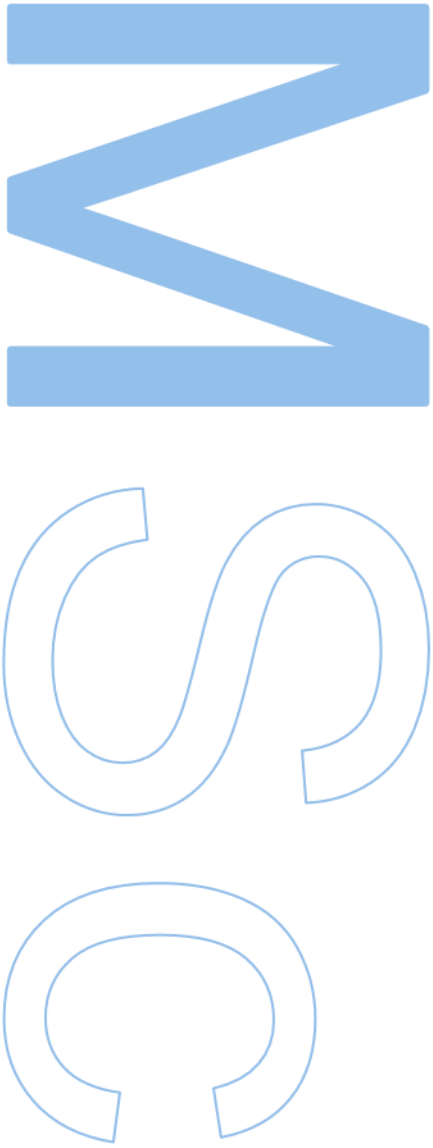
Understanding the impact of *Mycobacterium tuberculosis* complex diversity in tuberculosis

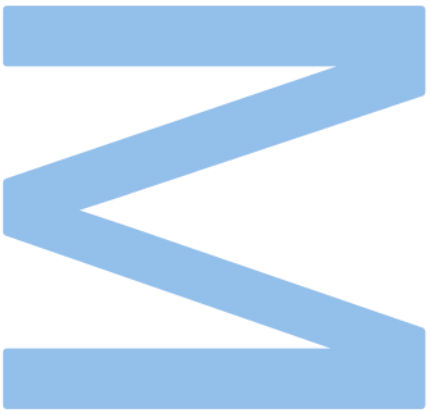
Ana Raquel Maceiras de Oliveira
Master's degree in Bioinformatics and Computational Biology
Department of Computer Science
2022

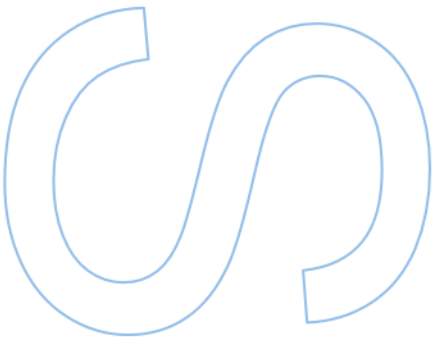
Supervisor
Margarida Saraiva, Group Leader
i3S – Instituto de Investigação e Inovação em Saúde

Co-supervisor
Pedro G. Ferreira, Professor Auxiliar
Faculdade de Ciências da Universidade do Porto




U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO





Dedicated to Carlos & Miguel

Sworn Statement

I, Ana Raquel Maceiras de Oliveira, enrolled in the Master's degree in Bioinformatics and Computational Biology at the Faculty of Sciences of the University of Porto hereby declare, in accordance with the provisions of paragraph a) of Article 14 of the Code of Ethical Conduct of the University of Porto, that the content of this dissertation reflects perspectives, research work and my own interpretations at the time of its submission.

By submitting this dissertation, I also declare that it contains the results of my own research work and contributions that have not been previously submitted to this or any other institution.

I further declare that all references to other authors fully comply with the rules of attribution and are referenced in the text by citation and identified in the bibliographic references section. This dissertation does not include any content whose reproduction is protected by copyright laws.

I am aware that the practice of plagiarism and self-plagiarism constitute a form of academic offence.

Ana Raquel Maceiras de Oliveira

24th November 2022

Acknowledgements

I would like to thank my supervisors, Dr Margarida Saraiva and Dr Pedro Ferreira, for giving me the possibility to develop this work. I would also like to acknowledge the Immune Regulation group at i3s, especially Dr Baltazar Ca, without whom this work would not have been possible, and Marta Silva. I would also like to thank everyone else that was involved in this work, from the team in Bissau that collected the samples, to the GenCore facility at EMBL where the sequencing was performed.

Lastly, I would like to thank my family for all the help and support without which I would not have been able to pursue this master's degree.

Resumo

A tuberculose (TB), uma das doenças com maior mortalidade a afetar a humanidade, é causada por duas bactérias altamente relacionadas e adaptadas aos humanos: o *Mycobacterium tuberculosis*, que se encontra distribuído por todo mundo e incluí as linhagens 1 (L1), L2, L3, L4, L7 e L8, e o *Mycobacterium africanum*, cuja incidência se encontra geograficamente restrita e que incluí as L5, L6 e L9. Apesar de o *M. africanum* ser muito menos estudado que o *M. tuberculosis*, várias diferenças entre estes dois patogénios foram já identificadas, as quais levantam a hipótese de o *M. africanum* ser menos virulento que o *M. tuberculosis*. No entanto, potenciais diferenças entre a resposta imune observada em indivíduos infetados com *M. africanum* ou *M. tuberculosis* não foram ainda estudadas.

Este trabalho tem como objetivo identificar essas diferenças tirando partido de um coorte de indivíduos com TB da Guiné-Bissau, um país com uma alta incidência de TB causada por estirpes pertencentes à L6 de *M. africanum*. Para isso, procedeu-se à análise do transcriptoma em amostras de sangue total de indivíduos com TB causada por estirpes pertencentes à L4 de *M. tuberculosis*, à L6 de *M. africanum*, ou co-infetados com estirpes pertencentes às duas linhagens anteriores, de forma a revelar diferenças entre as respostas imunes correspondentes.

A análise inicial dos dados de RNAseq obtidos revelou que o perfil transcripcional do sangue total era semelhante para os três grupos de infeção, uma vez que a maioria dos genes diferencialmente expressos (DEGs) era comum entre os três grupos. O perfil transcripcional observado era também semelhante ao perfil previamente reportado em indivíduos com TB. No entanto, o número de DEGs para o grupo L4 era superior ao observado para os outros grupos, o que poderá indicar uma resposta imune mais acentuada nos indivíduos infetados com estirpes pertencentes à L4 de *M. tuberculosis*. Seguidamente, efetuou-se uma análise de vias alteradas (“pathway analysis”) que permitiu confirmar o diferente nível de ativação do sistema imune entre o grupo L4 e o L6. Para além disso, verificou-se que a via de sinalização por recetores nucleares (no original, “signaling by nuclear receptors pathway”) apresentava diferentes níveis de ativação entre os grupo L4 e L6, ao mesmo tempo que apresentava uma relação positiva com o grau de severidade da TB observada. Por último, uma vez que o gene PDK1 da via de sinalização por recetores nucleares é também um constituinte da cascata de sinalização TLR/MyD88/PDK1, realizou-se um ensaio *in vitro* cujos resultados indicam que, enquanto a infeção com uma bactéria pertencente à L4 de *M. tuberculosis* parece

resultar na ativação da cascata de sinalização TLR/MyD88, e na consequente produção de TNFα pelas células infetadas, o mesmo não se verificou após infeção com bactéria pertencente à L6 de *M. africanum*, uma vez que a infeção não resultou na produção de TNFα pelas células infetadas.

Em conclusão, este estudo permitiu a identificação de diferenças importantes entre a resposta imune observada em indivíduos infetados com estirpes pertencentes às L4 de *M. tuberculosis* e L6 de *M. africanum*, as quais podem explicar diferenças previamente reportadas entre a TB causada por estas duas linhagens. Além disso, as diferenças reveladas entre a resposta imune observada em infeções por estirpes das duas linhagens fornecem novos caminhos para a descoberta de alvos terapêuticos para a TB.

Palavras-chave: Tuberculose, *Mycobacterium africanum*, *Mycobacterium tuberculosis*, RNAseq de sangue total

Abstract

Tuberculosis (TB), one of the deadliest threats to human health, is caused by two highly related and human-adapted bacteria: the widely spread *Mycobacterium tuberculosis*, comprising lineage 1 (L1), L2, L3, L4, L7, and L8, and the geographically restricted *Mycobacterium africanum*, that includes L5, L6, and L9. Even though *M. africanum* remains less studied than *M. tuberculosis,* several differences have been identified between these two pathogens, which raise the hypothesis that *M. africanum* might be attenuated when compared to *M. tuberculosis*. However, the differences in the immune responses triggered by *M. africanum* vs *M. tuberculosis* in humans are still elusive.

This work aimed to reveal differences in the immune response triggered by *M. tuberculosis* vs *M. africanum* by taking advantage of a new TB cohort from Guinea-Bissau, a country with a high incidence of TB caused by *M. africanum* L6 strains. For that, transcriptomic analysis was performed on whole blood samples from individuals infected with *M. tuberculosis* L4 strains (L4 group), *M. africanum* L6 strains (L6 group) or coinfected with both *M. tuberculosis* L4 strains and *M. africanum* L6 strains (L4+L6 group), to reveal differences in the corresponding immune responses.

Initial analysis of the RNAseq data showed that the overall blood transcriptomic profile was similar between the three groups of infection, with many of the differentially expressed genes (DEGs) being common between the three groups. In addition, the obtained whole blood transcriptomic profile was also compatible with that previously reported for another cohort of TB patients. Nevertheless, a considerable number of DEGs in L4-infected patients were not observed in the other infections, indicating that the immune system was more activated in individuals infected with *M. tuberculosis* L4 strains than in individuals infected with *M. africanum* L6 strains. Through pathway analysis, the differential activation status of the immune response in individuals from these two groups was further confirmed. In addition, pathway analysis allowed the identification of the signaling by nuclear receptors pathway, which distribution of activation was different for the L4 and the L6 group. This pathway also presented a significant positive relation with disease severity. Lastly, since the PDK1 gene from the signaling by nuclear receptors pathway is part of the TLR/MyD88/PDK1 signaling cascade, an *in vitro* experiment was performed which results indicate that infection with an *M. tuberculosis* L4 strain leads to the activation of this pathway, as it resulted in the production of TNFα by the infected cell. However, infection with an *M. africanum* L6 strain

does not seem to lead to the activation of this signaling cascade as it did not result in the production of TNFα by the infected cells.

In conclusion, this study allowed the identification of key differences in the immune response observed in individuals infected with *M. tuberculosis* L4 or *M. africanum* L6 strains, which may explain the previously reported differences in TB caused by strains belonging to these two lineages. In addition, the differences in the host immune response triggered by *M. tuberculosis* vs *M. africanum*, here reported, provide new insights into potential therapeutic targets for TB, albeit requiring further experimental investigation.

Keywords: Tuberculosis, *Mycobacterium africanum*, *Mycobacterium tuberculosis*, whole-blood RNAseq

Table of Contents

List of Figures

List of Abbreviations

BCR	B cell receptor
BHP	Bandim Health Project
BMDMs	Bone marrow-derived macrophages
bp	Base pairs
COVID-19	Coronavirus disease-19
CPM	Counts per million
DCs	Dendritic cells
DEG	Differentially expressed gene
FDR	False discovery rate
GSEA	Gene set enrichment analysis
HD	Healthy donor
HIV	Human immunodeficiency virus
IGRA	Interferon gamma release assays
IFNα/β	Interferon alpha/beta
IFNγ	Interferon gamma
IL	Interleukin
L	Lineage
LTBI	Latent tuberculosis infection
L4+L6	Coinfection with L4 and L6 strains
MDR	Multidrug-resistant
MOI	Multiplicity of infection
MTBC	*Mycobacterium tuberculosis* complex
NGS	Next-generation sequencing
NK cell	Natural killer cell
NKT cell	Natural killer T cell
ORA	Pathway over-representation analysis
PC	Principal component
PCA	Principal component analysis
RD	Region of difference
RNAseq	RNA sequencing
RORγt	Retinoic acid receptor-related orphan receptor gamma-T
RQS	RNA quality score
scRNAseq	Single-cell RNA sequencing
SNP	Single nucleotide polymorphism

TB	Tuberculosis
T-BET	T-box expressed in T cells
TCR	T cell receptor
Th1	Type 1 T helper
Th17	Type 17 T helper
TNFα	Tumour necrosis factor alpha
Treg	Regulatory T cell
WGS	Whole genome sequencing
WHO	World Health Organization
XDR	Extreme drug-resistant

Introduction[1]

Tuberculosis (TB) is one of the oldest and deadliest diseases affecting mankind. It is caused by gram-positive bacteria belonging to the *Mycobacterium tuberculosis* complex (MTBC), identified for the first time in 1882 by Robert Koch (Koch, 1882). TB has affected the human population since ancient times and remains a serious public health problem despite the fact of being, in most cases, a curable disease. In 2020 alone, it was responsible for 1.5 million deaths worldwide (WHO, 2021). The annual number of deaths caused by TB had been decreasing over the years (WHO, 2022). However, due to the Coronavirus disease-19 (COVID-19) pandemic, this number increased in 2020 even though the number of reported cases decreased substantially (WHO, 2021). Indeed, the pandemic has compromised access to TB diagnosis and treatment, especially in low-income countries (McQuaid et al., 2020). Thus, the global TB burden may increase in the next years, jeopardizing the efforts made to eradicate TB, as determined by the United Nations' goals.

Besides the COVID-19 pandemic, several other elements compromise the successful eradication of TB. For once, it is estimated that one-fourth of the world population is infected with *M. tuberculosis* but remains asymptomatic, a condition termed latent TB infection (LTBI). Of these latently infected individuals, 5-10% will develop TB during their lifetime (Vynnycky and Fine, 2000). In addition, the rising number of TB cases caused by multidrug-resistant (MDR) and extreme drug-resistant (XDR) *M. tuberculosis* strains, for which the standard antibiotic treatments are not effective, the co-infections with the human immunodeficiency virus (HIV), and the lack of an effective vaccine further hamper the goal of TB elimination (WHO, 2021). Hence, the development of new strategies to prevent, diagnose, and treat TB is of major importance, which in turn depends on a better knowledge of this disease at all levels.

1.1 The immune response in tuberculosis

The immune response to *Mycobacterium tuberculosis* infection in the lung is very complex and involves several cascade events (Figure 1) (O'Garra et al., 2013). Even though *M. tuberculosis* can disseminate into other organs such as lymph nodes, bone, and meninges, and cause extramedullary TB, pulmonary TB is the main form of TB, accounting for ~70% of the total TB cases and will be the solo focus of this section.

[1]Section partly published in Silva et al., 2022

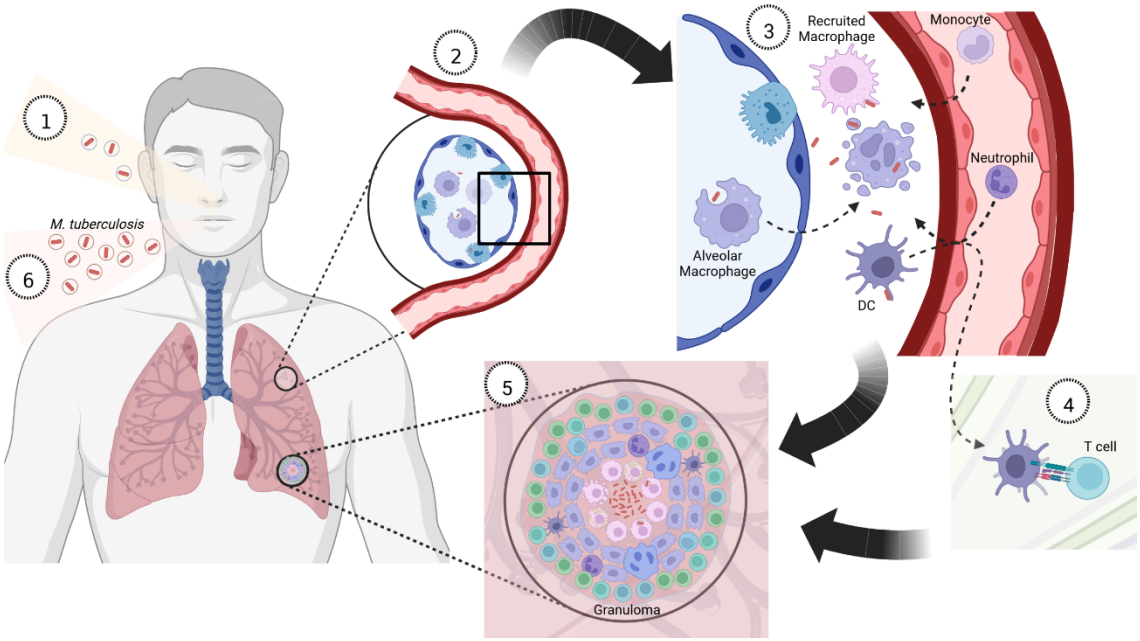


Figure 1. Schematics of the immune response to *M. tuberculosis* infection.
M. tuberculosis bacteria are uptake by inhalation (1) and travel into the lower airways and alveoli. At this location, the pathogen starts infecting different cell types including alveolar macrophages (2). Infected alveolar macrophages then migrate into the alveolar interstitium where undergo apoptosis and release bacteria that are uptake by monocyte-derived recruited macrophages, neutrophils, and DCs; more innate cells are recruited into the local of infection (3). DCs that have uptaken *M. tuberculosis* migrate into draining lymph nodes where they present bacterial antigens to T cells to activate these cells and initiate the adaptive immune response (4). The accumulation of both innate and adaptive immune cells then leads to the formation of granulomas (5). If the host is not able to control bacteria proliferation and consequent necrosis at the core of the granuloma, the granuloma disrupts leading to the expel of *M. tuberculosis* when coughing and, consequently, infecting other individuals (6). Figure created with BioRender.com

Much knowledge has been obtained from studying the infection in humans. However, obtaining human samples for such studies is challenging. Most of the studies only have access to blood samples and, to a less extent, bronchoalveolar lavage samples, from individuals with active TB. Animal models, especially murine ones, have been used through the years to unveil the complex mechanisms ongoing during *M. tuberculosis* infection. Nevertheless, the disease in these models is slightly different from the one observed in humans, as animal models are not able to mimic latent infection, the same *M. tuberculosis* strain may cause very different disease severities in different mouse strains, and the severity observed in mice may not reflect the one observed in humans (Moreira-Teixeira et al., 2020; O'Garra et al., 2013). Nevertheless, these models have been key to studying TB and many of the most important characteristics of TB are common between these models and the disease observed in humans.

TB is an airborne disease that is transmitted when an individual with active TB coughs. The initial symptoms of TB include a persistent cough and weight loss and the diagnosis is confirmed by a lung X-ray showing evidence of lung lesions and/or positive culture of *M. tuberculosis* from sputum. The lung-originated droplets containing *M. tuberculosis* that are released into the air can then be inhaled by a bystander individual (Figure 1,

section 1), which becomes infected when the bacteria reach the lower airways and alveoli (Figure 1, section 2). There are three main possible outcomes from initial contact with *M. tuberculosis* (Ainley and Kon, 2020). Some individuals may eliminate all bacteria very rapidly, thus not presenting any signals of ever being exposed. Others, as mentioned previously, become latently infected (latent TB infection - LTBI) where the bacteria may present in the system, but in a controlled state that in most cases does not evolve to an active TB state. LTBI can be identified by performing an interferon gamma (IFNγ) release assay (IGRA) or a tuberculin skin test. Both tests rely on determining whether the individual has been effectively infected and, consequently, has mounted a T cell-specific against *M. tuberculosis* antigens. Lastly, some individuals are not able to control *M. tuberculosis* infection, develop active TB, and become infectious, i.e., can spread the disease to others.

In the very early stages of infection, *M. tuberculosis* can then infect several cell types present in the lung including lung epithelia (mainly alveolar epithelial type II cells and to a less extent alveolar epithelial type I) and immune cells (alveolar macrophages, dendritic cells (DCs), neutrophils, and monocytes) (O'Garra et al., 2013; Scordo et al., 2016). Nevertheless, the initial disease outcome has been reported as being mainly dependent on alveolar macrophages (Cohen et al., 2018). In some individuals, the innate immune cells present in the lung, especially alveolar macrophages, are able to rapidly eliminate all bacteria. However, when the initial infection is not cleared out rapidly, alveolar macrophages then migrate into the lung interstitium where they undergo apoptosis leading to the release of bacteria at this site (Figure 1, section 3) (Cohen et al., 2018). These bacteria are in turn uptake by resident and recruited monocyte-derived macrophages, neutrophils, and DCs, fully initiating the innate immune response (Cohen et al., 2018; Eruslanov et al., 2005; Eum et al., 2010; Wolf et al., 2007). While recruited macrophages and neutrophils will stay on site trying to eliminate the bacteria present, DCs migrate into draining lymph nodes where they will present antigen peptides to both CD4 and CD8 T cells (Figure 1, section 4). T cells whose T cell receptors (TCR) successfully recognize the peptides being presented by DCs will become activated and differentiate, thus starting the adaptive immune response (Wolf et al., 2008). As an intracellular pathogen, the adaptive immune response triggered is characterized by the differentiation of naïve CD4 T cells into type 1 T helper (Th1) cells. Th1 cells are characterized by the expression of T-box expressed in T cells (T-BET) as the master transcription factor and the production of its major effector cytokine IFNγ, which is responsible for activating macrophages and potentiating their capacity to kill *M. tuberculosis* (Lugo-Villarino et al., 2003). The importance of CD4 T cells in TB is

demonstrated by the higher susceptibility presented by HIV patients and the incapacity of CD4-deficient mice to control infection by *M. tuberculosis* (Havlir and Barnes, 1999; Mogues et al., 2001). In the same line, IFNγ production by Th1 cells is required for host resistance to *M. tuberculosis* infection (Green et al., 2013). Besides Th1, some CD4 T cells differentiate into Th17. Indeed, Th17 cells have also been implicated in the ongoing immune response during TB although data collected so far indicate that these cells have a dual role. The master transcription factor expressed in these cells is the retinoic acid receptor-related orphan receptor gamma-T (RORγt) and their main effector cytokine is interleukin 17 (IL17) (Harrington et al., 2005; Park et al., 2005). While IL17 produced by Th17 cells can have a protective effect by promoting the recruitment of neutrophils, excessive IL17 and, consequently, excess of neutrophils results in tissue damage (Torrado and Cooper, 2010). Hence, both CD4 Th1 and Th17 have major roles in the immune response in TB since the cytokines produced by these cells further activate and potentiate the action of the effector cells. CD8 T cells are also key players in the immune response to *M. tuberculosis*. After becoming activated CD8 T cells produce IL2, IFNγ, and tumour necrosis factor alpha (TNFα), which promote innate immune cells' effector properties (Lin and Flynn, 2015; Silva et al., 2014). Moreover, CD8 T cells can recognise and kill infected cells via perforin, granzymes, and granulysin release, and the latest is also able to kill *M. tuberculosis* directly (Cho et al., 2000; Serbina et al., 2000; Stenger et al., 1998). Of note, natural killer (NK) cells and other T cells, such as γδT cells, natural killer T (NKT) cells, and CD4 regulatory T (Treg) Treg cells, also migrate into the lung where they play distinct roles in the established immune response (O'Garra et al., 2013). B cells are the other branch of the adaptive immune system. The main function of B cells is to produce antigen-specific antibodies that, in the context of infection, target pathogens or toxins produced by them. In the case of TB, the role of B cells is controversial, and their importance is yet not clear (Rijnink et al., 2021). Nevertheless, B cells expressing a B cell receptor (BCR) capable of successfully binding *M. tuberculosis* antigens will become activated and migrate into the lung, thus participating in the ongoing immune response (O'Garra et al., 2013).

The accumulation of both innate and adaptive immune cells in the lung leads to the formation of an immune structure denominated granuloma (Ramakrishnan, 2012). These structures aim to confine the bacteria and, consequently, control the spread of infection. Granulomas have a layered structure with the bacteria at the centre surrounded by innate immune cells: macrophages (including epithelioid, foamy, and multinucleated giant cells), neutrophils, DCs, and NK cells (Figure 1, section 5). The granulomas' outer part is composed of T and B lymphocytes. If the immune system can keep the bacteria under

control, i.e., without the uncontrolled proliferation of *M. tuberculosis* and the consequent growth of the granuloma, the individual will become latently infected. On the other hand, in individuals with active TB, the immune system is not able to control the proliferation of *M. tuberculosis*. This leads to the uncontrolled growth of the granulomas and massive cell death by necrosis at their centre. The necrotic centre of the granuloma, denominated caseous necrosis, ultimately leads to granuloma disruption. Individuals with disrupted granulomas become infectious, as *M. tuberculosis* bacteria are no longer contained within this structure and may be expelled from the lung when coughing.

1.2 High throughput transcriptomics to study TB

Since the first two studies that applied high throughput transcriptomics to study TB in 2001 (Ehrt et al., 2001; Ragno et al., 2001), both the transcriptomics field and the available resources and techniques have most evolved. While the first studies of TB transcriptomics used microarray to evaluate gene expression, the arise of next-generation sequencing (NGS) and the possibility of sequencing mRNA in a high through put manner (RNAseq) has provided a more complete and sensitive analysis tool to study TB. More recently, RNAseq at the single-cell level (scRNAseq) has also acquired a relevant place in the field while spatial transcriptomics has the potential of further contributing to the TB field in the next years.

The first two studies in the field of TB focused on uncovering gene expression changes in mouse and human macrophages infected *in vitro* with *M. tuberculosis* (Ehrt et al., 2001; Ragno et al., 2001). Since then, many other studies have used microarrays, RNAseq and scRNAseq to study transcriptomic changes in macrophages both *in vivo* and *in vitro* upon *M. tuberculosis* infection (Martinez-Perez et al., 2022). In addition, a great effort has been made to better understand human TB by studying the transcriptomic changes in blood samples from individuals with active TB or LTBI. Such studies' objectives were not only to characterize the host immune response, but also to identify better diagnostic tools using blood biomarkers, to find blood signatures that identify, at an early stage, individuals that will progress from an LTBI to an active TB state, and/or to characterize transcriptomic profile evolution during antibiotics treatment.

One of the pioneer works that accomplished several of these objectives was published in 2010 (Berry et al., 2010). In this study, by analysing whole blood transcriptome using microarrays, it was demonstrated that the blood transcriptome reflected disease extension (as determined by a radiological evaluation). Moreover, this study provided the

first evidence of the detrimental role of type I IFN, which includes IFN alpha and IFN beta (IFNα/β), in TB pathogenesis, as a type I IFN-inducible gene signature correlated with the extent of lung disease. In addition, the study revealed a 393-gene signature that discriminates between LTBI and active TB. Since then, many other studies that further characterized the host immune response and/or determined better biomarkers for diagnosis and disease progression have been published, most of them by performing whole blood RNAseq (Ahmed et al., 2020; Anderson et al., 2014; Bloom et al., 2012; Burel et al., 2021; Cai et al., 2014; Esaulova et al., 2021; Estevez et al., 2020; Gupta et al., 2020; Hoang et al., 2021; Jacobsen et al., 2007; Kaforou et al., 2013; Kim et al., 2010; Maertzdorf et al., 2011a; Maertzdorf et al., 2011b; Nathan et al., 2021; Roe et al., 2020; Singhania et al., 2018; Suliman et al., 2018; Tabone et al., 2021; Warsinske et al., 2019; Zak et al., 2016). More recently, another study compared TB signatures in mice (blood and lung) and human (blood) to establish links between mouse and human disease patterns and to provide further evidence on how the blood transcriptome may reflect the ongoing immune response in the lung (Moreira-Teixeira et al., 2020). Such comparisons are of major importance to establish better animal mouse models while providing evidence on how informative whole-blood transcriptome is on lung TB disease. Besides RNAseq, several studies have now used scRNAseq to provide further insights into the landscape of TB by characterizing the different immune populations' heterogeneity (Ahmed et al., 2020; Akter et al., 2022; Esaulova et al., 2021; Gideon et al., 2022). In addition, dual scRNAseq has also been used to study the transcriptome of host and pathogen simultaneously (Fontan et al., 2008; Pisu et al., 2020; Pisu et al., 2021; Rienksma et al., 2015; Zimmermann et al., 2017). These studies have mostly focused on macrophages and have provided insights, for example, into the heterogeneity of lung macrophage populations and the link between these and bacterial phenotypes (Pisu et al., 2020; Pisu et al., 2021). Lastly, one study has taken advantage of spatial transcriptomics to further dissect the granuloma structure in mice by analysing the development of this structure over time, in mice bearing encapsulated (as the ones observed in humans) or non-encapsulated granulomas (Carow et al., 2019).

Importantly, even though results from many studies have been made available over the years, most of them do not take into consideration the strain of *M. tuberculosis* causing the disease and how the *M. tuberculosis* diversity impacts the whole-blood transcriptome and, consequently, the immune response and disease observed.

1.3 The *Mycobacterium tuberculosis* Complex

TB-causing bacteria are not one restricted bacterial group but a wide range of phylogenetically related bacteria. Together, these bacteria form the MTBC (Brites and Gagneux, 2017), which is composed of nine human-adapted lineages (L) and animal-adapted bacteria (Brites and Gagneux, 2017; Napier et al., 2020). The human-adapted lineages can be separated into two main groups: *Mycobacterium tuberculosis sensu stricto* (from now on referred to as *M. tuberculosis*) which comprises L1, 2, 3, 4, 7, and 8, and *Mycobacterium africanum* which comprises L5, 6, and 9. In general, *M. tuberculosis* bacteria are considered generalist as the vast majority are spread throughout the globe (Comas et al., 2013; Gagneux et al., 2006). L4 is the most widespread lineage, being responsible for most TB cases in Europe, America, and Africa while still being found in Asia (Gagneux et al., 2006; Stucki et al., 2016). L1 and L3 are commonly found in South Asia and L2 is mostly found in Southeast Asia (Gagneux, 2018). The only *M. tuberculosis* L exceptions are L7 and L8 which are considered specialist lineages as they are restricted to Ethiopia and the Great Lakes regions, respectively (Firdessa et al., 2013; Ngabonziza et al., 2020). Interestingly, some sublineages of the generalist lineages are also geographically restricted and, therefore, considered specialists. This is the case of L4.1.3, L4.5, L4.6.1, and L4.6.2 (Stucki et al., 2016). On the other hand, L5, 6, and 9 comprise *M. africanum*. strains and are all classified as specialist lineages. L5 strains are more abundant on the east side of West Africa, whereas L6 strains appear mainly in the west part of West Africa (de Jong et al., 2009; Gehre et al., 2013; Groenheit et al., 2011; Homolka et al., 2008; Yeboah-Manu et al., 2016). The recently described L9 has so far only been identified in Somalia (Coscolla et al., 2021).

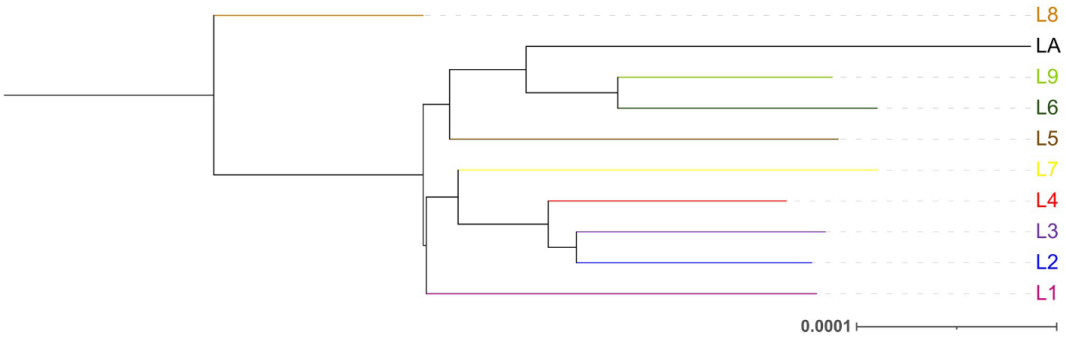


Figure 2. The phylogenetic structure of the MTBC.
Phylogenetic analysis of 10 MTBC genomes selected to include 1 genome from each of the known MTBC lineages (accession numbers: SRR1162469, ERR2704812, ERR181314, SRR10828835, ERR1193734, SRR8237291, ERR3470572, ERR3470655, ERR756344, and ERR015582. Figure from (Silva et al., 2022).

Besides their genetic and geographical differences, other differences have been observed between the different MTBC bacteria. It is the case of their different capacity to evade the immune system, the immune response triggered and, consequently, the disease severity observed (Bastos et al., 2017). Such differences have been observed not only at the lineage level but also between genetically related bacteria from the same sublineage.

1.4 *M. tuberculosis* vs *M. africanum*

Even though *M. africanum* remains a lot less studied than *M. tuberculosis*, several differences have been identified between these two pathogens (Silva et al., 2022). Genetically, *M. tuberculosis* and *M. africanum* strains are readily distinguishable by the deletion of the region of difference (RD) 9 in the genome of *M. africanum* strains. In addition, strains of the L5 have a further deletion in RD711, while L6 strains lack RD702 (Mostowy et al., 2004). Highly polymorphic regions of the genome can also be used to differentiate *M. tuberculosis* from *M. africanum*: L5 does not have spacers 8 to 12 and 37 to 39 and L6 does not have spacers 7 to 9 and 39 (de Jong et al., 2009; Mostowy et al., 2004). Moreover, several specific single nucleotide polymorphisms (SNPs) specific to strains belonging to L5 and L6 have been identified from whole genome sequencing (WGS) studies, that are now used in SNP assays for rapid identification of strains belonging to L5/L6 *M. africanum* lineages (Hershberg et al., 2008; Stucki et al., 2012). WGS has also revealed the existence of genetic diversity within *M. africanum* L5 and L6 strains and determined the phylogenetic closeness of different lineages through the construction of detailed MTBC phylogeny trees (Ates et al., 2018; Coscolla et al., 2021; Otchere et al., 2018; Sanoussi et al., 2021). Lastly, *M. africanum* strains also present important variations in some known virulence factors (e.g., DosR and phoP/R) of *M. tuberculosis,* which may account for their lower virulence (Broset et al., 2015; Gonzalo-Asensio et al., 2014; Ofori-Anyinam et al., 2017).

1.4.1 TB caused by *M. africanum* versus *M. tuberculosis*

There are no marked differences in TB caused by *M. tuberculosis* or *M. africanum*. Individuals infected with either *M. tuberculosis* or *M. africanum* have the same chest X-ray presentation and the response to the standard four first-line drugs in TB treatment is similar (de Jong et al., 2010; Diarra et al., 2018).

Despite these similarities in disease presentation, some differences have also been observed. Individuals with TB caused by L4 *M. tuberculosis* strains usually responded

faster to antibiotic treatment than those infected with L6 *M. africanum* strains (Diarra et al., 2018). Furthermore, TB caused by *M. africanum* strains is usually associated with more vulnerable hosts. *M. africanum* infections have been reported to be more common in HIV-coinfected individuals, older individuals, and individuals presenting severe malnutrition (Acquah et al., 2021; Baya et al., 2020; de Jong et al., 2010). Importantly, there seem to be no significant differences between the prevalence of *M. tuberculosis* or *M. africanum* infections in individuals with diabetes, another important comorbidity in TB (Ates et al., 2018). In addition, *M. africanum* infections have a slower progression to active TB (Baya et al., 2020; de Jong et al., 2008). In line with these findings, infection of mouse models with *M. africanum* strains showed a slower progression of the disease (Bold et al., 2012; Ca et al., 2019; Wiens and Ernst, 2016b) with mild lung pathology even in mice lacking IFNγ, which are highly susceptible to *M. tuberculosis* infection (Ca et al., 2019). Lastly, it is not clear whether transmission rates are different between *M. tuberculosis* and *M. africanum*. The two published studies that addressed this question reported contradictory results: while in the Gambia, similar transmission rates were reported for both *M. tuberculosis* and *M. africanum*, in Ghana, the transmission rate observed for *M. africanum* was lower than the one observed for *M. tuberculosis* (Asare et al., 2018; de Jong et al., 2008). Nonetheless, it may be the case that *M. africanum* L5 (predominant in Ghana) and L6 (predominant in the Gambia) have different transmission rates, which leads to the contradictory results published.

Altogether, several differences in *M. tuberculosis* vs *M. africanum* infections have been reported and *M. africanum* L5 and L6 strains (there is not much data available for the recently identified L9) are considered in general less virulent than strains of *M. tuberculosis* lineages. The reasons for the geographic restriction of *M. africanum* are still not clear to this day, but a host-pathogen mutual adaptation may have occurred, confining the ability of *M. africanum* to cause TB, and spread predominantly on a local ethnic group.

1.4.2 Host immune responses to *M. africanum*

As for other MTBC lineages, different *M. africanum* strains have been reported to induce different cytokine responses in macrophages infected *in vitro* (Reiling et al., 2013). However, it is still not clear whether *M. africanum* strains induce a weaker response from macrophages than *M. tuberculosis* (Reiling et al., 2013; Wiens and Ernst, 2016a). Moreover, it has been shown that *M. africanum* L6 strains induce fewer IFNβ *in vitro* than *M. tuberculosis* L2 and L4 strains (Wiens and Ernst, 2016a). However, the detrimental role of type I IFN in TB is still valid for lower amounts of type I IFN as mice lacking

type I IFN receptor signaling presented lower lung bacterial burdens and less severe histopathological findings upon *M. africanum* infection (Moreira-Teixeira et al., 2018; Wiens and Ernst, 2016b). The mechanisms used by macrophages to recognize *M. africanum* strains are still not very understood. One study reported that individuals infected with *M. africanum* isolates presented higher expression levels of TLR9 when compared with other lineages of the MTBC (Tientcheu et al., 2016). While these results indicate a role of TLR9 in the recognition of *M. africanum* strains by immune cells, another study that compared peripheral blood gene expression profiles between *M. africanum*- and *M. tuberculosis*-infected individuals reported no differences at diagnosis (Tientcheu et al., 2015).

In the same line as for macrophages, it is not clear if CD4 T cell responses in *M. tuberculosis* vs. *M. africanum* infections are different. The authors of the previously mentioned study, which compared peripheral blood gene expression profiles between individuals infected with *M. africanum* or *M. tuberculosis*, did not find differences in gene expression compatible with different T cell responses (Tientcheu et al., 2015). However, in this study, gene expression levels were measured using probe arrays, thus excluding from the analysis genes do not present in the array. In addition, another study showed that there were no differences in the frequencies of PPD-specific T cells in the blood of individuals with active TB caused by *M. tuberculosis* and *M. africanum* (Tientcheu et al., 2014). Once again, it is worth bearing in mind that L6 strains present an increased number of predicted epitopes that may result in more CD4 T cells with different TCRs being able to be activated and participate in the immune response (Coscolla et al., 2015). Interestingly, although frequencies of PPD-specific T cells in the blood were similar, the authors of this study did observe differences in the percentages of single-TNFα (increased) and single-IL2-producing CD4 T cells (decreased) after overnight *in vitro* stimulation with ESAT6 of whole blood samples from individuals infected with *M. africanum* vs *M. tuberculosis*. In another study, using the same system of whole blood stimulation *in vitro*, the authors observed higher levels of IFNγ being produced by blood samples from *M. tuberculosis* vs *M. africanum* after 2 and 6 months of anti-tuberculosis therapy (Tientcheu et al., 2016).

Although evidence points to *M. africanum* strains being less virulent and disease progression being slower after infection with these strains, whether there are differences in the immune response triggered by *M. tuberculosis* than *M. africanum* strains is still unclear.

2 Objectives

Despite being a very old and, in most cases, curable disease, TB continues to affect millions of people worldwide every year. To prevail against this disease, it is necessary to develop new strategies for prevention, diagnosis, and treatment. For that, it is essential to better understand the complex immune response mounted by the host upon infection with human-adapted bacteria of the MTBC and how bacteria from different lineages impact the onset and progression of the disease. In this study, we aimed to shed light on the first issue by studying the difference in host immune responses to bacteria of different lineages. In particular, this study focused on *M. tuberculosis* L4 and *M. africanum* L6, two lineages composed by pathogens associated with different geographic distributions and virulence.

Hence, the present work had the following objectives:

- Unveil the differences between the immune response observed in individuals infected with *M. tuberculosis* L4 strains and *M. africanum* L6 strains by performing RNAseq on whole-blood samples from infected and non-infected individuals;
- Identify pathways differentially expressed and, consequently, with different activation statuses in L4 vs L6 infections.

These objectives were accomplished by resorting to a new cohort of individuals with TB from Guinea-Bissau, a country with a high burden of TB cases, many of which are caused by *M. tuberculosis* L4 and *M. africanum* L6. The lineages of the bacteria causing the disease were established, after which RNAseq was performed in 58 whole-blood samples from healthy individuals (controls) and individuals with TB. Analysis of the transcriptome from these 58 individuals then allowed the characterization of the differences in the immune responses to *M. tuberculosis* vs *M africanum* and the identification of key pathways differentially expressed in the two groups of infection. Upon further testing and confirmation, these data are instrumental to guide hypotheses to be experimentally tested to guide novel therapies for TB.

3 Materials and Methods

3.1 Ethics statement

The study was approved by the *Comité Nacional de Ética na Saúde* (CNES), Guinea Bissau, reference number 003-CNES/INASA/2018. All participants signed an informed consent. To ensure confidentiality, and all data were anonymized by the assignment of a random identification number to each case. All experiments were conducted according to the principles expressed in the Declaration of Helsinki and the Oviedo Convention.

3.2 Study cohort and samples collection

A total of 372 individuals with active TB was recruited for the study. Sputum was collected and used to confirm the disease diagnosis and to perform lineage determination by SNP genotyping. In addition, for some patients, HIV testing was performed as part of the diagnostic procedure. Several clinical parameters were recorded to obtain a score of disease (TB score) (Rudolf et al., 2013). Blood samples of 13 individuals were also collected to be included as controls in the whole-blood RNAseq analysis.

3.3 RNA extraction and cDNA library preparation for RNAseq

Fifty-eight blood samples from HIV- individuals (45 from individuals with active TB and 13 from healthy individuals) were selected for RNAseq analysis. A volume of approximately 9ml of blood was collected into Tempus™ Blood RNA Tube (Applied Biosystems, cat. #4342792) followed by vigorous mixing and immediate storage at -80° C. Samples were shipped from Bissau to i3s in dry ice and stored at -80°C until further use. Total RNA was isolated using the Tempus™ Spin RNA Isolation Kit (Invitrogen, cat. #4380204), followed by depletion of globin RNA using the GLOBINclear™-Human Kit (Invitrogen, cat. #AM1980), both according to the manufacturer's instructions. The RNA yield, quality, and integrity of the globin-depleted RNA were assessed by running each sample on a Bioanalyzer and assigning an RNA Quality Score (RQS) to each sample. The 58 samples that presented an RQS above 6 were selected for sequencing. cDNA libraries were prepared using the Nexterra kit (Illumina, cat. #FC-131-1024). The libraries were then sequenced on an Illumina NextSeq sequencer at a targeted 35 million single-end reads with 75 base pairs (bp) per sample. Library preparation and high-throughput sequencing were performed at the EMBL Gene Core facility (Heidelberg, Germany).

3.4 RNAseq data analysis

The quality of raw sequencing data and the presence of adaptors were initially assessed using *FastQC*. Low-quality reads with less than 36 nucleotides long, adapters, leading bases with quality below 3, and trailing bases when the average quality per base is below 15 for regions of 4 bases were then removed using *Trimmomatic* v0.39 (Bolger et al., 2014). HISAT2 v2.2.1 was then used to align the remaining reads to the Homo sapiens genome Ensembl GRCh38 (release 84) followed by the usage of *StringTie* V2.1.6 to obtain gene-level counts (Kim et al., 2019; Pertea et al., 2015). The obtained counting table was imported into R v4.1 for further analysis. Raw counts were processed using *edgeR* v3.36.0 and *limma* v3.50.0 packages (Ritchie et al., 2015; Robinson et al., 2010). Initial processing of raw counts was performed with *edgeR* and genes with counts per million (CPM) lower than 10 in at least 12 samples were discarded from further analysis. Differential expression analysis on the remaining genes was performed using generalized linear models from the *limma* package, accounting for the effect of the sex covariate by adding it as a variable to the model. Genes were deemed significant if log2 fold change was >1 or <-1 and the false discovery rate (FDR) p-value < 0.05 after Benjamini–Hochberg correction for multiple testing. Pathway over-representation analysis (ORA) and gene set enrichment analysis (GSEA) was performed using the *ReactomePA* package (Yu and He, 2016). Activation of the previously described modules or the identified relevant pathways for each sample was computed using the *QuSAGE* package (Meng et al., 2019). Cell type composition was deconvoluted using the *CYBERSORTx* software (Steen et al., 2020) and differences for each cell type between the four groups were tested by performing Kruskal-Wallis tested followed by a Wilcoxon post-hoc test for the cell types that presented significant results. Plots were obtained using *ggplot2*, *pheatmap*, *ggcorrplot*, and *ReactomePA* packages (Cai, 2022; Kolde, 2019; Wickham et al., 2016; Yu and He, 2016). Whole-blood RNAseq data from the Berry South-Africa cohort (Singhania et al., 2018) was downloaded from SRA and the fastq files were processed using the aforementioned analysis pipeline. For the previously published dataset generated in the lab of *in vitro* infected mouse bone marrow-derived macrophages (BMDMs) (Sousa et al., 2020), count matrices obtained were imported into R and processed as described above.

3.5 BMDMs *in vitro* differentiation and infection

BMDMs differentiation *in vitro* was performed as previously described (Moreira-Teixeira et al., 2016). Briefly, mice tibias and femurs were collected, and bone marrow flushed. Cells were counted and $4x10^6$ cells were plated in Petri dishes in 8ml of complete DMEM with 20% L-cell conditioned medium (V/V). On day 4, fresh media was added to the cells. On day 7, cells were harvested, counted, and 500 000 cells were plated per well in 24-well plates. Before infection, mycobacterial clumps were disaggregated by gentle passaging through a 25G needle. BMDMs were infected with either L4 or L6 clinical strains at a multiplicity of infection (MOI) of 2. in the presence or absence of the PKN1 inhibitor staurosporine (cat #S4400, Merck Life Science S.L.U.). Twenty-four hours post-infection, culture supernatants were filter sterilized and stored for later protein analysis.

3.6 Cytokine detection

TNFα and IL10 were detected in supernatants of *in vitro* infected cultures by ELISA (cat. #BMS607 and #BMS614, respectively; both Invitrogen) following manufacturer instructions.

4 Results

4.1 Characterization of the study cohort

The West African country Guinea-Bissau has been reported to have a high incidence of TB cases caused by *M. africanum*, specifically L6 (Groenheit et al., 2011). To gain insights into the current prevalence of TB cases due to infection with *M. africanum* L6 strains, a cohort of 372 individuals with active TB was gathered. Individuals were recruited from different areas in the region of Bissau: within the Bandim Health Project (BHP) area, which includes the urban area of Bissau, or from outside the BHP area, which can be considered a rural area. Lineage determination of the *Mycobacterium* strain causing disease was then performed by SNP typing (Figure 3A). As expected, the incidence of L6 infections was high with 139 out of the 347 genotyped TB cases being due to strains belonging to this lineage (Figure 3B). However, the percentage of L6 infections observed was lower than the one reported in the last studies, which is in line with the decrease in infections caused by *M. africanum* L6 and increase in infections caused by *M. tuberculosis* L4 strains previously reported (Groenheit et al., 2011). Indeed, infections by *M. tuberculosis* L4 strains accounted for 44% of the total genotyped TB cases, being the lineage presenting the highest incidence. A small percentage of individuals (4%) were coinfected with L4 and L6 strains, an occurrence that has not been reported before. A small proportion of TB cases due to infection with L1, L2, and L3

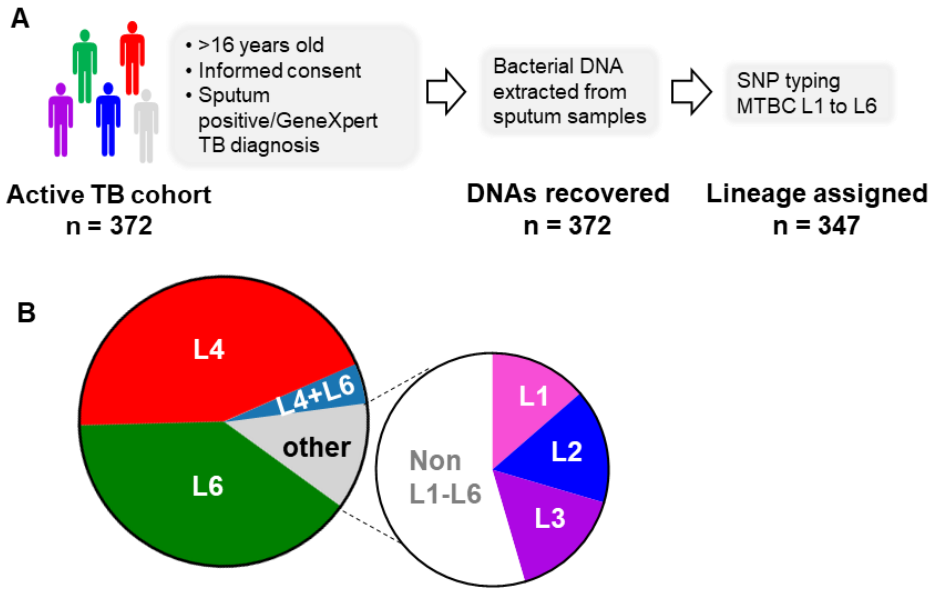


Figure 3. Characterization of the Guinea-Bissau cohort.
(A) Schematics of the recruitment of individuals with active TB followed by the procedure for MTBC lineage identification. (B) Pie charts with the proportion of TB cases caused by the different MTBC lineages.
Kindly provided by M. L. Silva and B. Cá.

strains was also identified. There was also a small percentage of patients infected with non-*M. tuberculosis* and non-*M. africanum* mycobacteria.

Altogether, these observations are in line with previous studies reporting that, even though there is a high burden of TB cases caused by L6 strains in Bissau, there was a decrease in the percentage since the last published report. These results further support the idea that the number of cases due to infection with L6 strains is decreasing over time. Furthermore, it had not been reported before the existence of TB cases where an individual was coinfected with strains from L4 and L6. From this study, however, it is not possible to determine whether the individual got infected with both strains at the same time or on separate occasions. Further studies are necessary to uncover how these double infections are established.

4.2 Whole blood RNAseq reveals a higher number of differentially expressed genes in individuals infected with L4 vs L6 strains

Since most of the TB cases were due to infections with *M. tuberculosis* L4 and *M. africanum* L6, whole-blood RNAseq was performed on blood samples collected at disease diagnosis from 45 individuals infected with strains belonging to L4, L6, or coinfected with strains of L4 and L6 (L4+L6). (Figure 4A). As a control, blood samples from 13 healthy individuals (from here on referred to as healthy donors - HD) were also sequenced.

Upon performing the alignment of the sequencing reads against the human genome reference, the quality of the alignment was confirmed. As shown in Figure 4B, all samples had a minimum of 26 x 10^6 reads, with the vast majority of which aligning to only one site of the genome. Next, principal component analysis (PCA) was performed to verify whether some batch effect was observed between samples or between runs (as the samples were run in 4 different sequencing runs). Analysis of the eight first principal components (PCs) did not indicate the existence of any batch effect (Figure 4C). However, there was a clear separation between samples of healthy individuals and individuals with active TB by the first PC that explained 25 % of the total variability, while the clusters of all infection groups were superimposed (Figure 4C).

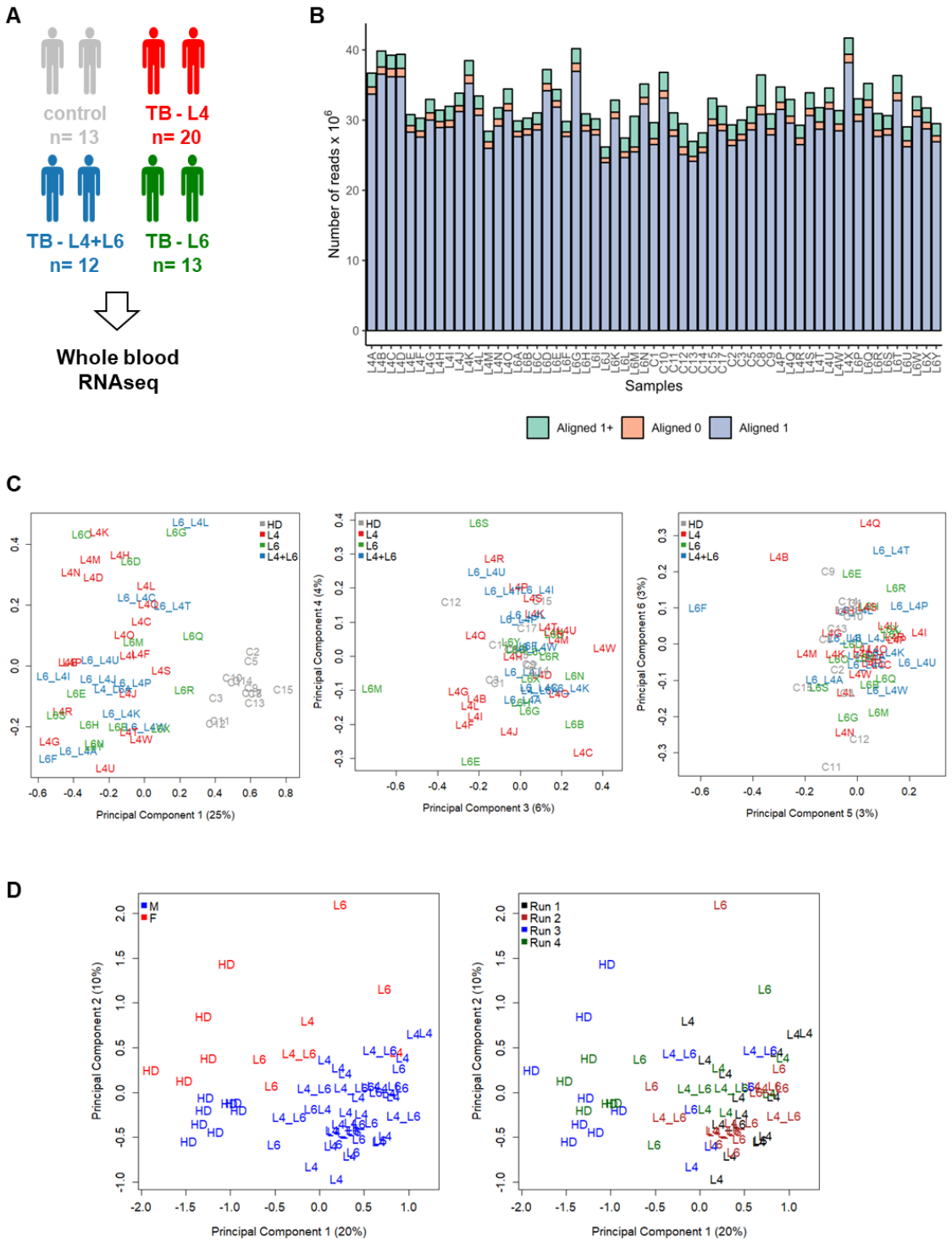


Figure 4. Overview of the whole blood RNAseq results.
(A) Schematics of the number of samples sequenced per group of infection (L4, L6, and L4+L6) or control. (B) Histogram showing the number of reads that did not aligned (Aligned 0) aligned exactly once (Aligned 1) or that aligned more than once (Aligned 1+) to the human reference genome for each sample. (C) Projection of the first six PC from the PCA analysis on the 58 samples. No separation is observed for the different groups except for the PC1 that clearly separates the HD from the other groups. (D) Projection of the PC 1 and 2, from a PCA analysis using the 500 genes presenting the largest standard deviations across the samples, reveals the existence of an effect for the sex (left) covariate but not for the sequencing run (right).

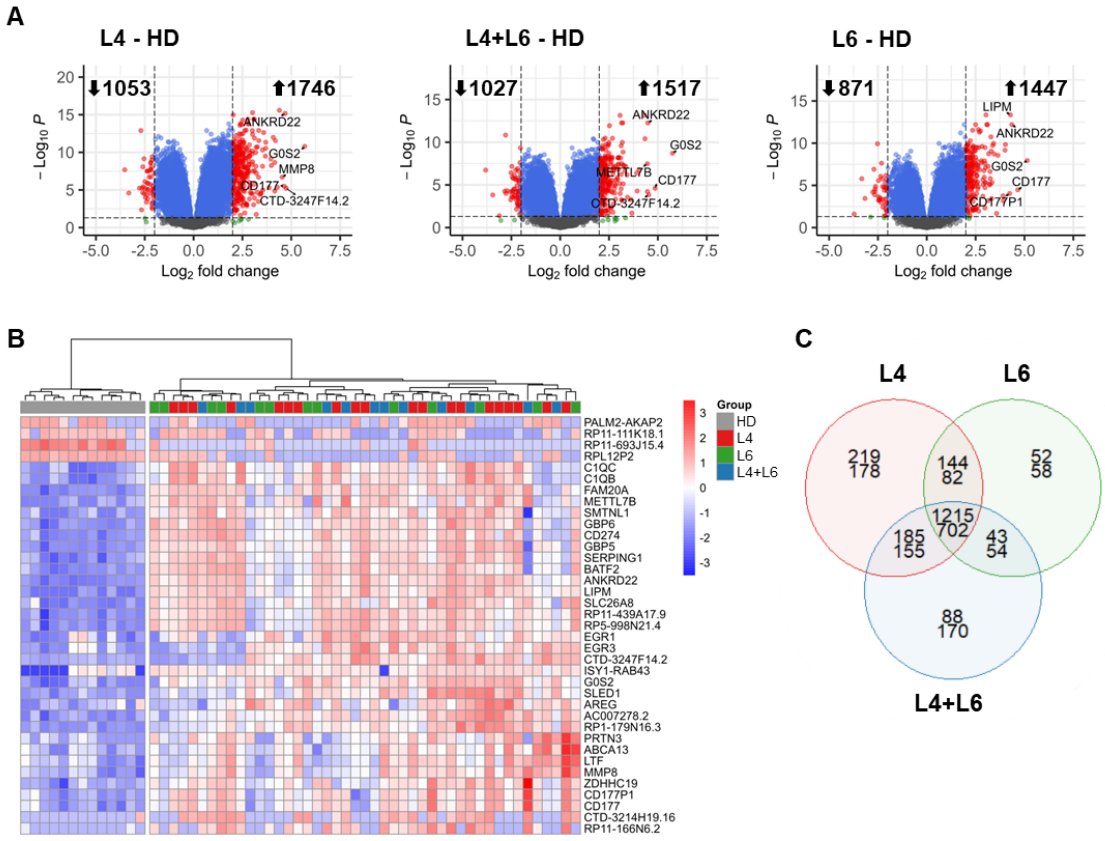


Figure 5 Comparison between the transcriptional profile presented by the three groups of infection.
(A) Volcano plots showing the distribution of genes by Log2 fold change and -log10 adjusted p-value genes for each infection group vs the control group and the corresponding number of downregulated (next to down arrows) and upregulated (next to up arrows) for each infection group when compared with the healthy controls. (B) Heatmap showing the top 25 protein coding DEGs for each infection group when compared to the control group. (C) Venn diagram showing the number of shared DEGs between the three infection groups. Top and bottom numbers correspond to the number of shared or non-shared upregulated and downregulated genes, respectively.

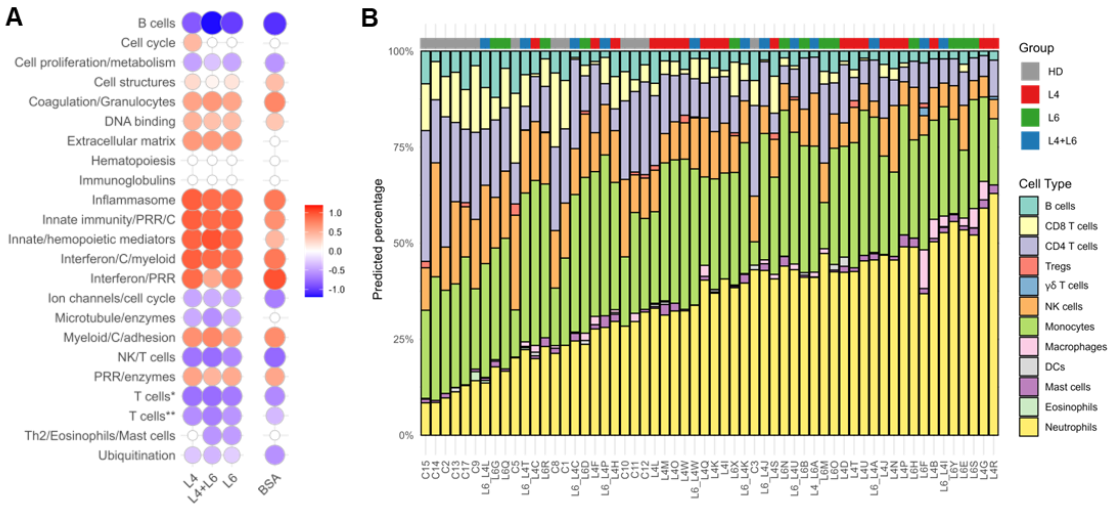
This observation was not surprising since profound transcriptomic alterations have been described in the whole blood of TB patients, which segregate these individuals from healthy controls. In addition, no major alterations between TB caused *M. tuberculosis* or *M. africanum* have been reported. Of note, the first two PCs together explained 33 % of the total variability existing between samples and groups, thus, not accounting for most of the existing differences between samples/groups. The following four PCs (PC3-PC6) only described 16 % of the remaining variability which indicates that most of the variability observed in the dataset originates from sample-specific variability. When performing PCA on the 58 samples using the 500 genes presenting the largest standard deviations across the samples, it was possible to observe that, while PC1 separated once again the samples of healthy individuals and individuals with active TB, PC2 segregated samples based on the individuals sex (Figure 4D). This observation revealed the need to account for the individuals sex in the downstream modelling used for determining differentially expressed genes (DEGs). No batch effect was observable for the sequencing run (Figure

4D). It was not possible to test a potential effect of the covariate age as the age range for the individuals that participated in this study was not available.

Analysis of DEGs, correcting for the effect of the sex covariate, revealed an overall altered transcriptomic profile in all groups (L4, L6, and L4+L6) when compared with the control group (Figure 5A). In addition, analysis of the top protein encoding DEGs revealed that most of these genes were immune system activation genes (Figure 5B). Accordingly, most of the upregulated and downregulated genes were shared between infections caused by L4, L6, and L4+L6 co-infections (Figure 5C). Interestingly, we observed that the number of non-common upregulated genes was higher (approx. four times) in the L4 group than in the L6 group, which may suggest a stronger immune response occurring in *M. tuberculosis* L4-infected individuals than in *M. africanum* L6 counterparts.

To verify if the transcriptional profile obtained in this cohort was in line with the profile observed in previous studies, the activation of the 23 previously described modules was measured in the three infection groups (Singhania et al., 2018). The same analysis was performed on a previously published dataset from a South Africa (Berry South Africa, BSA) TB cohort (Singhania et al., 2018). Of note, for the BSA cohort, whole-blood samples from LTBI, and not samples from healthy individuals, were used as controls. As shown in Figure 6A, the pattern of upregulation and downregulation observed in all groups and the external cohort was largely overlapping, with only four exceptions. The module “Cell cycle” was upregulated in L4 but not in L6 or L4+L6 groups; however, this module was also not upregulated in the group of samples from the BSA, which we would expect to have a more similar profile to the L4 group. Interestingly, the “Th2/Eosinophils/Mast cells” module was downregulated for the L4+L6 and the L6 groups but not L4 and BSA groups, indicating that downregulation of this module may be specific to infections/co-infections with *M. africanum* L6 strains. Modules “Extracellular matrix” and “Microtubule/enzymes” were altered in all groups from the Guinea-Bissau cohort but not for the BSA. The differences observed for these two modules may be due to the different type of samples used as control.

Lastly, we performed deconvolution of our data to identify the percentages of different immune cell types present in the blood (Figure 6B). Most of the individuals with active TB presented neutrophilia in the blood with neutrophils accounting for 30-60% of the total blood immune cells, whereas in healthy individuals the predicted percentage of neutrophils was between 8 and 30%. This was expected, taking into consideration that



C

	Kruskal-Wallis Test	Wilcoxon post-hoc test					
		HD vs L4	HD vs L4+L6	HD vs L6	L4 vs L4+L6	L4 vs L6	L4+L6 vs L6
B cells	** (0.0019)	** (0.0047)	** (0.0047)	* (0.0057)	(0.2543)	(0.5033)	(0.5033)
CD8 T cells	*** (8.7×10^{-05})	*** (0.0001)	** (0.0020)	** (0.0012)	(0.7610)	(0.7610)	(0.7610)
CD4 T cells	*** (4.0×10^{-05})	*** (4.1×10^{-06})	*** (0.0002)	*** (0.0002)	(0.7737)	(0.7737)	(0.7737)
Tregs	0.8498						
γδ T cells	0.5861						
NK cells	* (0.0180)	** (0.0084)	(0.1194)	(0.0573)	(0.7909)	(0.6506)	(0.8100)
Monocytes	** (0.0016)	* (0.0270)	*** (0.0002)	(0.0512)	(0.1284)	(0.8417)	(0.1284)
Macrophages	(0.279)						
DCs	(0.441)						
Mast cells	(0.0582)						
Eosinophils	(0.1728)						
Neutrophils	*** (6.5×10^{-04})	*** (0.0001)	** (0.0074)	** (0.0057)	(0.4512)	(0.9566)	(0.4512)

Figure 6. Overall transcriptional and cellular profile of the different groups of infection.
(A) Activation of the 23 previously described modules for the three infection groups of the Guinea-Bissau cohort (left) and the active TB group of the public available Berry South Africa Cohort (BSA). (B). Predicted blood cell type composition for each sample. (C) Results of the Kruskal-Wallis statistical test on differences between the four groups for each blood cell type, followed by Wilcoxon post-hoc test for the cell types that presented significant p-value further confirm the different blood composition between healthy individuals and individuals with TB. * p-value < 0.05, ** p-value < 0.01, and *** p-value < 0.01.

blood neutrophilia is characteristic of patients with active TB (Lowe et al., 2013; Lowe et al., 2012). Further statistical testing between the four groups for each cell type not only confirmed the difference in neutrophils between individuals with TB and the healthy counterparts but also revealed statistical differences in B cells, CD8 and CD4 T cells, and NK cells, monocytes between infected and non-infected individuals.

In summary, all three infection groups presented a whole blood transcriptome profile compatible with the ongoing immune response to infection and in line with the transcriptional profile previously described in TB. Nevertheless, the results indicate the response observed in the L4 group is potentially stronger than the one observed for the other two groups.

4.3 Differences in nuclear receptor pathway transcriptional activation in *M. tuberculosis* L4- vs *M. africanum* L6-infected individuals

To further dissect the differences in the host immune response against L4 vs L6 strains, we performed two types of pathway analysis in the DEGs of L4 and L6 groups: ORA, which determines if a list of genes contains genes belonging to sets/pathways more represented than expected by chance; and GSEA, which determines if a list of genes and their corresponding differential expression is in line with predefined gene expression changes (pre-determined by comparison of two conditions) in predefined gene sets or pathways (Figure 7). Most of the top pathways (8/10) revealed by ORA analysis (Figure 7A) were pathways involved in immune responses common between

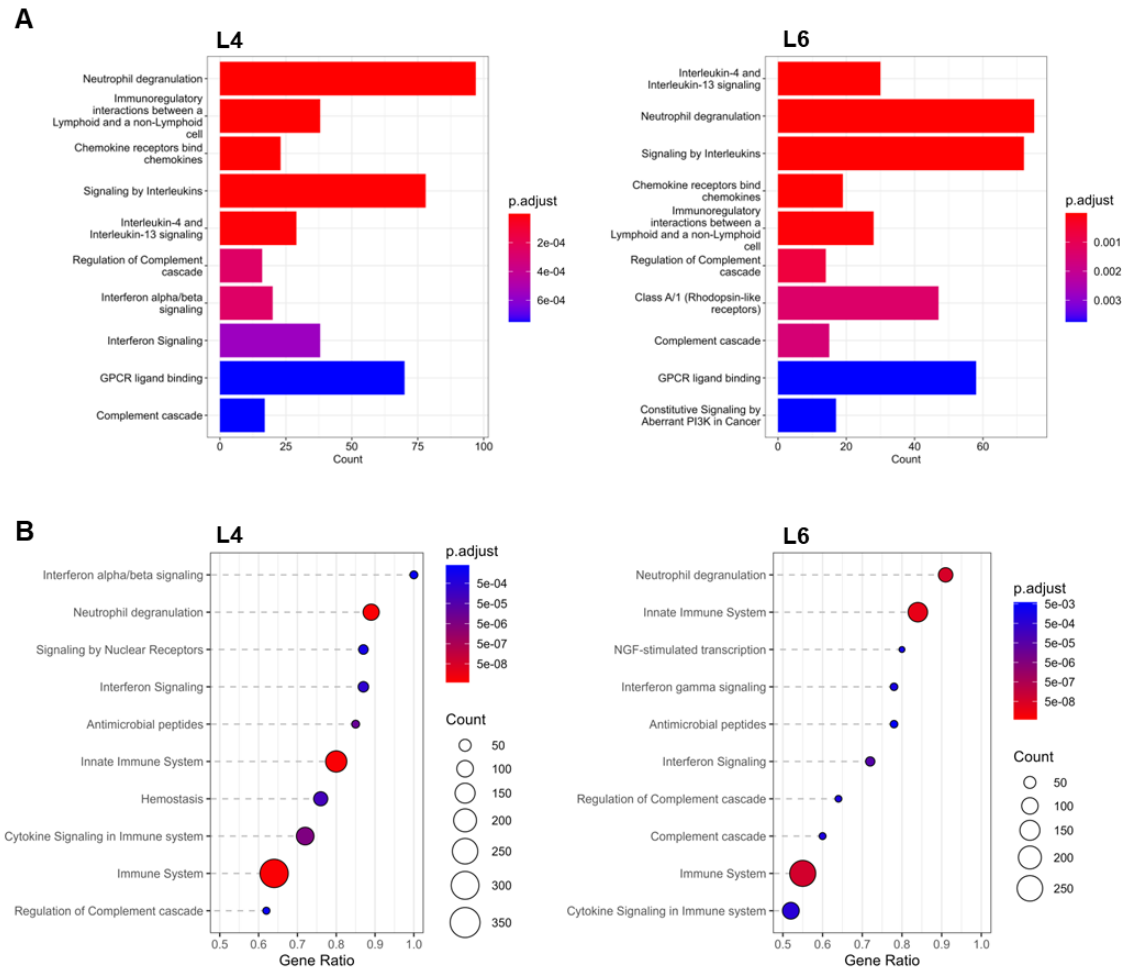


Figure 7. Pathway analysis in L4 vs L6 groups.
The top 10 altered pathways for the L4 and L6 infection groups (left and right, respectively) were obtained by ORA (A) and GSEA (B) on the DEGs of each group. Both upregulated and downregulated genes were used in the analysis giving a total of 2799 and 2318 genes for the L4 and L6 groups, respectively used for the analysis.

Figure 8. The number of DEGs per pathway is higher in L4 than in L6 groups.
Linkage plots of DEGs associated with the pathways (top to bottom) (1) Neutrophil degranulation, (2) Cytokine signaling, IFNα/β signaling, and IFNγ signaling, (3) TLR cascades and signaling by nuclear receptors, and (4) Hemostasis for L4 (left) and L6 (right) groups.

the two groups. However, the number of genes identified per pathway was higher for the L4 group and the range of adjusted p-values was twenty times lower for the L4 group vs the L6 group. The results obtained from the GSEA analysis (Figure 7B) showed a similar trend with most of the top pathways being common between groups (and common with the ORA results). Furthermore, in the GSEA the number of genes was also higher while the range adjusted p-values were lower for L4 vs L6, respectively. These results indicate that while both L4 and L6 groups presented the same altered pathways, the degree of activation was different between the two groups.

A closer look into the DEGs associated with selected identified pathways, including neutrophil degranulation, IFNα/β signaling, and IFNγ signaling which are known to play key roles in TB, showed once more that the web of genes was higher for the L4 than for the L6 group (Figure 8). This observation is in line with the initial results that, overall, the group of samples from patients with TB resulting from *M. tuberculosis* L4 infections presented more upregulated and downregulated genes than those infected with *M. africanum* L6 (Figure 5C). One exception, to the commonly altered gene pathways, was the signaling by nuclear receptors pathway. This pathway was only detected as transcriptionally altered for the L4 group and, interestingly, with more than 85% of the pathway genes altered (Figure 7B and Figure 8). The same was not true for the hemostasis pathway which was altered for both groups (Figure 8), only to a less extent for the L6 group (only the top 10 altered pathways are presented the L6 panel of Figure 7B and hemostasis pathway is the 18th altered pathway for the L6 group).

To further understand how these selected pathways were differently activated at transcriptional levels in individuals infected with *M. tuberculosis* L4 or *M. africanum* L6 strains, a closer analysis of their activation at the individual level was performed (Figure 9, Figure 10, and Figure 11). By clustering the samples based on the expression of the DEGs associated with the signaling by the nuclear receptors pathway, we could observe that 5 out of the 13 samples from L6 infected individuals formed a cluster (with only a sample from the L4 group) with low expression levels of these genes (Figure 9A). Five of the samples from the L6 group grouped in a cluster with intermediate expression levels, while only two presented high expression levels of the genes associated with this pathway. This pattern was only observed for this pathway, as for the pathways IFNα/β signaling, IFNγ signaling, and neutrophil degranulation pathways, both groups were scattered over the 3 main clusters observed (Figure 10A,B and Figure 11A). To confirm the different levels of activation of the signaling by the nuclear receptors pathway and the other altered pathways, a pathway activation score (eigengene) was obtained for

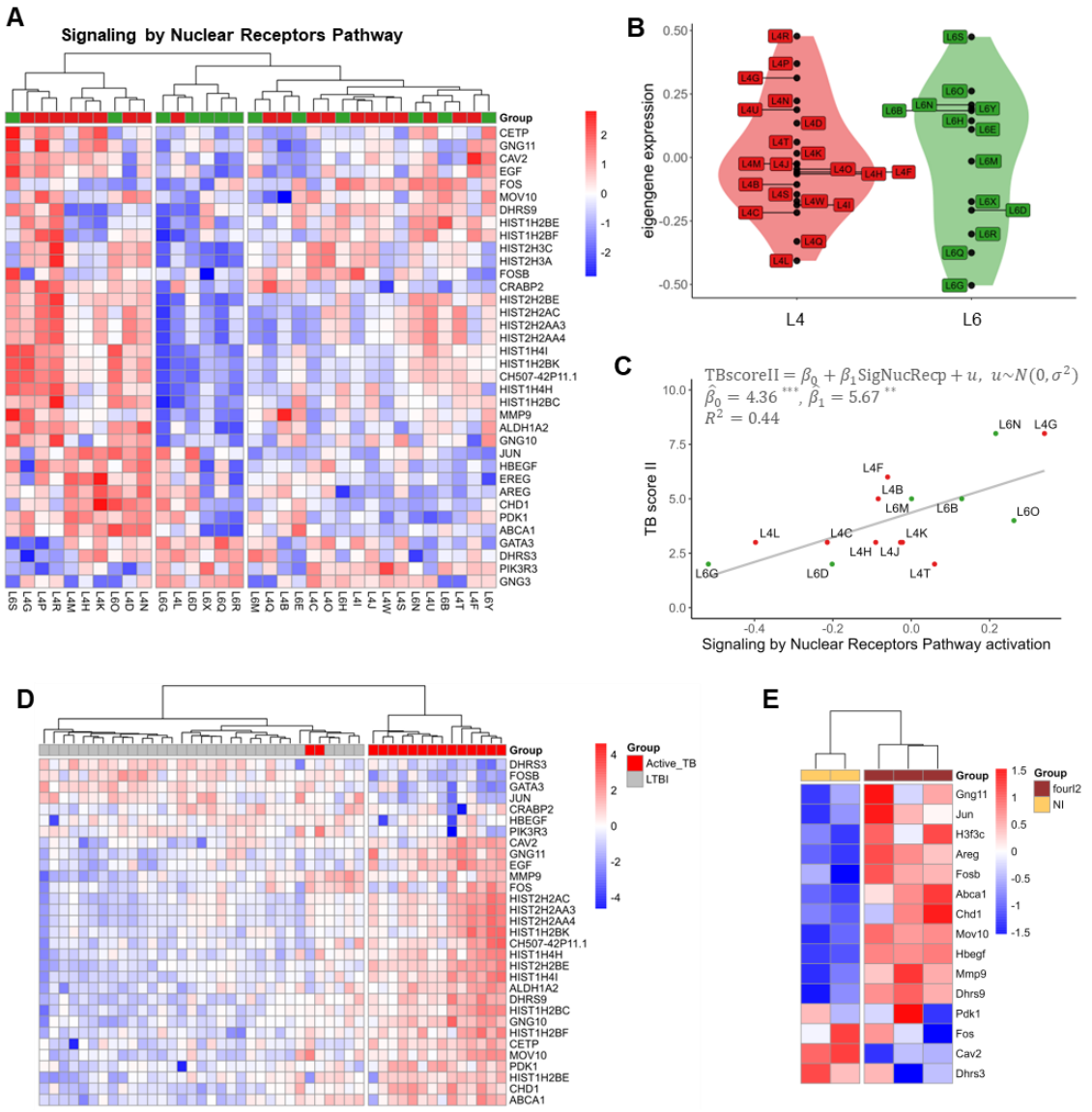


Figure 9. Signalling by nuclear receptors pathway is differently activated in L4 and L6 groups and has a positive relation with disease score.
(A) Heatmap showing the relative expression DEGs that are part of the signaling by nuclear receptors pathway. (B) Eigengene expression of the signaling by nuclear receptors pathway for each sample of the L4 and L6 groups. (C) Regression analysis between TB score and the eigengene expression of the signaling by nuclear receptors pathway. (D) Heatmap showing the relative expression of DEGs (active TB vs LTBI) that are part of the signaling by nuclear receptors pathway for the public available Berry South Africa cohort (Singhania et al., 2018). (E) Heatmap showing the relative expression of DEGs (BMDMs infected vs non-infected *in vitro* with L4 strain) (Sousa et al., 2020).

each pathway for each sample of L4 and L6 groups (Figure 9B, Figure 10C,D, and Figure 11B). In general, we could observe that the L6 group distributions were wider than the ones observed for the L4 group. Indeed, while L4 group distributions resembled skewed distributions (right-skewed for pathways signaling by nuclear receptors and neutrophil degranulation, and left-skewed for IFNα/β signaling and IFNγ signaling), distributions of the L6 group resembled uniform distributions (Figure 9B, Figure 10C,D, and Figure 11B). These results suggest that while the different pathways have a similar pattern of

activation in individuals with active TB caused by L4 bacteria, individuals with an active TB caused by L6 bacteria have a wide range of activation levels of these pathways and consequently different levels of activation of the immune system. Lastly, to verify whether the disease severity observed was dependent on the altered expression of any of the identified pathways, we performed a linear regression analysis (Figure 9C, Figure 10E,F, and Figure 11C). Regression analysis was calculated between the TB score II and the previously calculated eigengenes. Of note, the regression was performed only on 15 samples (9 L4 and 6 L6) as the TB score data was only available for the samples whose individuals lived within BHP. Interestingly, only the signaling by nuclear receptors pathway presented a significant coefficient ($\hat{\beta}_1$ p-value = 0.007) with a fit R^2 of 0.44.

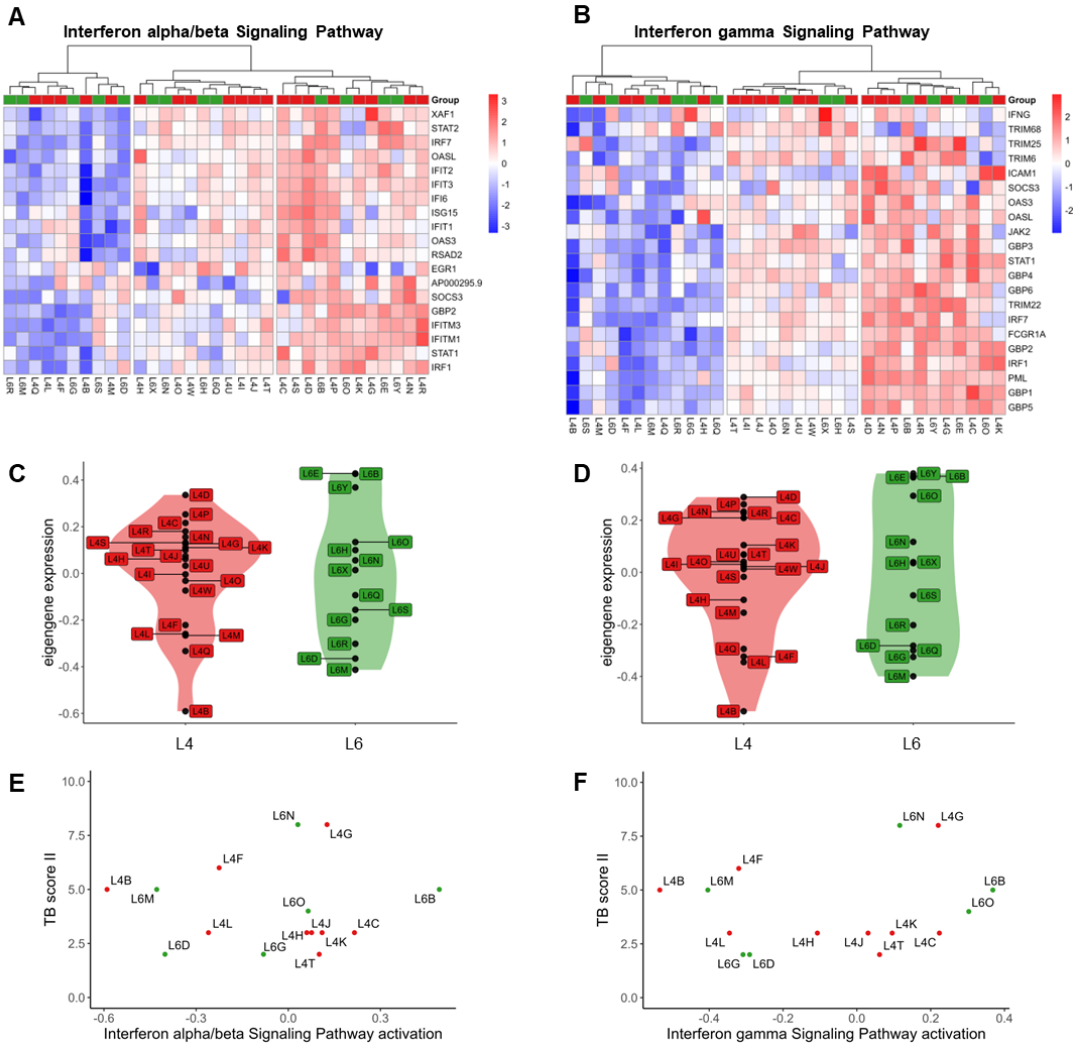


Figure 10. The activation level of IFNα/β and IFNγ pathways for L4 and L6 samples and their relation with TB score. (A/B) Heatmap showing the relative expression of DEGs that are part of the IFNα/β (A) and IFNγ (B) signaling pathways. (C/D) Eigengene expression of the IFNα/β (C) and IFNγ (D) signaling pathways for each sample of the L4 and L6 groups. (E/F) Regression analysis between TB score and the eigengene expression of the IFNα/β (E) and IFNγ (F) signaling pathways.

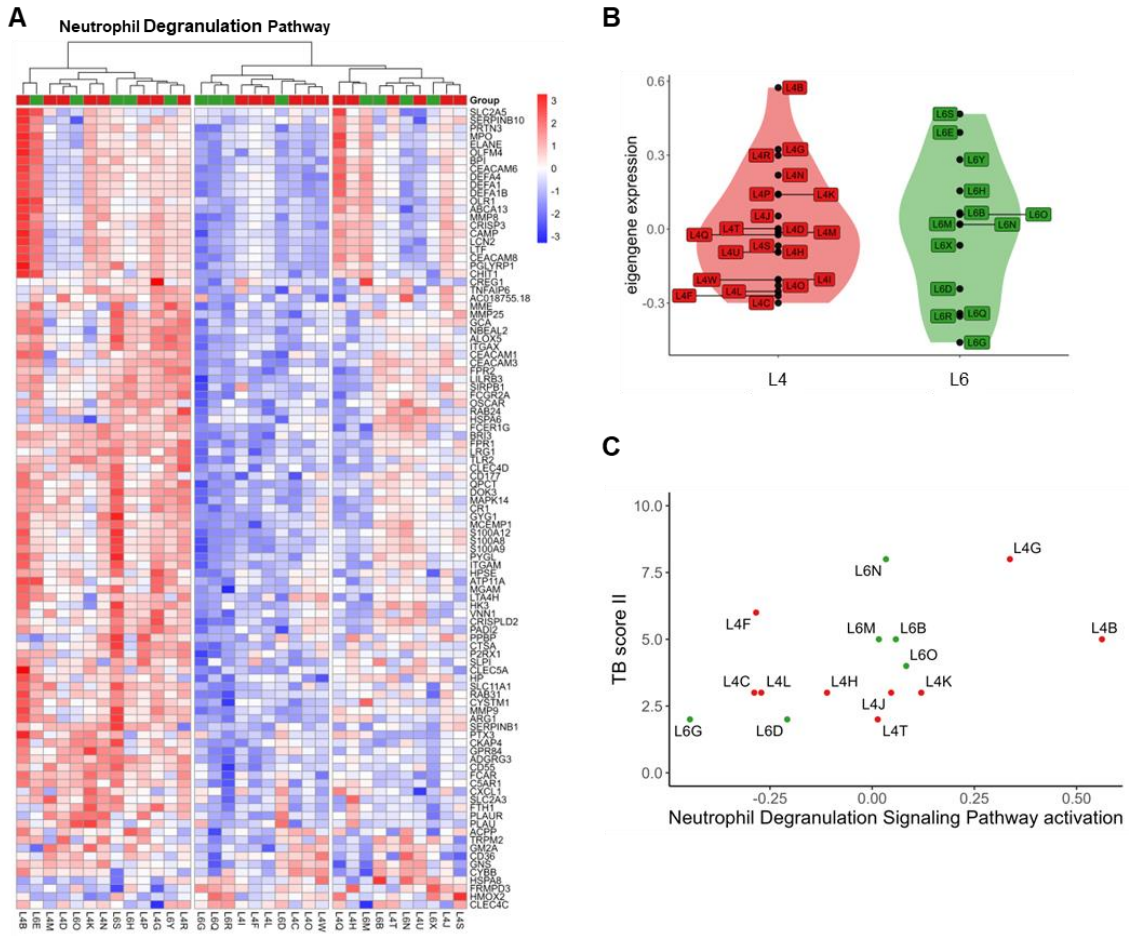


Figure 11. The activation level of neutrophil degranulation pathway for L4 and L6 samples and relation with TB score. (A) Heatmap showing the relative expression DEGs that are part of neutrophils degranulation pathway. (B) Eigengene expression of the neutrophil degranulation pathway for each sample of the L4 and L6 groups. (E/F) Regression analysis between TB score and the eigengene expression of neutrophil degranulation pathway.

To investigate whether the activation of the signaling by nuclear receptors pathways in *M. tuberculosis* infections observed was not exclusive to the Guinea-Bissau cohort, we interrogated the whole-blood RNAseq samples from the BSA cohort (Singhania et al., 2018). In this geographical location, L4 is the lineage with higher incidence (approx. 70% of the cases), and no cases of *M. africanum* infection are expected (Chihota et al., 2018) In agreement with our data, 14 out of the 16 samples from individuals with active TB presented an overall higher expression of the genes associated with this pathway than the LTBI samples (Figure 9D). Indeed, the two groups formed separate and distinct clusters with exception of two active TB samples. Lastly, to verify if this upregulation of genes associated with the signaling by nuclear receptors was transversal to species and could also be observed in mice, we resorted to a dataset previously published by our lab (Sousa et al., 2020). The dataset consisted of BMDMs infected with an L4 strain *in vitro* and non-infected BMDMs as controls. As observed for the whole-blood human samples,

BMDMs infected with an L4 strain presented an upregulation of some of the corresponding mouse genes of the signaling by nuclear receptors pathway.

Altogether, these results revealed that most of the altered pathways are common between L4 and L6 infection. However, while most samples of the L4 group present a similar degree of activation, L6 samples present a wide range of activation statuses. This difference was more striking for the signaling by nuclear receptors pathway, which expression presented a positive relationship with the disease severity observed.

4.4 Differential impact of PKN1 abrogation in macrophages infected with L4 or L6 isolates

Most immune cells possess special pattern recognition receptors (PRRs) that recognize pathogens, more specifically, pathogen-associated molecular patterns (PAMPs) (Li and Wu, 2021). One family of these PRRs is called toll-like receptors (TLRs). TLRs can be divided into two groups depending on their cellular localization: TLRs 1, 2, 4, and 5 localize on the cell surface and recognize PAMPs present in the extracellular space, and TLR3, 7, 8, and 9 are mainly found in endosomal vesicles and are activated when PAMPs are present within this cellular structure (Gay and Gangloff, 2007; Li and Wu, 2021). The role of TLRs has been extensively studied in *M. tuberculosis* infection (Faridgohar and Nikoueinejad, 2017; Jo et al., 2007; Natarajan et al., 2011). The main TLRs involved in the recognition of *M. tuberculosis* PAMPS and consequent activation of innate immune cells are TLRs 2, 4, 8, and 9 (Faridgohar and Nikoueinejad, 2017). The binding of *M. tuberculosis* PAMPs to these TLRs activates several signaling cascades that ultimately lead to the production of cytokines, chemokines, and co-stimulatory molecules (Figure 12A). PDK1, which gene was part of the nuclear receptor pathway, is one of the kinases downstream TLRs (Figure 12A). This information led to the hypothesis that the differences in activation of the immune response observed between L4 and L6 strains could be due, at least in part, to a differential activation of the TLR/MyD88/PDK1 signaling pathway. To investigate this, BMDMs were infected *in vitro* with an *M. tuberculosis* L4 or an *M. africanum* L6 clinical isolate, both recovered from TB patients from Bissau. The amount of TNFα produced was used as an indirect measure of the signaling cascade activation (Figure 12B). In agreement with the proposed hypothesis, infection with an *M. africanum* L6 strain led to lower production of TNFα than infection with *M. tuberculosis* L4 strain (Figure 12B). Moreover, infection in the presence of

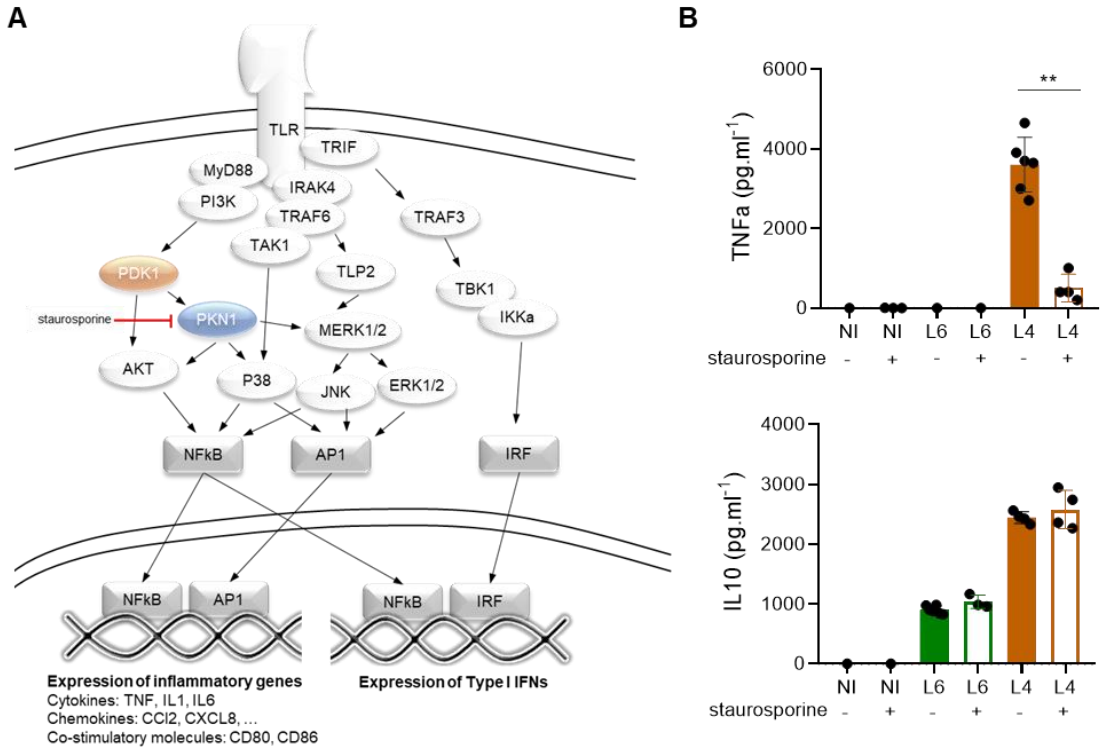


Figure 12. PDK1/PKN1 pathway is activated in BMDMs infected with L4 but not L6.
(A) Schematics of the TLR signaling pathways, including the PI3K/PDK1/PKN1 pathway, which leads to the expression of inflammatory genes and type I IFNs. (B) TNFα and IL10 production by BMDMs 24 h after infection with an L4 or an L6 strain, and in the presence/absence of the PKN1 inhibitor staurosporine.
Panel B kindly provided by M. L. Silva.

staurosporine, a PKN1 (kinase downstream of PDK1 in the signaling cascade; Figure 12A) inhibitor, resulted in a reduction in TNFα produced after infection with an L4 strain, further confirming the involvement of this signaling cascade in the BMDMs activation and consequent cytokine production. Lastly, even though the amount of IL10 produced was higher after infection with the *M. tuberculosis* L4 than the *M. africanum* L6 strain, the presence of the inhibitor did not affect the amount of cytokine produced, as the cascades leading to IL10 expression were not affected.

Hence, these data suggest that the *M. africanum* L6 strain does not seem to activate TLR/PDK1/PKN1 signaling cascades to the same degree as the *M. tuberculosis* L4 strain. These results are in line with the overall lower activation of the immune response observed in individuals infected with *M. africanum* L6 vs *M. tuberculosis* L4 strains on the RNAseq data, providing an initial functional validation.

5 Discussion

TB is still one of the infectious diseases with a higher mortality burden worldwide. To face this disease, it is necessary to develop new preventive, diagnostic and treatment approaches which, in turn, requires a better knowledge of TB-causing bacteria and the immune response triggered in the host.

TB is caused by a wide range of bacterial strains that have been divided into 9 lineages (Brites and Gagneux, 2017; Napier et al., 2020). Bacteria belonging to different lineages have been described to trigger different immune responses and associate with TB of different severities (Bastos et al., 2017). Thus, unveiling such differences in the immune response may allow a better understanding of the disease and reveal potential therapeutic targets to facilitate an effective immune response while preventing tissue damage and destruction. Hence, the objective of this study was to reveal differences in the immune response observed in TB caused by *M. tuberculosis* or *M. africanum*.

As the first step, a new TB cohort was gathered in Guinea-Bissau, a country with a high burden of TB and in which many of the cases are due to infection with *M. africanum* L6 strains. Upon determining the prevalence of the main 6 MTBC lineages (L1-L6), it was observed that *M. tuberculosis* L4 was the lineage with the highest incidence, followed by *M. africanum* L6. Thus, whole-blood RNAseq from individuals infected with *M. tuberculosis* L4 and *M. africanum* L6 strains was performed. In addition, whole-blood RNAseq was also performed in blood samples from individuals coinfected with *M. tuberculosis* L4 and *M. africanum* L6 strains, an occurrence that had not been previously reported, and from healthy individuals who were used as controls in this study.

Initial analysis of the RNAseq results revealed a profound alteration of the blood transcriptomic profile, with many immune-related genes being upregulated in the three groups of infection. Indeed, the main induced alterations were common between the three groups and the profile observed was in line with transcriptomic profiles previously published (Singhania et al., 2018). Interestingly, there were a considerable number of upregulated or downregulated genes that were specific to the L4, L6 and L4+L6 groups. Furthermore, the L4 group was the group presenting the highest number of upregulated genes, which strongly indicates that the immune response was stronger for this group when compared with the other two groups. Thus, while an overall similar immune response was observed for the three groups, specific differences in the immune response mounted were in place. Importantly, although it was possible to determine that some individuals were coinfected with (at least) two different strains belonging to L4 and

L6, it is not possible to establish which infection occurred first, neither the impact of the co-infection to the disease trajectory. This fact prevents further dissecting the L4+L6 group, as similarities or differences observed may be due to the relative burden of *M. tuberculosis* L4 and *M. africanum* L6 strains. Thus, further studies are necessary to reveal how the bacteria of each lineage may impact the infection of their counterparts.

The results obtained in this study are in line with 23 previously described modules activation levels (Berry et al., 2010). An interesting observation was that the "Th2/Eosinophils/Mast cells" module was downregulated for the L6 and L4+L6 groups, whilst not altered in the L4 group or for the BSA cohort. While this observation rises the hypothesis that the immune response mounted in TB cases due to infection with *M. africanum* L6 strains may associate with this downregulation, this pattern was also observed in published results for the Leicester cohort, where cases due to infection *M. africanum* L6 strains are not expected (Singhania et al., 2018). Moreover, it was not possible to explore this hypothesis since the predicted percentages for eosinophils and mast cells were very low or null for most samples and the potential Th2 cells that may exist in the blood of these individuals with TB are very much obscured by the Th1 nature of the ongoing immune response. Thus, resorting to scRNAseq of similar techniques would be a better strategy to (1) characterize the heterogeneous cell populations present in the blood, (2) uncover specific transcriptomic profiles of blood cells and, (3) establish similarities and predominant clones of T and B cells by sequencing and analysing their TCR and BCR repertoire, respectively. Of note, it was not possible to analyse the TCR/BCR repertoire from the sequencing data generated in this study due to the short length of the sequencing reads.

Pathway analysis was then performed on the DEGs of the L4 and L6 groups. While most of the altered pathways were associated with the expected immune response in place and common between the groups, the pathway signaling by nuclear receptors was only statistically altered for the L4 group. Closer analysis of this and other pathways revealed that the levels of pathway alteration for the samples from the L6 group covered a wider distribution, while for the individuals from the L4 group the values presented a less dispersed distribution. From these observations rises the hypothesis that while TB caused by *M. tuberculosis* L4 strains may progress faster than TB caused by *M. africanum* L6 strains (Baya et al., 2020; Bold et al., 2012; Ca et al., 2019; de Jong et al., 2008; Wiens and Ernst, 2016b), most individuals will progress to a 'stopping' point where the immune response observed is similar; however, this 'stopping' point does not seem to exist for TB caused *M. africanum* L6 strains. Moreover, it may be the case that

although the initial disease progression is slower in individuals infected with *M. africanum* L6 strains, the progression to worse disease states may be faster than the one observed in individuals infected with *M. tuberculosis* L4 strains.

One of the DEGs belonging to the signaling by nuclear receptors pathway was the PDK1 gene. The kinase encoded by this gene takes part in the TLR/MyD88 signaling cascade that leads to the activation of innate immune cells and consequent transcription of inflammatory genes and type I IFNs. By performing *in vitro* experiments, it was possible to confirm that while infection of BMDMs with an L4 clinical isolate led to the transcription of such inflammatory genes (measured by the amount of TNFα production), infection with an L6 clinical isolate did not lead to the production of TNFα. The involvement of the TLR/MyD88/PDK1 signaling pathway was further confirmed by the reduction of TNFα production when an inhibitor of a downstream kinase was added to the culture. Interestingly, these observations may also explain the lower levels of IFNβ produced *in vitro* after infection with *M. africanum* L6 strains, when compared with the amount produces after infection with *M. tuberculosis* L2 and L4 strains, since this signaling cascade is also involved in the transcription of type I IFN genes (Wiens and Ernst, 2016a). Nevertheless, further experiments are necessary to confirm if this pathway is indeed responsible for the differential production of type I IFN genes by cells infected with *M. tuberculosis* L2/L4 or *M. africanum* L6 strains.

Altogether, the results here presented provide new insights into differences in the immune response observed in individuals infected with *M. tuberculosis* L4 strains vs. *M. africanum* L6 strains that may be used in the future as targets for new TB therapies.

6 References

Acquah, S.K.E., P. Asare, S. Osei-Wusu, P. Morgan, T. Afum, D.A. Asandem, E.K. Danso, I.D. Otchere, L.A. Ofori, K. Obiri-Danso, R. Kock, A. Asante-Poku, and D. Yeboah-Manu. 2021. Molecular epidemiology and drug susceptibility profiles of Mycobacterium tuberculosis complex isolates from Northern Ghana. *Int J Infect Dis* 109:294-303.

Ahmed, M., S. Thirunavukkarasu, B.A. Rosa, K.A. Thomas, S. Das, J. Rangel-Moreno, L. Lu, S. Mehra, S.K. Mbandi, L.B. Thackray, M.S. Diamond, K.M. Murphy, T. Means, J. Martin, D. Kaushal, T.J. Scriba, M. Mitreva, and S.A. Khader. 2020. Immune correlates of tuberculosis disease and risk translate across species. *Sci Transl Med* 12:eaay0233.

Ainley, A., and O.M. Kon. 2020. Clinical tuberculosis. *Medicine* 48:356-362.

Akter, S., K.S. Chauhan, M.D. Dunlap, J.A. Choreno-Parra, L. Lu, E. Esaulova, J. Zuniga, M.N. Artyomov, D. Kaushal, and S.A. Khader. 2022. Mycobacterium tuberculosis infection drives a type I IFN signature in lung lymphocytes. *Cell Rep* 39:110983.

Anderson, S.T., M. Kaforou, A.J. Brent, V.J. Wright, C.M. Banwell, G. Chagaluka, A.C. Crampin, H.M. Dockrell, N. French, M.S. Hamilton, M.L. Hibberd, F. Kern, P.R. Langford, L. Ling, R. Mlotha, T.H.M. Ottenhoff, S. Pienaar, V. Pillay, J.A.G. Scott, H. Twahir, R.J. Wilkinson, L.J. Coin, R.S. Heyderman, M. Levin, and B. Eley. 2014. Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Engl J Med* 370:1712-1723.

Asare, P., A. Asante-Poku, D.A. Prah, S. Borrell, S. Osei-Wusu, I.D. Otchere, A. Forson, G. Adjapong, K.A. Koram, S. Gagneux, and D. Yeboah-Manu. 2018. Reduced transmission of Mycobacterium africanum compared to Mycobacterium tuberculosis in urban West Africa. *Int J Infect Dis* 73:30-42.

Ates, L.S., A. Dippenaar, F. Sayes, A. Pawlik, C. Bouchier, L. Ma, R.M. Warren, W. Sougakoff, L. Majlessi, J.W.J. van Heijst, F. Brossier, and R. Brosch. 2018. Unexpected Genomic and Phenotypic Diversity of Mycobacterium africanum Lineage 5 Affects Drug Resistance, Protein Secretion, and Immunogenicity. *Genome Biol Evol* 10:1858-1874.

Bastos, H.N., N.S. Osorio, S. Gagneux, I. Comas, and M. Saraiva. 2017. The Troika Host-Pathogen-Extrinsic Factors in Tuberculosis: Modulating Inflammation and Clinical Outcomes. *Front Immunol* 8:1948.

Baya, B., B. Diarra, S. Diabate, B. Kone, D. Goita, Y.D.S. Sarro, K. Cohen, J.L. Holl, C.J. Achenbach, M. Tolofoudie, A.C.G. Togo, M. Sanogo, A. Kone, O. Kodio, D. Dabitao, N.

Coulibaly, S. Siddiqui, S. Diop, W. Bishai, S. Dao, S. Doumbia, R.L. Murphy, S. Diallo, and M. Maiga. 2020. Association of Mycobacterium africanum Infection with Slower Disease Progression Compared with Mycobacterium tuberculosis in Malian Patients with Tuberculosis. *Am J Trop Med Hyg* 102:36-41.

Berry, M.P., C.M. Graham, F.W. McNab, Z. Xu, S.A. Bloch, T. Oni, K.A. Wilkinson, R. Banchereau, J. Skinner, R.J. Wilkinson, C. Quinn, D. Blankenship, R. Dhawan, J.J. Cush, A. Mejias, O. Ramilo, O.M. Kon, V. Pascual, J. Banchereau, D. Chaussabel, and A. O'Garra. 2010. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 466:973-977.

Bloom, C.I., C.M. Graham, M.P. Berry, K.A. Wilkinson, T. Oni, F. Rozakeas, Z. Xu, J. Rossello-Urgell, D. Chaussabel, J. Banchereau, V. Pascual, M. Lipman, R.J. Wilkinson, and A. O'Garra. 2012. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. *PLoS One* 7:e46191.

Bold, T.D., D.C. Davis, K.K. Penberthy, L.M. Cox, J.D. Ernst, and B.C. de Jong. 2012. Impaired fitness of Mycobacterium africanum despite secretion of ESAT-6. *J Infect Dis* 205:984-990.

Bolger, A.M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.

Brites, D., and S. Gagneux. 2017. The Nature and Evolution of Genomic Diversity in the Mycobacterium tuberculosis Complex. *Adv Exp Med Biol* 1019:1-26.

Broset, E., C. Martin, and J. Gonzalo-Asensio. 2015. Evolutionary landscape of the Mycobacterium tuberculosis complex from the viewpoint of PhoPR: implications for virulence regulation and application to vaccine development. *mBio* 6:e01289-01215.

Burel, J.G., A. Singhania, P. Dubelko, J. Muller, R. Tanner, E. Parizotto, M. Dedicoat, T.E. Fletcher, J. Dunbar, A.F. Cunningham, C.S. Lindestam Arlehamn, D.G. Catanzaro, A. Catanzaro, T. Rodwell, H. McShane, M.K. O'Shea, and B. Peters. 2021. Distinct blood transcriptomic signature of treatment in latent tuberculosis infected individuals at risk of developing active disease. *Tuberculosis (Edinb)* 131:102127.

Ca, B., K.L. Fonseca, J. Sousa, A.R. Maceiras, D. Machado, L. Sanca, P. Rabna, P.N.S. Rodrigues, M. Viveiros, and M. Saraiva. 2019. Experimental Evidence for Limited in vivo Virulence of Mycobacterium africanum. *Front Microbiol* 10:2102.

Cai, J.M., G.; Daniël, S.L. 2022. ggcorrplot2: Visualize a Correlation Matrix using ggplot2. In https://github.com/caijun/ggcorrplot2. Last access 29 September 2022.

Cai, Y., Q. Yang, Y. Tang, M. Zhang, H. Liu, G. Zhang, Q. Deng, J. Huang, Z. Gao, B. Zhou, C.G. Feng, and X. Chen. 2014. Increased complement C1q level marks active disease in human tuberculosis. *PLoS One* 9:e92340.

Carow, B., T. Hauling, X. Qian, I. Kramnik, M. Nilsson, and M.E. Rottenberg. 2019. Spatial and temporal localization of immune transcripts defines hallmarks and diversity in the tuberculosis granuloma. *Nat Commun* 10:1823.

Chihota, V.N., A. Niehaus, E.M. Streicher, X. Wang, S.L. Sampson, P. Mason, G. Kallenius, S.G. Mfinanga, M. Pillay, M. Klopper, W. Kasongo, M.A. Behr, N.C. Gey van Pittius, P.D. van Helden, D. Couvin, N. Rastogi, and R.M. Warren. 2018. Geospatial distribution of Mycobacterium tuberculosis genotypes in Africa. *PLoS One* 13:e0200632.

Cho, S., V. Mehra, S. Thoma-Uszynski, S. Stenger, N. Serbina, R.J. Mazzaccaro, J.L. Flynn, P.F. Barnes, S. Southwood, E. Celis, B.R. Bloom, R.L. Modlin, and A. Sette. 2000. Antimicrobial activity of MHC class I-restricted CD8+ T cells in human tuberculosis. *Proc Natl Acad Sci U S A* 97:12210-12215.

Cohen, S.B., B.H. Gern, J.L. Delahaye, K.N. Adams, C.R. Plumlee, J.K. Winkler, D.R. Sherman, M.Y. Gerner, and K.B. Urdahl. 2018. Alveolar Macrophages Provide an Early Mycobacterium tuberculosis Niche and Initiate Dissemination. *Cell Host Microbe* 24:439-446 e434.

Comas, I., M. Coscolla, T. Luo, S. Borrell, K.E. Holt, M. Kato-Maeda, J. Parkhill, B. Malla, S. Berg, G. Thwaites, D. Yeboah-Manu, G. Bothamley, J. Mei, L. Wei, S. Bentley, S.R. Harris, S. Niemann, R. Diel, A. Aseffa, Q. Gao, D. Young, and S. Gagneux. 2013. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nat Genet* 45:1176-1182.

Coscolla, M., R. Copin, J. Sutherland, F. Gehre, B. de Jong, O. Owolabi, G. Mbayo, F. Giardina, J.D. Ernst, and S. Gagneux. 2015. M. tuberculosis T Cell Epitope Analysis Reveals Paucity of Antigenic Variation and Identifies Rare Variable TB Antigens. *Cell Host Microbe* 18:538-548.

Coscolla, M., S. Gagneux, F. Menardo, C. Loiseau, P. Ruiz-Rodriguez, S. Borrell, I.D. Otchere, A. Asante-Poku, P. Asare, L. Sanchez-Buso, F. Gehre, C.N. Sanoussi, M. Antonio, D. Affolabi, J. Fyfe, P. Beckert, S. Niemann, A.S. Alabi, M.P. Grobusch, R. Kobbe, J. Parkhill, C. Beisel, L. Fenner, E.C. Bottger, C.J. Meehan, S.R. Harris, B.C. de

Jong, D. Yeboah-Manu, and D. Brites. 2021. Phylogenomics of Mycobacterium africanum reveals a new lineage and a complex evolutionary history. *Microb Genom* 7:000477.

de Jong, B.C., I. Adetifa, B. Walther, P.C. Hill, M. Antonio, M. Ota, and R.A. Adegbola. 2010. Differences between tuberculosis cases infected with Mycobacterium africanum, West African type 2, relative to Euro-American Mycobacterium tuberculosis: an update. *FEMS Immunol Med Microbiol* 58:102-105.

de Jong, B.C., M. Antonio, T. Awine, K. Ogungbemi, Y.P. de Jong, S. Gagneux, K. DeRiemer, T. Zozio, N. Rastogi, M. Borgdorff, P.C. Hill, and R.A. Adegbola. 2009. Use of spoligotyping and large sequence polymorphisms to study the population structure of the Mycobacterium tuberculosis complex in a cohort study of consecutive smear-positive tuberculosis cases in The Gambia. *J Clin Microbiol* 47:994-1001.

de Jong, B.C., P.C. Hill, A. Aiken, T. Awine, M. Antonio, I.M. Adetifa, D.J. Jackson-Sillah, A. Fox, K. Deriemer, S. Gagneux, M.W. Borgdorff, K.P. McAdam, T. Corrah, P.M. Small, and R.A. Adegbola. 2008. Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in The Gambia. *J Infect Dis* 198:1037-1043.

Diarra, B., M. Kone, A.C.G. Togo, Y.D.S. Sarro, A.B. Cisse, A. Somboro, B. Degoga, M. Tolofoudie, B. Kone, M. Sanogo, B. Baya, O. Kodio, M. Maiga, M. Belson, S. Orsega, M. Krit, S. Dao, Maiga, II, R.L. Murphy, L. Rigouts, S. Doumbia, S. Diallo, and B.C. de Jong. 2018. Mycobacterium africanum (Lineage 6) shows slower sputum smear conversion on tuberculosis treatment than Mycobacterium tuberculosis (Lineage 4) in Bamako, Mali. *PLoS One* 13:e0208603.

Ehrt, S., D. Schnappinger, S. Bekiranov, J. Drenkow, S. Shi, T.R. Gingeras, T. Gaasterland, G. Schoolnik, and C. Nathan. 2001. Reprogramming of the macrophage transcriptome in response to interferon-gamma and Mycobacterium tuberculosis: signaling roles of nitric oxide synthase-2 and phagocyte oxidase. *J Exp Med* 194:1123-1140.

Eruslanov, E.B., I.V. Lyadova, T.K. Kondratieva, K.B. Majorov, I.V. Scheglov, M.O. Orlova, and A.S. Apt. 2005. Neutrophil responses to Mycobacterium tuberculosis infection in genetically susceptible and resistant mice. *Infect Immun* 73:1744-1753.

Esaulova, E., S. Das, D.K. Singh, J.A. Choreno-Parra, A. Swain, L. Arthur, J. Rangel-Moreno, M. Ahmed, B. Singh, A. Gupta, L.A. Fernandez-Lopez, M. de la Luz Garcia-Hernandez, A. Bucsan, C. Moodley, S. Mehra, E. Garcia-Latorre, J. Zuniga, J. Atkinson,

D. Kaushal, M.N. Artyomov, and S.A. Khader. 2021. The immune landscape in tuberculosis reveals populations linked to disease and latency. *Cell Host Microbe* 29:165-178 e168.

Estevez, O., L. Anibarro, E. Garet, A. Pallares, L. Barcia, L. Calvino, C. Maueia, T. Mussa, F. Fdez-Riverola, D. Glez-Pena, M. Reboiro-Jato, H. Lopez-Fernandez, N.A. Fonseca, R. Reljic, and A. Gonzalez-Fernandez. 2020. An RNA-seq Based Machine Learning Approach Identifies Latent Tuberculosis Patients With an Active Tuberculosis Profile. *Front Immunol* 11:1470.

Eum, S.Y., J.H. Kong, M.S. Hong, Y.J. Lee, J.H. Kim, S.H. Hwang, S.N. Cho, L.E. Via, and C.E. Barry, 3rd. 2010. Neutrophils are the predominant infected phagocytic cells in the airways of patients with active pulmonary TB. *Chest* 137:122-128.

Faridgohar, M., and H. Nikoueinejad. 2017. New findings of Toll-like receptors involved in Mycobacterium tuberculosis infection. *Pathog Glob Health* 111:256-264.

Firdessa, R., S. Berg, E. Hailu, E. Schelling, B. Gumi, G. Erenso, E. Gadisa, T. Kiros, M. Habtamu, J. Hussein, J. Zinsstag, B.D. Robertson, G. Ameni, A.J. Lohan, B. Loftus, I. Comas, S. Gagneux, R. Tschopp, L. Yamuah, G. Hewinson, S.V. Gordon, D.B. Young, and A. Aseffa. 2013. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *Emerg Infect Dis* 19:460-463.

Fontan, P.A., V. Aris, M.E. Alvarez, S. Ghanny, J. Cheng, P. Soteropoulos, A. Trevani, R. Pine, and I. Smith. 2008. Mycobacterium tuberculosis sigma factor E regulon modulates the host inflammatory response. *J Infect Dis* 198:877-885.

Gagneux, S. 2018. Ecology and evolution of Mycobacterium tuberculosis. *Nat Rev Microbiol* 16:202-213.

Gagneux, S., K. DeRiemer, T. Van, M. Kato-Maeda, B.C. de Jong, S. Narayanan, M. Nicol, S. Niemann, K. Kremer, M.C. Gutierrez, M. Hilty, P.C. Hopewell, and P.M. Small. 2006. Variable host-pathogen compatibility in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* 103:2869-2873.

Gay, N.J., and M. Gangloff. 2007. Structure and function of Toll receptors and their ligands. *Annu Rev Biochem* 76:141-165.

Gehre, F., M. Antonio, F. Faihun, M. Odoun, C. Uwizeye, P. de Rijk, B.C. de Jong, and D. Affolabi. 2013. The first phylogeographic population structure and analysis of

transmission dynamics of M. africanum West African 1--combining molecular data from Benin, Nigeria and Sierra Leone. *PLoS One* 8:e77000.

Gideon, H.P., T.K. Hughes, C.N. Tzouanas, M.H. Wadsworth, 2nd, A.A. Tu, T.M. Gierahn, J.M. Peters, F.F. Hopkins, J.R. Wei, C. Kummerlowe, N.L. Grant, K. Nargan, J.Y. Phuah, H.J. Borish, P. Maiello, A.G. White, C.G. Winchell, S.K. Nyquist, S.K.C. Ganchua, A. Myers, K.V. Patel, C.L. Ameel, C.T. Cochran, S. Ibrahim, J.A. Tomko, L.J. Frye, J.M. Rosenberg, A. Shih, M. Chao, E. Klein, C.A. Scanga, J. Ordovas-Montanes, B. Berger, J.T. Mattila, R. Madansein, J.C. Love, P.L. Lin, A. Leslie, S.M. Behar, B. Bryson, J.L. Flynn, S.M. Fortune, and A.K. Shalek. 2022. Multimodal profiling of lung granulomas in macaques reveals cellular correlates of tuberculosis control. *Immunity* 55:827-846 e810.

Gonzalo-Asensio, J., W. Malaga, A. Pawlik, C. Astarie-Dequeker, C. Passemar, F. Moreau, F. Laval, M. Daffe, C. Martin, R. Brosch, and C. Guilhot. 2014. Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. *Proc Natl Acad Sci U S A* 111:11491-11496.

Green, A.M., R. Difazio, and J.L. Flynn. 2013. IFN-gamma from CD4 T cells is essential for host survival and enhances CD8 T cell function during Mycobacterium tuberculosis infection. *J Immunol* 190:270-277.

Groenheit, R., S. Ghebremichael, J. Svensson, P. Rabna, R. Colombatti, F. Riccardi, D. Couvin, V. Hill, N. Rastogi, T. Koivula, and G. Kallenius. 2011. The Guinea-Bissau family of Mycobacterium tuberculosis complex revisited. *PLoS One* 6:e18601.

Gupta, R.K., C.T. Turner, C. Venturini, H. Esmail, M.X. Rangaka, A. Copas, M. Lipman, I. Abubakar, and M. Noursadeghi. 2020. Concise whole blood transcriptional signatures for incipient tuberculosis: a systematic review and patient-level pooled meta-analysis. *Lancet Respir Med* 8:395-406.

Harrington, L.E., R.D. Hatton, P.R. Mangan, H. Turner, T.L. Murphy, K.M. Murphy, and C.T. Weaver. 2005. Interleukin 17-producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. *Nat Immunol* 6:1123-1132.

Havlir, D.V., and P.F. Barnes. 1999. Tuberculosis in patients with human immunodeficiency virus infection. *N Engl J Med* 340:367-373.

Hershberg, R., M. Lipatov, P.M. Small, H. Sheffer, S. Niemann, S. Homolka, J.C. Roach, K. Kremer, D.A. Petrov, M.W. Feldman, and S. Gagneux. 2008. High functional diversity

in Mycobacterium tuberculosis driven by genetic drift and human demography. *PLoS Biol* 6:e311.

Hoang, L.T., P. Jain, T.D. Pillay, M. Tolosa-Wright, U. Niazi, Y. Takwoingi, A. Halliday, L.C. Berrocal-Almanza, J.J. Deeks, P. Beverley, O.M. Kon, and A. Lalvani. 2021. Transcriptomic signatures for diagnosing tuberculosis in clinical practice: a prospective, multicentre cohort study. *Lancet Infect Dis* 21:366-375.

Homolka, S., E. Post, B. Oberhauser, A.G. George, L. Westman, F. Dafae, S. Rusch-Gerdes, and S. Niemann. 2008. High genetic diversity among Mycobacterium tuberculosis complex strains from Sierra Leone. *BMC Microbiol* 8:103.

Jacobsen, M., D. Repsilber, A. Gutschmidt, A. Neher, K. Feldmann, H.J. Mollenkopf, A. Ziegler, and S.H. Kaufmann. 2007. Candidate biomarkers for discrimination between infection and disease caused by Mycobacterium tuberculosis. *J Mol Med (Berl)* 85:613-621.

Jo, E.K., C.S. Yang, C.H. Choi, and C.V. Harding. 2007. Intracellular signalling cascades regulating innate immune responses to Mycobacteria: branching out from Toll-like receptors. *Cell Microbiol* 9:1087-1098.

Kaforou, M., V.J. Wright, T. Oni, N. French, S.T. Anderson, N. Bangani, C.M. Banwell, A.J. Brent, A.C. Crampin, H.M. Dockrell, B. Eley, R.S. Heyderman, M.L. Hibberd, F. Kern, P.R. Langford, L. Ling, M. Mendelson, T.H. Ottenhoff, F. Zgambo, R.J. Wilkinson, L.J. Coin, and M. Levin. 2013. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med* 10:e1001538.

Kim, D., J.M. Paggi, C. Park, C. Bennett, and S.L. Salzberg. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37:907-915.

Kim, M.J., H.C. Wainwright, M. Locketz, L.G. Bekker, G.B. Walther, C. Dittrich, A. Visser, W. Wang, F.F. Hsu, U. Wiehart, L. Tsenova, G. Kaplan, and D.G. Russell. 2010. Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism. *EMBO Mol Med* 2:258-274.

Koch, R. 1882. The etiology of tuberculosis. *Berliner Klinische Wochenschrift* 15:221-230.

Kolde, R. 2019. pheatmap: Pretty Heatmaps. In https://CRAN.R-project.org/package=pheatmap. Last access 29 September 2022.

Li, D., and M. Wu. 2021. Pattern recognition receptors in health and diseases. *Signal Transduct Target Ther* 6:291.

Lin, P.L., and J.L. Flynn. 2015. CD8 T cells and Mycobacterium tuberculosis infection. *Semin Immunopathol* 37:239-249.

Lowe, D.M., A.K. Bandara, G.E. Packe, R.D. Barker, R.J. Wilkinson, C.J. Griffiths, and A.R. Martineau. 2013. Neutrophilia independently predicts death in tuberculosis. *Eur Respir J* 42:1752-1757.

Lowe, D.M., P.S. Redford, R.J. Wilkinson, A. O'Garra, and A.R. Martineau. 2012. Neutrophils in tuberculosis: friend or foe? *Trends Immunol* 33:14-25.

Lugo-Villarino, G., R. Maldonado-Lopez, R. Possemato, C. Penaranda, and L.H. Glimcher. 2003. T-bet is required for optimal production of IFN-gamma and antigen-specific T cell activation by dendritic cells. *Proc Natl Acad Sci U S A* 100:7749-7754.

Maertzdorf, J., M. Ota, D. Repsilber, H.J. Mollenkopf, J. Weiner, P.C. Hill, and S.H. Kaufmann. 2011a. Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis. *PLoS One* 6:e26938.

Maertzdorf, J., D. Repsilber, S.K. Parida, K. Stanley, T. Roberts, G. Black, G. Walzl, and S.H. Kaufmann. 2011b. Human gene expression profiles of susceptibility and resistance in tuberculosis. *Genes Immun* 12:15-22.

Martinez-Perez, A., O. Estevez, and A. Gonzalez-Fernandez. 2022. Contribution and Future of High-Throughput Transcriptomics in Battling Tuberculosis. *Front Microbiol* 13:835620.

McQuaid, C.F., N. McCreesh, J.M. Read, T. Sumner, C.C.-W. Group, R. Houben, R.G. White, and R.C. Harris. 2020. The potential impact of COVID-19-related disruption on tuberculosis burden. *Eur Respir J* 56:2001718.

Meng, H., G. Yaari, C.R. Bolen, S. Avey, and S.H. Kleinstein. 2019. Gene set meta-analysis with Quantitative Set Analysis for Gene Expression (QuSAGE). *PLoS Comput Biol* 15:e1006899.

Mogues, T., M.E. Goodrich, L. Ryan, R. LaCourse, and R.J. North. 2001. The relative importance of T cell subsets in immunity and immunopathology of airborne Mycobacterium tuberculosis infection in mice. *J Exp Med* 193:271-280.

Moreira-Teixeira, L., K. Mayer-Barber, A. Sher, and A. O'Garra. 2018. Type I interferons in tuberculosis: Foe and occasionally friend. *J Exp Med* 215:1273-1285.

Moreira-Teixeira, L., J. Sousa, F.W. McNab, E. Torrado, F. Cardoso, H. Machado, F. Castro, V. Cardoso, J. Gaifem, X. Wu, R. Appelberg, A.G. Castro, A. O'Garra, and M. Saraiva. 2016. Type I IFN Inhibits Alternative Macrophage Activation during Mycobacterium tuberculosis Infection and Leads to Enhanced Protection in the Absence of IFN-gamma Signaling. *J Immunol* 197:4714-4726.

Moreira-Teixeira, L., O. Tabone, C.M. Graham, A. Singhania, E. Stavropoulos, P.S. Redford, P. Chakravarty, S.L. Priestnall, A. Suarez-Bonnet, E. Herbert, K.D. Mayer-Barber, A. Sher, K.L. Fonseca, J. Sousa, B. Ca, R. Verma, P. Haldar, M. Saraiva, and A. O'Garra. 2020. Mouse transcriptome reveals potential signatures of protection and pathogenesis in human tuberculosis. *Nat Immunol* 21:464-476.

Mostowy, S., A. Onipede, S. Gagneux, S. Niemann, K. Kremer, E.P. Desmond, M. Kato-Maeda, and M. Behr. 2004. Genomic analysis distinguishes Mycobacterium africanum. *J Clin Microbiol* 42:3594-3599.

Napier, G., S. Campino, Y. Merid, M. Abebe, Y. Woldeamanuel, A. Aseffa, M.L. Hibberd, J. Phelan, and T.G. Clark. 2020. Robust barcoding and identification of Mycobacterium tuberculosis lineages for epidemiological and clinical studies. *Genome Med* 12:114.

Natarajan, K., M. Kundu, P. Sharma, and J. Basu. 2011. Innate immune responses to M. tuberculosis infection. *Tuberculosis (Edinb)* 91:427-431.

Nathan, A., J.I. Beynor, Y. Baglaenko, S. Suliman, K. Ishigaki, S. Asgari, C.C. Huang, Y. Luo, Z. Zhang, K. Lopez, C.S. Lindestam Arlehamn, J.D. Ernst, J. Jimenez, R.I. Calderon, L. Lecca, I. Van Rhijn, D.B. Moody, M.B. Murray, and S. Raychaudhuri. 2021. Multimodally profiling memory T cells from a tuberculosis cohort identifies cell state associations with demographics, environment and disease. *Nat Immunol* 22:781-793.

Ngabonziza, J.C.S., C. Loiseau, M. Marceau, A. Jouet, F. Menardo, O. Tzfadia, R. Antoine, E.B. Niyigena, W. Mulders, K. Fissette, M. Diels, C. Gaudin, S. Duthoy, W. Ssengooba, E. Andre, M.K. Kaswa, Y.M. Habimana, D. Brites, D. Affolabi, J.B. Mazarati, B.C. de Jong, L. Rigouts, S. Gagneux, C.J. Meehan, and P. Supply. 2020. A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region. *Nat Commun* 11:2917.

O'Garra, A., P.S. Redford, F.W. McNab, C.I. Bloom, R.J. Wilkinson, and M.P. Berry. 2013. The immune response in tuberculosis. *Annu Rev Immunol* 31:475-527.

Ofori-Anyinam, B., G. Dolganov, T. Van, J.L. Davis, N.D. Walter, B.J. Garcia, M. Voskuil, K. Fissette, M. Diels, M. Driesen, C.J. Meehan, D. Yeboah-Manu, M. Coscolla, S. Gagneux, M. Antonio, G. Schoolnik, F. Gehre, and B.C. de Jong. 2017. Significant under expression of the DosR regulon in M. tuberculosis complex lineage 6 in sputum. *Tuberculosis (Edinb)* 104:58-64.

Otchere, I.D., M. Coscolla, L. Sanchez-Buso, A. Asante-Poku, D. Brites, C. Loiseau, C. Meehan, S. Osei-Wusu, A. Forson, C. Laryea, A.I. Yahayah, A. Baddoo, G.A. Ansa, S.Y. Aboagye, P. Asare, S. Borrell, F. Gehre, P. Beckert, T.A. Kohl, S. N'Dira, C. Beisel, M. Antonio, S. Niemann, B.C. de Jong, J. Parkhill, S.R. Harris, S. Gagneux, and D. Yeboah-Manu. 2018. Comparative genomics of Mycobacterium africanum Lineage 5 and Lineage 6 from Ghana suggests distinct ecological niches. *Sci Rep* 8:11269.

Park, H., Z. Li, X.O. Yang, S.H. Chang, R. Nurieva, Y.H. Wang, Y. Wang, L. Hood, Z. Zhu, Q. Tian, and C. Dong. 2005. A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. *Nat Immunol* 6:1133-1141.

Pertea, M., G.M. Pertea, C.M. Antonescu, T.C. Chang, J.T. Mendell, and S.L. Salzberg. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33:290-295.

Pisu, D., L. Huang, J.K. Grenier, and D.G. Russell. 2020. Dual RNA-Seq of Mtb-Infected Macrophages In Vivo Reveals Ontologically Distinct Host-Pathogen Interactions. *Cell Rep* 30:335-350 e334.

Pisu, D., L. Huang, V. Narang, M. Theriault, G. Le-Bury, B. Lee, A.E. Lakudzala, D.T. Mzinza, D.V. Mhango, S.C. Mitini-Nkhoma, K.C. Jambo, A. Singhal, H.C. Mwandumba, and D.G. Russell. 2021. Single cell analysis of M. tuberculosis phenotype and macrophage lineages in the infected lung. *J Exp Med* 218:e20210615.

Ragno, S., M. Romano, S. Howell, D.J. Pappin, P.J. Jenner, and M.J. Colston. 2001. Changes in gene expression in macrophages infected with Mycobacterium tuberculosis: a combined transcriptomic and proteomic approach. *Immunology* 104:99-108.

Ramakrishnan, L. 2012. Revisiting the role of the granuloma in tuberculosis. *Nat Rev Immunol* 12:352-366.

Reiling, N., S. Homolka, K. Walter, J. Brandenburg, L. Niwinski, M. Ernst, C. Herzmann, C. Lange, R. Diel, S. Ehlers, and S. Niemann. 2013. Clade-specific virulence patterns of Mycobacterium tuberculosis complex strains in human primary macrophages and aerogenically infected mice. *mBio* 4:e00250-00213.

Rienksma, R.A., M. Suarez-Diez, H.J. Mollenkopf, G.M. Dolganov, A. Dorhoi, G.K. Schoolnik, V.A. Martins Dos Santos, S.H. Kaufmann, P.J. Schaap, and M. Gengenbacher. 2015. Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing. *BMC Genomics* 16:34.

Rijnink, W.F., T.H.M. Ottenhoff, and S.A. Joosten. 2021. B-Cells and Antibodies as Contributors to Effector Immune Responses in Tuberculosis. *Front Immunol* 12:640168.

Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, and G.K. Smyth. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47.

Robinson, M.D., D.J. McCarthy, and G.K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139-140.

Roe, J., C. Venturini, R.K. Gupta, C. Gurry, B.M. Chain, Y. Sun, J. Southern, C. Jackson, M.C. Lipman, R.F. Miller, A.R. Martineau, I. Abubakar, and M. Noursadeghi. 2020. Blood Transcriptomic Stratification of Short-term Risk in Contacts of Tuberculosis. *Clin Infect Dis* 70:731-737.

Rudolf, F., G. Lemvik, E. Abate, J. Verkuilen, T. Schon, V.F. Gomes, J. Eugen-Olsen, L. Ostergaard, and C. Wejse. 2013. TBscore II: refining and validating a simple clinical score for treatment monitoring of patients with pulmonary tuberculosis. *Scand J Infect Dis* 45:825-836.

Sanoussi, C.N., M. Coscolla, B. Ofori-Anyinam, I.D. Otchere, M. Antonio, S. Niemann, J. Parkhill, S. Harris, D. Yeboah-Manu, S. Gagneux, L. Rigouts, D. Affolabi, B.C. de Jong, and C.J. Meehan. 2021. Mycobacterium tuberculosis complex lineage 5 exhibits high levels of within-lineage genomic diversity and differing gene content compared to the type strain H37Rv. *Microb Genom* 7:000437.

Scordo, J.M., D.L. Knoell, and J.B. Torrelles. 2016. Alveolar Epithelial Cells in Mycobacterium tuberculosis Infection: Active Players or Innocent Bystanders? *J Innate Immun* 8:3-14.

Serbina, N.V., C.C. Liu, C.A. Scanga, and J.L. Flynn. 2000. CD8+ CTL from lungs of Mycobacterium tuberculosis-infected mice express perforin in vivo and lyse infected macrophages. *J Immunol* 165:353-363.

Silva, B.D.S., M.M. Trentini, A.C. da Costa, A. Kipnis, and A.P. Junqueira-Kipnis. 2014. Different phenotypes of CD8+ T cells associated with bacterial load in active tuberculosis. *Immunol Lett* 160:23-32.

Silva, M.L., B. Ca, N.S. Osorio, P.N.S. Rodrigues, A.R. Maceiras, and M. Saraiva. 2022. Tuberculosis caused by Mycobacterium africanum: Knowns and unknowns. *PLoS Pathog* 18:e1010490.

Singhania, A., R. Verma, C.M. Graham, J. Lee, T. Tran, M. Richardson, P. Lecine, P. Leissner, M.P.R. Berry, R.J. Wilkinson, K. Kaiser, M. Rodrigue, G. Woltmann, P. Haldar, and A. O'Garra. 2018. A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. *Nat Commun* 9:2308.

Sousa, J., B. Ca, A.R. Maceiras, L. Simoes-Costa, K.L. Fonseca, A.I. Fernandes, A. Ramos, T. Carvalho, L. Barros, C. Magalhaes, A. Chiner-Oms, H. Machado, M.I. Veiga, A. Singh, R. Pereira, A. Amorim, J. Vieira, C.P. Vieira, A. Bhatt, F. Rodrigues, P.N.S. Rodrigues, S. Gagneux, A.G. Castro, J.T. Guimaraes, H.N. Bastos, N.S. Osorio, I. Comas, and M. Saraiva. 2020. Mycobacterium tuberculosis associated with severe tuberculosis evades cytosolic surveillance systems and modulates IL-1beta production. *Nat Commun* 11:1949.

Steen, C.B., C.L. Liu, A.A. Alizadeh, and A.M. Newman. 2020. Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods Mol Biol* 2117:135-157.

Stenger, S., D.A. Hanson, R. Teitelbaum, P. Dewan, K.R. Niazi, C.J. Froelich, T. Ganz, S. Thoma-Uszynski, A. Melian, C. Bogdan, S.A. Porcelli, B.R. Bloom, A.M. Krensky, and R.L. Modlin. 1998. An antimicrobial activity of cytolytic T cells mediated by granulysin. *Science* 282:121-125.

Stucki, D., D. Brites, L. Jeljeli, M. Coscolla, Q. Liu, A. Trauner, L. Fenner, L. Rutaihwa, S. Borrell, T. Luo, Q. Gao, M. Kato-Maeda, M. Ballif, M. Egger, R. Macedo, H. Mardassi, M. Moreno, G. Tudo Vilanova, J. Fyfe, M. Globan, J. Thomas, F. Jamieson, J.L. Guthrie, A. Asante-Poku, D. Yeboah-Manu, E. Wampande, W. Ssengooba, M. Joloba, W. Henry Boom, I. Basu, J. Bower, M. Saraiva, S.E.G. Vaconcellos, P. Suffys, A. Koch, R. Wilkinson, L. Gail-Bekker, B. Malla, S.D. Ley, H.P. Beck, B.C. de Jong, K. Toit, E. Sanchez-Padilla, M. Bonnet, A. Gil-Brusola, M. Frank, V.N. Penlap Beng, K. Eisenach, I. Alani, P. Wangui Ndung'u, G. Revathi, F. Gehre, S. Akter, F. Ntoumi, L. Stewart-Isherwood, N.E. Ntinginya, A. Rachow, M. Hoelscher, D.M. Cirillo, G. Skenders, S. Hoffner, D. Bakonyte, P. Stakenas, R. Diel, V. Crudu, O. Moldovan, S. Al-Hajoj, L. Otero,

F. Barletta, E. Jane Carter, L. Diero, P. Supply, I. Comas, S. Niemann, and S. Gagneux. 2016. Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat Genet* 48:1535-1543.

Stucki, D., B. Malla, S. Hostettler, T. Huna, J. Feldmann, D. Yeboah-Manu, S. Borrell, L. Fenner, I. Comas, M. Coscolla, and S. Gagneux. 2012. Two new rapid SNP-typing methods for classifying Mycobacterium tuberculosis complex into the main phylogenetic lineages. *PLoS One* 7:e41253.

Suliman, S., E.G. Thompson, J. Sutherland, J. Weiner, 3rd, M.O.C. Ota, S. Shankar, A. Penn-Nicholson, B. Thiel, M. Erasmus, J. Maertzdorf, F.J. Duffy, P.C. Hill, E.J. Hughes, K. Stanley, K. Downing, M.L. Fisher, J. Valvo, S.K. Parida, G. van der Spuy, G. Tromp, I.M.O. Adetifa, S. Donkor, R. Howe, H. Mayanja-Kizza, W.H. Boom, H.M. Dockrell, T.H.M. Ottenhoff, M. Hatherill, A. Aderem, W.A. Hanekom, T.J. Scriba, S.H.E. Kaufmann, D.E. Zak, G. Walzl, and T.A.C.S.c.s.t. Gc6-74 cohort study team. 2018. Four-Gene Pan-African Blood Signature Predicts Progression to Tuberculosis. *Am J Respir Crit Care Med* 197:1198-1208.

Tabone, O., R. Verma, A. Singhania, P. Chakravarty, W.J. Branchett, C.M. Graham, J. Lee, T. Trang, F. Reynier, P. Leissner, K. Kaiser, M. Rodrigue, G. Woltmann, P. Haldar, and A. O'Garra. 2021. Blood transcriptomics reveal the evolution and resolution of the immune response in tuberculosis. *J Exp Med* 218:e20210915.

Tientcheu, L.D., M.C. Haks, S.C. Agbla, J.S. Sutherland, I.M. Adetifa, S. Donkor, E. Quinten, M. Daramy, M. Antonio, B. Kampmann, T.H. Ottenhoff, H.M. Dockrell, and M.O. Ota. 2016. Host Immune Responses Differ between M. africanum- and M. tuberculosis-Infected Patients following Standard Anti-tuberculosis Treatment. *PLoS Negl Trop Dis* 10:e0004701.

Tientcheu, L.D., J. Maertzdorf, J. Weiner, I.M. Adetifa, H.J. Mollenkopf, J.S. Sutherland, S. Donkor, B. Kampmann, S.H. Kaufmann, H.M. Dockrell, and M.O. Ota. 2015. Differential transcriptomic and metabolic profiles of M. africanum- and M. tuberculosis-infected patients after, but not before, drug treatment. *Genes Immun* 16:347-355.

Tientcheu, L.D., J.S. Sutherland, B.C. de Jong, B. Kampmann, J. Jafali, I.M. Adetifa, M. Antonio, H.M. Dockrell, and M.O. Ota. 2014. Differences in T-cell responses between Mycobacterium tuberculosis and Mycobacterium africanum-infected patients. *Eur J Immunol* 44:1387-1398.

Torrado, E., and A.M. Cooper. 2010. IL-17 and Th17 cells in tuberculosis. *Cytokine Growth Factor Rev* 21:455-462.

Vynnycky, E., and P.E. Fine. 2000. Lifetime risks, incubation period, and serial interval of tuberculosis. *Am J Epidemiol* 152:247-263.

Warsinske, H., R. Vashisht, and P. Khatri. 2019. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. *PLoS Med* 16:e1002786.

WHO. 2021. Global Tuberculosis Report.

WHO. 2022. Summary of tuberculosis data. In https://worldhealthorg.shinyapps.io/TBrief/. Last access 29 September 2022.

Wickham, H., C. Sievert, and SpringerLink. 2016. ggplot2 : Elegant Graphics for Data Analysis. In Use R! Springer, Springer International Publishing.

Wiens, K.E., and J.D. Ernst. 2016a. The Mechanism for Type I Interferon Induction by Mycobacterium tuberculosis is Bacterial Strain-Dependent. *PLoS Pathog* 12:e1005809.

Wiens, K.E., and J.D. Ernst. 2016b. Type I Interferon is Pathogenic During Chronic Mycobacterium africanum Infection. *J Infect Dis* 214:1893-1896.

Wolf, A.J., L. Desvignes, B. Linas, N. Banaiee, T. Tamura, K. Takatsu, and J.D. Ernst. 2008. Initiation of the adaptive immune response to Mycobacterium tuberculosis depends on antigen production in the local lymph node, not the lungs. *J Exp Med* 205:105-115.

Wolf, A.J., B. Linas, G.J. Trevejo-Nunez, E. Kincaid, T. Tamura, K. Takatsu, and J.D. Ernst. 2007. Mycobacterium tuberculosis infects dendritic cells with high frequency and impairs their function in vivo. *J Immunol* 179:2509-2519.

Yeboah-Manu, D., P. Asare, A. Asante-Poku, I.D. Otchere, S. Osei-Wusu, E. Danso, A. Forson, K.A. Koram, and S. Gagneux. 2016. Spatio-Temporal Distribution of Mycobacterium tuberculosis Complex Strains in Ghana. *PLoS One* 11:e0161892.

Yu, G., and Q.Y. He. 2016. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst* 12:477-479.

Zak, D.E., A. Penn-Nicholson, T.J. Scriba, E. Thompson, S. Suliman, L.M. Amon, H. Mahomed, M. Erasmus, W. Whatney, G.D. Hussey, D. Abrahams, F. Kafaar, T. Hawkridge, S. Verver, E.J. Hughes, M. Ota, J. Sutherland, R. Howe, H.M. Dockrell, W.H.

Boom, B. Thiel, T.H.M. Ottenhoff, H. Mayanja-Kizza, A.C. Crampin, K. Downing, M. Hatherill, J. Valvo, S. Shankar, S.K. Parida, S.H.E. Kaufmann, G. Walzl, A. Aderem, W.A. Hanekom, Acs, and G.C.c.s. groups. 2016. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* 387:2312-2322.

Zimmermann, M., M. Kogadeeva, M. Gengenbacher, G. McEwen, H.J. Mollenkopf, N. Zamboni, S.H.E. Kaufmann, and U. Sauer. 2017. Integration of Metabolomics and Transcriptomics Reveals a Complex Diet of Mycobacterium tuberculosis during Early Macrophage Infection. *mSystems* 2:e00057-00017.