in regulated financial technology environments. Int J Comput Appl Technol Res. 2022;11(12):566–577. doi:10.7753/IJCATR1112.1021.

[38] BoseDasgupta S, Pieters J. Macrophage-microbe interaction: lessons learned from the pathogen Mycobacterium tuberculosis. InSeminars in immunopathology 2018 Nov (Vol. 40, No. 6, pp. 577-591). Berlin/Heidelberg: Springer Berlin Heidelberg.

[39] Dey B, Bishai WR. Crosstalk between Mycobacterium tuberculosis and the host cell. InSeminars in immunology 2014 Dec 1 (Vol. 26, No. 6, pp. 486-496). Academic Press.

[40] Stutz MD, Clark MP, Doerflinger M, Pellegrini M. Mycobacterium tuberculosis: Rewiring host cell signaling to promote infection. Journal of leukocyte biology. 2018 Feb;103(2):259-68.

[41] Natarajan K, Kundu M, Sharma P, Basu J. Innate immune responses to M. tuberculosis infection. Tuberculosis. 2011 Sep 1;91(5):427-31.

[42] Khan A, Singh VK, Hunter RL, Jagannath C. Macrophage heterogeneity and plasticity in tuberculosis. Journal of leukocyte biology. 2019 Aug;106(2):275-82.

[43] Boom WH, Canaday DH, Fulton SA, Gehring AJ, Rojas RE, Torres M. Human immunity to M. tuberculosis: T cell subsets and antigen processing. Tuberculosis. 2003 Feb 1;83(1-3):98-106.

[44] Raymond Antwi Boakye, George Gyamfi, Cindy Osei Agyemang. DEVELOPING REAL-TIME SECURITY ANALYTICS FOR EHR LOGS USING INTELLIGENT BEHAVIORAL AND ACCESS PATTERN ANALYSIS. International Journal of Engineering Technology Research & Management (IJETRM). 2023Jan21;07(01):144–62.

[45] Ekundayo F. Strategies for managing data engineering teams to build scalable, secure REST APIs for real-time FinTech applications. Int J Eng Technol Res Manag. 2023 Aug;7(8):130. Available from: https://doi.org/10.5281/zenodo.15486520

[46] Ugwueze VU, Chukwunweike JN. Continuous integration and deployment strategies for streamlined DevOps in software engineering and application delivery. Int J Comput Appl Technol Res. 2024;14(1):1–24. doi:10.7753/IJCATR1401.1001.

[47] Bhat KH, Yaseen I. Mycobacterium tuberculosis: macrophage takeover and modulation of innate effector responses. Mycobacterium-Research and Development. 2018 Jun 20.

[48] Flynn JL, Chan J. Immune evasion by Mycobacterium tuberculosis: living with the enemy. Current opinion in immunology. 2003 Aug 1;15(4):450-5.

[49] Russell DG. Mycobacterium tuberculosis: here today, and here tomorrow. Nature reviews Molecular cell biology. 2001 Aug;2(8):569-78.

[50] Mueller P, Pieters J. Modulation of macrophage antimicrobial mechanisms by pathogenic mycobacteria. Immunobiology. 2006 Sep 14;211(6-8):549-56.