University of Southampton

Faculty of Medicine

Clinical and Experimental Sciences

Investigating tuberculosis immunopathogenesis by studying clinical samples and 3-dimensional cell culture using unbiased methodology

by

Michaela Thiranan Reichmann

ORCHID ID 0000-0002-3015-9827

Thesis for the degree of Doctor of Philosophy

July 2021

University of Southampton

Abstract

Faculty of Medicine

Clinical and Experimental Sciences

Doctor of Philosophy

Investigating tuberculosis immunopathogenesis by studying clinical samples and

3-dimensional cell culture using unbiased methodology

by

Michaela Thiranan Reichmann

The global burden of tuberculosis continues to be of pandemic proportions, currently with a quarter of the world's population infected with *Mycobacterium tuberculosis* (Mtb), and 1.4 million people dying from Mtb infection in 2019. Sarcoidosis is another human granulomatous condition primarily affecting the lung and lymph nodes. TB and sarcoidosis share histological and clinical features which can be indistinguishable, including immune-related phenomena which suggest shared immunological processes. Here, I investigate TB pathogenesis by studying untreated TB and sarcoidosis clinical samples, alongside control tissue. Twenty-four formalin-fixed paraffin-embedded biopsies were selected, and granulomas excised using laser capture microdissection, with an equivalent area of control tissue excised. Total RNA was isolated, and sequenced by Ion Torrent Sequencing. In parallel, three cell culture models of TB were studied; 2-dimensional (2D), 3-dimensional (3D) with Alginate, and 3-dimensional with Alginate-Collagen. Peripheral blood mononuclear cells from six healthy donors were infected overnight with the Mtb strain H37Rv. Total RNA was extracted and sequenced by Illumina HiSeq. For bioinformatic analysis, both RNA sequencing data sets were aligned to human reference genome GRCh38 using kallisto, and differentially expressed genes identified using a linear model (R: limma, TMM normalised voom transformed data) with false discovery rate less than 0.05 and absolute $\log_2$ fold change at least 1.5. The overall strategy was to analyse pathways common to TB and sarcoid, and those unique to each condition, and then investigate these in a cell culture model whose gene expression best reflects that of human TB. Using a systematic selection process, comparison of clinical TB samples with infected 3D Alginate-Collagen microspheres identified twelve potential targets, including sphingosine kinase 1 (SphK1). Inhibition of SphK1 resulted in control of Mtb growth, concurrently reducing intracellular pH in infected monocytes and suppressing inflammatory mediator secretion. Immunohistochemistry of clinical TB specimens confirmed the presence of SphK1, demonstrating a potential novel host-directed therapy target to improve outcomes in TB. Therefore, by combining unbiased analysis of human clinical samples with a biomimetic model, I have established a translational pipeline to identify new therapeutic approaches.

Table of Contents

Table of Tables

Table of Figures

Table of Figures

Research Thesis: Declaration of Authorship

Print name: Michaela Reichmann

Title of thesis: Investigating tuberculosis immunopathogenesis by studying clinical samples and 3-dimensional cell culture using unbiased methodology

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help ;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:
   https://www.atsjournals.org/doi/10.1164/ajrccm-conference.2020.201.1_MeetingAbstracts.A1048
   Gene Expression Omnibus (GEO) Accession number GSE174443
   Gene Expression Omnibus (GEO) Accession number GSE174566

Signature: ..... Date: 6th July 2021

Acknowledgements

I would like to express my immense gratitude to Professor Paul Elkington, Dr Marta Polak, and Dr Liku Tezera for training and supporting me through these years of research, understanding my needs as a clinician and as a parent, and without whom this research would not have been possible. Particular thanks to Andres Vallejo for your patience and help with bioinformatic training. A big thank you also to the TB-CD1 and Systems Immunology research groups for valuable feedback, and other members of Level E for your help and comradery. Special thanks to Dr Susan Wilson and her team in the Histochemistry Research Unit (Jon Ward, Jenny Norman and Jamila McRobb) for accommodating my early starts and preparing my samples urgently. Thank you also to lab managers Jenny Russell and Regina Teo who have gone out of their way to help me. Finally, huge thanks to Dr James Reynolds whose unwavering care and support, and technical expertise, has been invaluable.

Chapter 1: Introduction

1.1 Tuberculosis

1.1.1 Global burden of TB

Tuberculosis (TB) remained the leading cause of mortality in the world due to a single infectious agent until the COVID-19 pandemic, and results from *Mycobacterium tuberculosis* (Mtb)-driven pathology (WHO, 2020). The World Health Organisation (WHO) declared TB a global health emergency, prompting the End TB Strategy, a worldwide resolve to reduce TB incidence by 80% and related mortality by 90% by 2030 (Uplekar *et al.*, 2015). In 2019, the global incidence of TB was 10 million, with a mortality of 1.4 million people (WHO, 2020). Despite recent advances in the diagnosis and treatment of other globally important infectious diseases such as malaria and HIV, TB morbidity and mortality has not significantly declined, with an estimated quarter of the world's population currently infected with TB (Houben and Dodd, 2016). A vast majority of those infected with TB are asymptomatic and have latent TB, otherwise known as "dormant TB". Latent TB persists in the human host for up to decades, with approximately 10% then going on to develop active TB disease (Vynnycky and Fine, 2000; Shea *et al.*, 2014).

1.1.2 TB life cycle

Mtb was first identified as a pathogen by Robert Koch in 1882 (Koch, 1882). Mtb is an obligate pathogen of man, where both host and pathogen have coevolved together over an estimated 70,000 years (Gagneux, 2012), though this estimate varies considerably. Symptoms of active TB include fever, night sweats and cough. Coughing is essential to the life cycle of Mtb by allowing transmission of the organism through aerosol droplets, hence TB is disseminated in the environment to other hosts before the diagnosis is made and treatment commenced (Fennelly *et al.*, 2012). Exposure to Mtb aerosol droplets leads to its implantation in the host tissue, usually in the lung bases (Frieden *et al.*, 2003). During this process Mtb exposure generates a spectrum of immune responses in the host (O'Garra and Berry, 2013). A small minority (2%) develop primary TB disease with widespread dissemination, often seen as miliary TB (Sharma *et al.*, 2005), and this occurs in the context of immunocompromise such as HIV infection or the first year of life. A further 2% instead develop extra-pulmonary disease. However, almost 90% develop latent TB, with the initial site of implantation residing in calcified granulomas at the lung base identifiable as a Ghon focus on chest x-ray (Lin and Flynn, 2010). In sharp contrast to the rather indolent nature of latent TB, active TB manifests with destruction of lung tissue and cavity formation, which is often irreversible, and occurs at the lung apices (Frieden *et al.*, 2003; Skoura, Zumla and Bomanji, 2015). Clinically, the reactivation of TB from its latent form can be observed after commencement of immunosuppressive

therapy such as anti-TNF drugs (Keane *et al.*, 2001) but worldwide the majority of TB occurs in the absence of any discernible immunocompromise.

1.1.3 Host response to M.tuberculosis

The host immune response plays a significant role in the destructive processes seen in TB. Mtb is phagocytosed by alveolar macrophages, but is able to delay phagosome fusion (Sturgill-Koszycki *et al.*, 1994), resulting in its intracellular survival. The adaptive immune response, particularly $CD4^+$ T cells, is important in granuloma formation (Peters and Ernst, 2003), with secretion of cytokines and matrix metalloproteinases (Salgame, 2011), as well as triggering inflammatory pathways such as TNF-α (Tumour Necrosis Factor alpha) and IFN-γ (Interferon gamma) (Flynn and Chan, 2001) which lead to macrophage activation and ultimately control of Mtb growth. Deficiency of these pathways leads to disseminated disease, though there is no evidence that an excess is protective. Correspondingly, disseminated TB is reported in a proportion of patients after receiving checkpoint inhibitor therapy to enhance T cell function in cancer treatment, including PD-1 (Programmed cell Death 1) blockers (Barber *et al.*, 2019), indicating host immune responses in TB are finely balanced.

1.1.4 Spectrum of TB disease

TB has long been distinguished by two clinical phenotypes: latent and active. However, a less binary spectrum of TB disease is now recognised, where different host and environmental factors may play their role (O'Garra and Berry, 2013). Indeed, post-mortem anatomic studies in the 1920s found almost half the patients had live Mtb in their mediastinal lymph nodes without evidence of clinical disease (Opie and Aronson, 1927), suggesting that individuals with "latent" TB actually have subclinical TB. The precise events that lead to the switch from latent TB to active disease are only partially understood and involve a change from containment to immunopathology (Drain *et al.*, 2018). Some have recently suggested that the events that lead from latency to reactivation are more due to a loss of tolerance by the host than pathogen-specific factors (Divangahi, Khan and Kaufmann, 2018; Olive and Sassetti, 2018), suggesting a change in perception of the disease process. Host-directed therapy, whereby modification of the host-mediated responses to pathogens is used to improve clinical outcomes, may play an exciting role in treating TB. If we can understand the key molecular pathways that allow susceptible individuals to progress to reactivation of TB, the pathways would be a potential target for host-directed therapy (Hawn and Vandal, 2013), and this knowledge is also essential to safely informing novel vaccination approaches.

1.2 The role of the extracellular matrix

TB is highly transmissible when lung cavitation occurs, which involves extensive destruction of the lung extracellular matrix (ECM) (Sathyamoorthy *et al.*, 2015). The ECM plays a pivotal role in

regulating the cell microenvironment (Mattila *et al.*, 2013). Cytokine signalling is dependent on the surrounding spatial organisation of cells, where proteomic studies have recently shown proinflammatory conditions within the centre of TB granulomas and an anti-inflammatory environment surrounding granulomas (Marakalala *et al.*, 2016). Type IV collagen is the main component of lung extracellular matrix, while type I collagen gives tensile strength. Collagen fibril destruction is an early key event in cavity formation (Al Shammari *et al.*, 2015), caused by upregulation of matrix metalloproteinases (MMPs) during Mtb infection. In particular, MMP-1 is a collagenase implicated in lung tissue destruction and cavity formation (Ong and Friedland, 2014). HIV patients with CD4 counts less than 200, who are co-infected with Mtb, do not develop lung cavities but instead systemic disseminated disease. Expression of MMP activity is reduced in these HIV co-infected patients (Walker and Elkington, 2012) and is a clinical example of how MMP-driven matrix destruction occurs dues to an excessive immune response (Elkington and Friedland, 2011b).

However, it is not fully understood how active TB disease develops, nor why pathology differs in each parts of the lung. What is evident is the human immune response to Mtb not only protects the host, but paradoxically can destroy the host tissue. There may be excessive pathogen-driven inflammation, or alternatively a failure of resolution of the host inflammatory response. Therefore, a better understanding of the immunopathological processes during Mtb infection, particularly those involved in the progression of the TB granuloma to the active destructive form of TB disease, is needed.

1.3 TB, sarcoidosis and autoimmunity

1.3.1 Similarities between TB and sarcoidosis

TB is characterised by granuloma formation, and the other "classical" human granulomatous disease is sarcoidosis. Sarcoidosis is a chronic inflammatory condition with pathognomonic features of non-caseating granuloma formation most commonly found in the lungs and affecting multiple organ systems (Mitchell *et al.*, 1977; Baughman *et al.*, 2001). Clinical features shared between TB and sarcoidosis include constitutional symptoms such as fever, malaise, anorexia and weight loss, immune-related conditions such as erythema nodosum, inflammatory arthritis and uveitis, and central nervous system involvement (Iannuzzi, Rybicki and Teirstein, 2007). Sarcoidosis is treated with steroids, while severe cases of TB are often treated with adjunctive steroids alongside antibacterial therapy (Judson, 1999). TB and sarcoidosis are also recognised adverse events of immune-checkpoint inhibitor therapy for cancer (Elkington *et al.*, 2018; Nishino *et al.*, 2018). Interestingly, spontaneous regression occurs in a significant proportion of patients with these conditions. A third of those infected with TB spontaneously regressed in the pre-antibiotic era

(Dubos and Dubos, 1987), compared to nearly two thirds of those with sarcoidosis (Costabel and Hunninghake, 1999). Histologically TB and sarcoidosis can appear identical with the formation of well defined granulomas containing multinucleate giant cells, surrounded by a $CD4^+$ T cell infiltrate (Elkington, Tebruegge and Mansour, 2016). Despite using Ziehl-Neelsen or Auramine-Rhodamine staining, the Mtb bacilli are often difficult to see microscopically in TB granulomas as they are present in very low abundance, despite subsequent culture positivity. Therefore, clinical observation demonstrates that TB and sarcoidosis mimic each other and subsequently may share underlying pathological mechanisms, though these are not yet understood.

1.3.2 Immunology of sarcoidosis

In the UK, the annual incidence of sarcoidosis is 5 per 100 000 person-years (Gribbin *et al.*, 2006). However, the aetiology of sarcoidosis is still not known, with host, infectious and environmental causative factors proposed including mycobacteria (Song *et al.*, 2005; Brownell *et al.*, 2011). Macrophage activation is thought to be the initial step in a cascade of immune reactions, with elevated inflammatory mechanisms such as TNF-α seen in alveolar macrophages (Ziegenhagen *et al.*, 2002) and IFN-γ in bronchoalveolar lavage fluid (Shigehara *et al.*, 2001) from sarcoidosis patients. Granuloma formation and persistent inflammation occurs involving both $CD8^+$ and $CD4^+$ T cells (Chen and Moller, 2015), with spontaneous regression occurring in a majority of patients, and progression to fibrosis ensuing in others. An abundance of Serum Amyloid A (SAA), an amyloid precursor protein, is also seen in sarcoidosis granulomas and likely to promote chronic granulomatous inflammation (Chen *et al.*, 2010). Clinically, there remains a lack of standardised care for sarcoidosis patients, primarily due to a wide range of clinical manifestations and uncertainty of disease trajectory (Spagnolo *et al.*, 2018), which requires a better understanding of underlying disease mechanisms.

1.3.3 Autoimmunity and TB overlap

Genetic susceptibility is a recognized feature of sarcoidosis (Valeyre *et al.*, 2014) and it has been suggested that sarcoidosis is an autoimmune disease (Zissel and Muller-Quernheim, 2016). Autoimmune diseases also share the symptoms of TB described above and this has led our group to propose, somewhat controversially, that there may be some shared underlying mechanisms (Elkington, Tebruegge and Mansour, 2016). Different lines of evidence have been collated to support this view. Clinically, circulating autoantibodies such as anti-CCP (anti–cyclic citrullinated peptide), anti-Scl-70 (anti-topoisomerase 1) and anti-cardiolipin, strongly associated with autoimmune conditions are found in up to 40% of TB patients (Kakumanu *et al.*, 2008; Shen *et al.*, 2013), and severe TB is correlated with autoimmune-associated genotypes (Pagan and Ramakrishnan, 2015). Experimentally, Freund's adjuvant contains heat-killed Mtb and is used to

stimulate models of autoimmunity in the mouse. Recent transcriptomic analysis of blood samples from TB patients found a greater proportion of genes shared with autoimmune conditions Systemic Lupus Erythematosus (SLE) and adult-onset Still's disease than with infectious conditions (Clayton *et al.*, 2017). Importantly, the TB cohort had culture-confirmed TB and had not received treatment, so immune pathways were not altered but instead taken in their native state. However, despite the theoretical arguments suggesting an overlap between TB, sarcoid and autoimmune disease, further experimental investigation is required to prove a link.

1.4 Experimental models of tuberculosis

It could be argued that many of the disappointments in advancing the TB field may be ascribed to failings of the model systems used to investigate disease processes. Model systems include cell culture, animal models and clinical investigation, all of which have contributed to understanding TB pathogenesis.

1.4.1 Two-dimensional cell culture

Mtb causes TB in the human species alone, necessitating the use of human cells to study the host immune response. The most widely used cell culture model of human cells is the 2-dimensional (2D) cell culture of human cell lines and consists of a monolayer of cells adherent to flat rigid substrates then infected with Mtb. This 2D cell culture model provides understanding of cell-cell interactions but cannot offer the essential microenvironment more representative of *in vivo* conditions than 3-dimensional (3D) models provide (Wenzel and Steigemann, 2014). The altered morphology and physiology of the cells cultured in a monolayer, together with the lack of extracellular matrix and diffusion gradients, greatly limits the 2D cell culture model (Elkington *et al.*, 2019).

1.4.2 Animal models

Compared to cell culture, animal models offer significant advantages to understanding the key biological processes involved in Mtb infection as they can represent the full complexity of *in vivo* infection. However, the animal models used to date lack key features integral to Mtb infection in the human host (Vilaplana and Cardona, 2014). Mice models are the most widely used, but do not form true caseous necrosis or cavities during Mtb infection (Cardona, 2010). The macaque model is most similar to humans in that caseous necrosis is observed in Mtb infection, though unlike in humans, subsequent lung tissue destruction and fibrosis appears to be limited. Furthermore, 50% of animals progress after infection, showing that even primate models of infection have significant differences. During Mtb infection, the guinea pig model rarely cavitates and has a much higher mortality than in humans, with all animals progressing after infection, making the study of Mtb immunopathology in the early inflammatory phase of active TB from latent TB difficult. Rabbit

models of Mtb share more features with humans, though show a stronger response in the formation of lung cavities to *Mycobacterium bovis* than Mtb, whereas *M.bovis* infection in humans causes greater extra-pulmonary than pulmonary disease (Gupta and Katoch, 2005). Lastly, the zebrafish model utilises *Mycobacterium marinum* infection, and has given many novel insights into early innate events, but takes up this infection through the gastrointestinal tract, does not have lungs and the larval models lack T cells. Therefore, despite the invaluable insights offered by animal models of TB, human cells are required to further improve our understanding of the complex host immune response to Mtb.

1.4.3 Human TB granuloma 3-dimensional cell culture model

In order to address the need for more advanced cell culture models of TB, the Southampton TB group have pioneered a 3-dimensional humanised cell culture model of the human granuloma in Mtb infection over the last seven years (Tezera *et al.*, 2017). The bioelectrospray (BES) model generates microspheres of Alginate gel in which human peripheral blood mononuclear cells (PBMCs) and Mtb form clusters and grow together. The BES system consists of a syringe and attached tube to a needle secured above a spinning gelling bath, with a voltage between the needle and bath to accelerate the spheres from the needle tip and control microsphere size (Workman and Jayasinghe, 2014). For each experiment, fresh PBMCs are isolated from a human blood donor and cultured with Mtb overnight. The following day, the infected PBMCs are mixed into an Alginate gel containing collagen, the major component in the human extracellular lung matrix. The Alginate-Collagen gel containing PBMCs are transferred into the syringe and flow is set at a rate of 10ml/hour into the BES needle. Here a voltage difference of 7kV is applied which together with $CaCl_2$ in the gelling bath permits the formation of uniform droplets shaped as spheres with a diameter of 500µm. The BES cell culture model of TB is potentially the most advanced in vitro model of the human granuloma developed to date.

1.5 Emerging methods to investigate TB immunopathology

1.5.1 Tissue compartments

Clearly a significant overlap in inflammatory mechanisms exists between TB and sarcoidosis, but this is poorly understood and requires investigation. Mtb is an obligate human pathogen, and capturing the disease state that is just evolving and naive to immune pathway modification by treatment is essential. Another consideration is the importance of the tissue compartment analysed. Many studies have used peripheral blood cells, while others have used lung or lymph node samples, and an arising theme is that differing results are observed from different compartments (Sathaliyawala *et al.*, 2013; Ogongo, Porterfield and Leslie, 2019; Ogongo *et al.*, 2021). Emerging methods to

investigate human disease may provide the opportunity to not only study tissue close to the site of disease, but also particular intracellular compartments such as the granuloma. Laser capture microdissection (LCM) dissects specific tissue areas of interest in a biological specimen under direct microscopic vision, and preserves their quality so nucleic acid extraction and analysis including RNA-seq can be performed (Vukmirovic *et al.*, 2017). Therefore, it permits targeting of specific lesions, without taking the entire RNA of a tissue biopsy.

1.5.2 RNA sequencing developments

Recently RNA-seq methods have developed which can sequence RNA fragments of low quality (RNA Integrity Number < 2) for downstream transcriptomic analysis (Li *et al.*, 2015). This is very relevant in the field of TB where obtaining biopsy samples from patients who have not received treatment for TB or mimicking conditions such as sarcoidosis is very limited due to ethical medical practice. Hence, the diagnosis of TB from histological specimens is often an incidental finding, where the biopsy specimen has already undergone formalin fixation, as it was taken to exclude other conditions such as malignancy. Together, LCM and RNA-seq provide an exciting opportunity to investigate pathological conditions in formalin-fixed paraffin-embedded (FFPE) samples, and recent work has used this combination of methods to explore gene expression in granulomatous diseases including four TB patients (Casanova *et al.*, 2020). However, to my knowledge this methodology has yet been utilised to investigate the pathogenesis of treatment naïve TB and sarcoidosis.

1.6 The analysis challenge generated by unbiased data

1.6.1 Bioinformatic analysis

RNA sequencing analyses the entire transcriptome, and therefore is an “unbiased” methodology. Unbiased bioinformatic analysis is a huge challenge facing biomedical science in the current era. The volume of data generated from sequencing is enormous, threatening to outstrip methods to analyse it, and can lead to difficulties in data storage and processing (Goodwin, McPherson and McCombie, 2016). Furthermore, there is an extraordinary wealth of information gained from the study of transcriptomics, such as gene and gene pathway identification, gene set enrichment analysis, and protein network and protein regulatory hub identification (Conesa *et al.*, 2016). However, bioinformatic analysis has its own setbacks, with an array of heterogeneous methods and pipelines to choose from, each with their own advantages and limitations (Sahraeian *et al.*, 2017). Indeed, comparisons of multiple software tools for differential expression for a given sample selection give a large variation in results and a significant proportion of results which do not overlap with other. Choosing the most appropriate bioinformatics pipeline and understanding its limitations is therefore

crucial to drawing meaningful biological conclusions that can directly translate to patient care. This unbiased data analysis requires insight from a combined clinical and bioinformatic perspective.

1.6.2 The importance of clinical sample selection

Along with bioinformatics challenges, sample selection is another crucial factor in generating biologically meaningful results from translational studies. Certain criteria need to be fulfilled to achieve this in order to understand TB. Firstly, human samples must be used, as TB and sarcoidosis are diseases exclusive to humans (Brites and Gagneux, 2015; Locke, Schlesinger and Crouser, 2020). Secondly, samples should be taken in a disease state whose immune response is naïve to modification by pharmacotherapy including immunosuppressants. Thirdly, samples should come from the main organ affected. In the case of TB and sarcoidosis, the predominant sites are the lung and regional lymph nodes. Lastly, in order to minimise any change in immune responses, the patients providing the samples should ideally not have HIV disease or other organ dysfunction that may confound analysis.

1.7 Hypothesis and Aims

TB and sarcoidosis are globally important human diseases with likely shared mechanisms, and the combination of unbiased methodology with the study of clinical samples and 3D cell culture provides an opportunity to understand disease pathogenesis.

1.7.1 Hypothesis

Investigation of TB pathogenesis by studying TB and sarcoidosis clinical samples and the 3D cell culture model of TB by unbiased methodology will identify key underlying mechanisms.

1.7.2 Aims

1. To compare differential gene expression in mediastinal/neck lymph nodes of untreated active TB, untreated sarcoidosis and normal patients. Communal inflammatory pathways are likely to exist between TB and sarcoidosis, and bioinformatic analysis of regulatory pathways may provide valuable insight into these granulomatous diseases.
2. To perform comprehensive analysis of gene regulatory pathways in 2- and 3-dimensional cell culture compared to that of untreated active TB from clinical samples to identify common signatures shared between systems.
3. To apply host-directed therapy (HDT) to key regulatory targets identified by bioinformatic analysis using a human model of TB. HDT may improve host survival during Mtb infection and therefore directly translate to improving clinical outcomes in Mtb infection.

Chapter 2: Materials and methods

All protocols were carried out by me, unless otherwise stated.

2.1 Ethical approval

Analysis of blood from healthy donors was approved by the National Research Ethics Service committee South Central - Southampton A (Reference 13/SC/0043). All donors gave written informed consent. Histological analysis of biopsy samples was approved by the University of Southampton Ethics and Research Governance Online (ERGO) and Southampton Research Biorepository (Reference 12/NW/0794 SRB04_14). Lymph node biopsy tissue was taken as part of routine clinical care for diagnostic purposes, and surplus tissue analysed in this study. The ethics committee approved the analysis of this tissue without individual informed consent since the tissue was taken as part of routine care and excess tissue stored with consent.

2.2 Clinical sample selection

Clinical sample selection was from adult patients undergoing a mediastinal or neck lymph node biopsy at University Hospital Southampton NHS Foundation Trust. At the time the biopsy was obtained, they were formalin-fixed and paraffin-embedded (FFPE). The histology demonstrated by each specimen was considered typical for the disease identified, and was verified by a Consultant Histopathologist. All patient's medical records were thoroughly screened to ensure patients were non-smokers at the time the biopsy was taken. Importantly, the screening process also excluded patients who had received antituberculous treatment or any immunosuppressant therapy, including systemic corticosteroids or chemotherapy, prior to biopsy excision. Consequently, all biopsy samples in this study are from treatment naive patients. Given the number of library chips available to me for RNA sequencing (RNA-seq), it was planned for a total of 24 samples to undergo RNA-seq, hence 11 samples from each of the 3 clinical sample groups were chosen to provide an excess and optimise the chances of obtaining adequate RNA yield for RNA-seq.

2.2.1 TB sample selection

Initially, 78 patients with positive microbiological or histological evidence of TB between 2011 and 2016 from the Microbiology Department or Pathology Department respectively were identified. Those patients who had fully sensitive Mtb positive cultures together with mediastinal or neck lymph node biopsies demonstrating typical TB caseating granulomas were selected. Each patient's medical record was screened to ensure all TB patients had negative HIV serology, and no significant comorbidities. The selection process resulted in 4 mediastinal and 8 neck lymph node samples available for analysis. From each biopsy, 1 section was taken and the area in mm^2 approximated.

After taking digital photos of the section from each biopsy, the proportion of each section containing granulomas was visually estimated, and the 11 samples with the largest areas of granulomas were selected. From the 11 samples, the patients' ages ranged from 23 to 56 years old, with mean age 36 years old, and consisted of 6 males and 5 females.

2.2.2 Sarcoidosis sample selection

Similarly, 44 patients with lymph node or lung histology resembling sarcoidosis reported by a Consultant Histopathologist in 2013-2014 (the average dates the TB biopsies were taken) were identified. Of the 44 patients, 29 were reported as demonstrating classical sarcoidosis granulomas. Those patients whose biopsies were taken from mediastinal lymph nodes were selected and screened to ensure each patient had Mtb culture negative results and no positive HIV serology (HIV serology was negative or not performed). As a result, 16 mediastinal lymph node samples were available for analysis. One section was taken from each biopsy, and the area in mm^2 approximated. Again, digital photos of the sections were taken and the proportion of each section containing granulomas was visually estimated. The 11 samples with the largest areas of granulomas were selected. Of note, further retrospective analysis found comorbidities in 3 of the 11 patients where each of the 3 patients either had mild ulcerative colitis, chronic kidney disease stage IV, or diabetes mellitus type 2, though none of the patients had received immunosuppressant therapy. The 11 samples were taken from patients with ages 39 to 76 years old, mean age of 59 years old, and consisted of 4 males and 7 females.

2.2.3 Control sample selection

Choosing an appropriate control sample group was done after discussion and with guidance from a Consultant Histopathologist, Sanjay Jogai. Given that granulomas are associated with disease and not found in healthy tissue, the best control is a 'normal' lymph node, and selecting my normal tissue was a central challenge of this project. However, these 'normal' lymph node biopsies most often display 'reactive changes'. Reactive lymph nodes in the context of a healthy person was considered more abnormal than those of a person with a diagnosis to explain this finding. Lymph nodes as normal with reactive changes from either healthy patients or patients with newly diagnosed untreated cancer were chosen as the control group. To obtain the control group samples, the time period of 2013 to 2014 (the average dates the TB biopsies were taken) was used, resulting in 61 patients with mediastinal lymph node or lung histology reported to be 'normal'. An age cut-off of 75 years old was chosen to improve the chances of obtaining good quality RNA. Of the 46 remaining patients, their medical records were screened to ensure the biopsies were only from mediastinal lymph nodes, and no positive Mtb culture or HIV serology obtained from that patient. A total of 24 patients with normal reactive lymph nodes had available specimens. From each biopsy, 1

section was taken and the area in mm^2 approximated. Digital photos of a section from each biopsy were taken, the proportion of each section with normal histology visually estimated (almost 100%), and the 11 samples with the largest areas of normal tissue were selected. From the 11 samples, the patients' ages ranged from 26 to 70 years old, with mean age 57 years old, and consisted of 7 males and 4 females.

2.3 Laser Capture Microdissection (LCM) of clinical samples

2.3.1 Sectioning and dewaxing

To ensure RNase-free conditions, all instruments and tools were cleaned with RNase*Zap*™ RNase Decontamination Solution (RNaseZap RNase decontamination wipes, Cat. Number AM9786, Life Technologies). FFPE biopsy blocks were cut with a Leica RM2135 microtome. After discarding the first 20µm, sections of 10µm thickness were cut, floated in RNase-free water (UltraPure DNase/RNase-Free Distilled Water, Cat. Number 10977049, Life Technologies) at 45°C, mounted on to PEN (polyethylene naphthalate) membrane glass slides (LCM0522, Life Technologies), and dried in a drying oven at 37°C overnight. Sectioning was performed by the Histochemistry Research Unit at the University of Southampton. To dewax, the sections were twice immersed in Xylene (Arcturus Paradise Plus Staining components, Life Technologies) for 5 minutes at a time. No staining of tissue was found to be required. To remove the Xylene, sections were twice immersed in 100% EtOH (Cat. Number 10644795, Absolute 200 Proof, Molecular Biology Grade, Fisher BioReagents) for 1 minute at a time, then left to air dry for up to 1 hour. Dry times of longer than 1 hour resulted in reduced RNA yield downstream.

2.3.2 Laser Capture Microdissection

The laser capture microdissection (LCM) instrument (Zeiss PALM MicroBeam) settings were chosen by trial and error to dissect the dried 10µm sections with minimum damage to the tissue. An excess combination of Energy, Speed and Focus settings visibly disintegrated the tissue, whereas insufficient force generated by the settings did not penetrate the full thickness of each section. The optimised LCM settings were Energy 66, Speed 10%, and Focus 75. The dried sections underwent LCM, dissecting $8mm^2$ per biopsy (8.0 – 8.19, mean $8.09mm^2$), and samples captured in the inverted caps of tubes (Zeiss AdhesiveCap 500 opaque (415190-9201-000)). Two tubes per biopsy were used as the maximum area of dissected area each cap could effectively capture was found to be 4.8 mm^2. The tubes were removed from the LCM and 20µl digestion buffer placed in each tube. After closing the caps, the tubes were inverted again to expose dissected samples to digestion buffer. After 5 minutes exposure to digestion buffer, the tubes were reverted and pulse spun to draw the samples down from the cap. A further 20µl digestion buffer was added to each cap to loosen any samples adhered to the cap, and a second pulse spin done. A final 60µl digestion buffer (i.e. 100µl total final

volume) was added per tube, ensuring no samples stuck to the sides of the tubes. 4µl protease was gently mixed into each tube to disrupt the protein-protein and protein-nucleic acid crosslinks formed as a result of the formaldehyde fixation and embedding process. The tubes were then incubated in heat blocks at 50°C for 30 min, and then 80°C for 14 minutes to inactivate the protease. Each tube was subsequently frozen at -80°C. The digestion buffer and protease were from RecoverAll Total Nucleic Acid Isolation Kit for FFPE (AM1975), Invitrogen.

2.4 Total RNA extraction from clinical samples

The protocol was adapted from a protocol developed in the Histochemistry Research Unit at University Hospital Southampton NHS Foundation Trust. Each step was optimised to ensure the maximum RNA yield was later obtained from the clinical samples.

2.4.1 Materials for RNA isolation

Samples were thawed at room temperature in 4 batches each consisting of 8 to 9 samples. The following reagents (Isolation Additive, Wash 1 Concentrate, Wash 2/3 Concentrate) were from RecoverAll Total Nucleic Acid Isolation Kit for FFPE (AM1975, Invitrogen). Wash 1 Solution was made by adding 42ml 100% ethanol to Wash 1 Concentrate, and Wash 2/3 Solution prepared by adding 48ml 100% ethanol to Wash 2/3 concentrate, as per kit instructions. All Micro Filter Cartridges were from RNAqueous-Micro Total RNA Isolation Kit (AM1931, Invitrogen). All centrifugation was under the same settings of 30 seconds at 10,000 rcf (10,000 xg) at 19°C.

2.4.2 Method for RNA isolation

Isolation buffer was prepared by mixing 120µl Isolation Additive and 275µl 100% ethanol per tube. Isolation buffer was mixed with each sample and up to 500µl loaded on to a prepared Micro Filter Cartridge Assembly. Tubes from the same biopsy sample shared the same Micro Filter Cartridge Assembly. Each Micro Filter Cartridge Assembly was centrifuged to bind the RNA to the filter. Flow-through was discarded. Centrifugation was repeated if there remained any sample not yet loaded on to the Micro Filter Cartridge Assembly. 700µl of Wash 1 Solution was placed on each filter and underwent centrifugation to pass solution through the filter to clean the RNA bound to the membrane. Flow-through was discarded. 500µl of Wash 2/3 Solution was placed on each filter and centrifuged a third time, with flow-through discarded. A final centrifugation was done to remove residual fluid from the filter.

2.4.3 Materials for total RNA extraction

The following reagents (10X DNase Buffer, DNase) were from RecoverAll Total Nucleic Acid Isolation Kit for FFPE (AM1975), Invitrogen. Unless otherwise stated, all centrifugation was performed at 10,000 xg at 19°C.

2.4.4 DNase application

DNase mix was prepared using 6µl of 10X DNase Buffer, 4µl DNase, and 50µl Nuclease-free H_20 per sample. 60µl of DNase mix was placed in the centre of each Micro Filter Cartridge Assembly to degrade any contaminating DNA bound to the membrane or in the sample. Incubation proceeded for 30 minutes at room temperature. 700µl of Wash 1 Solution was added and incubated at room temperature for 30 seconds to clean the RNA bound to the filter. The samples underwent centrifugation for 30 seconds and flow-through discarded. 500µl of Wash 2/3 Solution was placed in the Micro Filter Cartridge Assembly, centrifugation for 30 seconds performed, and flow-through discarded. A further 500µl of Wash 2/3 Solution was added to the Micro Filter Cartridge Assembly and centrifugation for 30 seconds repeated, discarding any flow-through. A final centrifugation was then performed for 1 minute to remove any residual fluid from the filter.

2.4.5 Elution

Elution took place in new micro elution tubes to which the Micro Filter Cartridge Assembly was transferred. 12µl of RNase-free water (UltraPure DNase/RNase-Free Distilled Water, Cat. Number 10977049, Life Technologies) was placed in the centre of each Micro Filter Cartridge Assembly and left at room temperature for 5 minutes. To elute the RNA, the Micro Filter Cartridge Assembly was centrifuged for 1 minute. For each sample, 3 aliquots were made: 8µl stored at -80°C for subsequent RNA sequencing; 1µl kept on ice for immediate Qubit analysis; and 3µl at -80°C for subsequent Bioanalyzer analysis.

2.5 Clinical sample RNA extraction analysis

2.5.1 QubitTM Fluorometer

Qubit™ 3.0 Fluorometer with Qubit™ RNA HS Assay Kit (Cat. Number Q32852, ThermoFisher Scientific) was used to measure RNA concentration from clinical samples. A probe with extremely low fluorescence is used. Upon RNA binding, the probe becomes intensely fluorescent. Buffer mix was made using 1 in 200 dilution of probe in buffer. 199µl buffer mix was added to 1µl of each sample in an Eppendorf and left in darkness for 2 minutes at room temperature, then analysed using Qubit™ 3.0 Fluorometer.

2.5.2 Agilent 2100 Bioanalyzer

Agilent 2100 Bioanalyzer systems measured the RNA Integrity Number (RIN) and RNA concentration using picochips (RNA 6000 Pico kit, Cat. Number 5067-1513, Agilent Technologies) and nanochips (RNA 6000 Nano kit, Cat. Number 5067-1511, Agilent Technologies). Agilent Technologies specify the range of total RNA detected by the Pico kit is 50 – 5000pg/µl, whereas that detected by the Nano kit is 25 – 500ng/µl. Due to the clinical samples being formalin fixed, RNA yield was low (6.2 – 36,

mean 15.2 ng/µl). Therefore, depending on the RNA quantity detected by Qubit, each RNA clinical sample was diluted by 1 in 2 or 1 in 3, and quality measured using picochips. Fluorescent dye molecules are used to intercalate into nucleic acid strands and later detected by fluorescence. The Gel-Dye mix was prepared by adding 1µl dye to 65µl filtered gel, then 9µl loaded on to the chip using the chip priming station. 5µl Marker (25 nucleotide DNA fragment) was loaded on to every well of each chip to align the electropherogram. 1µl of heat denatured ladder was pipetted in to the appropriate well of the chip to act as a reference for RNA sizing and quantitation, and 9µl RNA conditioning solution into a different dedicated well in the picochip. Finally, 1µl of each sample was loaded into a sample well, vortexed for 1 minute in the IKA vortexer at 2400 rpm, and the chip run immediately in the Agilent 2100 Bioanalyzer instrument.

2.6 RNA sequencing from clinical samples

Sequencing of clinical samples was performed at the Kaminski laboratory, Department of Pulmonary, Critical Care, and Sleep Medicine, Yale School of Medicine, where I travelled to do this and subsequent analysis (funded by Scadding Morriston Davies travel award). The next generation sequencing approach used was Ion Torrent sequencing, a method based on ion semiconductor sequencing where the template DNA strand is flooded with a single type of deoxyribonucleotide triphosphate (dNTP) and, if complementary to the template nucleotide, a hydrogen ion is released and difference in pH detected.

2.6.1 Library preparation

Reagents were all purchased from Life Technologies. A total of 24 samples was prepared to make the cDNA libraries using Ion AmpliSeq™ Human Gene Expression kit (Ion Torrent™). The Ion PI™ Hi-Q™ Chef Reagents cartridge was thawed at room temperature for 45 minutes. Samples were diluted with nuclease-free water to 1:10 dilution to finish with 25µl solution per chip. Each chip can hold 4 samples, so 25µl from 4 diluted libraries was pipetted into an Ion Chef™ Library Sample Tube and stored on ice until loading on to the Ion Chef™ instrument. Library template construction using Ion PI™ Hi-Q™ Chef Kit (Cat. No. A27198) on Ion Chef™ instrument was performed, constructing 2 chips (from 8 diluted samples) at a time.

The Agilent 2100 Bioanalyzer system was used to repeat the RNA Integrity Number (RIN) and RNA concentration measurements using nanochips (RNA 6000 Nano kit, Cat. Number 5067-1511, Agilent Technologies).

2.6.2 Sequencing

The Ion Proton™ System was cleaned and initialised following on-screen prompts. dNTP stock solutions were thawed on ice. From the Ion Pi™ Controls 200 Kit, 32µl 10M sodium hydroxide was

added to Wash 1 Reagent Tube, and 45ml W3 Solution was added to Wash 3 Reagent Tube. Sippers were replaced, and Ion Proton™ Wash 2 Bottle rinsed twice with 18 MΩ water before filling it. 70µl of each nucleotide stock solution was pipetted into each corresponding Reagent Tube. The Ion Proton™ System sequenced 1 library chip at a time, and 2 library chips sequenced per day. Total RNA of 15ng per sample was used to produce 3 to 30 million (mean average of 18 million) single end 100 base pair reads per sample and transferred in FASTQ format.

2.7 Bioinformatics pipeline for clinical sample RNA sequencing analysis

The same optimised bioinformatics pipeline was used for analysis of RNA sequencing data from clinical samples as that from the microsphere samples. A detailed explanation for each stage of the pipeline is provided in Section 2.12.

2.7.1 Quality control

Quality control was performed on the raw sequenced data, which were in the form of FASTQ files. FastQC software was utilised to execute quality control checks including per base sequence quality, per sequence quality scores, per base sequence content, per sequence GC content, and sequence length distribution.

2.7.2 Alignment

Alignment was performed using kallisto software, with sequence based bias correction. Both the human transcriptome and the transcriptome of the Mtb strain H37Rv, used for infection of microspheres, were utilised as index transcriptomes to which alignment was performed.

2.7.3 Annotation

The tximport program was used to import the reads aligned by kallisto, and ensembldb program annotated them to gene level. Sleuth was also used to annotate the data aligned by kallisto, to confirm similar results.

2.7.4 Exploratory Data Analysis

Exploratory data analysis was undertaken to evaluate characteristics of RNA sequencing data from the clinical samples. Principal component analysis (PCA), and interquartile range (IQR) vs median analysis were performed.

2.7.5 Normalisation

Inter-sample normalisation was performed using TMM (Trimmed mean of M-value) normalisation. Given that the absolute expression levels and true lengths are not known for every gene, the relative expression levels are determined using an empirical formula. TMM normalisation uses a weighted trimmed mean of M values, where the M value is the log expression ratio for each gene in the library, and must assume the majority of genes are not differentially expressed.

2.7.6 Differential gene analysis

Differential gene analysis was performed using limma, with its voomWithQualityWeights function. The voom transformation is required for limma to use the Gaussian distribution, and converts raw counts to log counts per million (log CPM) values by extracting library sizes and normalisation factors from the raw data. Thereafter voomWithQualityWeights is able to estimate the mean-variance trend for log counts, and assign sample-specific weights to each observation based on their predicted variance. The weights are used in linear modelling to adjust for heteroscedasticity. A Benjamini-Hochberg [false discovery rate (FDR)]- adjusted P value of less than 0.05 was applied. Comparison was made between TB and control samples, sarcoidosis and control samples, and TB and sarcoidosis samples. Filter values were optimised for numbers of expressed genes across the study cohort in order to yield the highest number of differentially expressed genes, and the same filter value was applied for all comparisons. Batch effect was not observed on principle component analysis.

2.7.7 Correlation analysis

TMM normalised voom transformed data was taken forward for gene coexpression analysis. Graphia Professional (version 2.1) used the Markov Cluster Algorithm to perform correlation analysis with Pearson's $r \geq 0.83$, and MCL inflation value of 1.7, pre inflation value of 3, scheme value of 6, and smallest cluster allowed of 5 genes.

2.7.8 Gene Ontology

Gene ontology enrichment analyses for differentially expressed genes was performed in several ways: ToppFun functional enrichment in ToppGene suite (Chen *et al.*, 2009); REVIGO online tool (Supek *et al.*, 2011); g:Profiler online tool (Raudvere *et al.*, 2019); and ReactomePA software program (version 1.26.0, R) (Yu and He, 2016).

2.7.9 Gene Set Enrichment Analysis

Gene set enrichment analysis, also known as functional enrichment analysis, is a computational tool that identifies overrepresentation of specified gene sets instead of individual genes. Gene sets are groups of genes that share common biological function, chromosomal location, or regulation. Gene sets from the Molecular Signatures Database (MSigDB) (version 7.0) were used (Subramanian *et al.*, 2005). Two methods were employed: Fast gene sent enrichment analysis (fgsea) program (version 1.8.0, R) was used to generate normalised enrichment scores and display enrichment plots; and Ensemble of Gene Set Enrichment Analyses (EGSEA) used eleven prominent algorithms (camera, gage, globaltest, gsva, ora, padog, plage, roast, safe, ssgsea, zscore) to visualise pathways and generate fold change values identified by gene set enrichment analysis (version 1.10.1, R) (Alhamdoosh *et al.*, 2017).

2.8 Three-dimensional cell culture model

To make the microspheres, a number of initial preparatory steps were required: The matrix for the spheres, Mtb cultured to an optical density (OD) of 0.60, and peripheral blood mononuclear cells (PBMCs). Then all were combined to make the microspheres. All work was performed under sterile conditions in a Class II microbiological safety cabinet as outlined below. I was trained in 3-dimensional cell culture by Liku Tezera , and the work was carried out by me.

2.8.1 Preparation of 3% Alginate

Alginate was prepared by adding HBSS with Ca/Mg to sodium alginate powder (PRONOVA UP MVG, Cat. No. 4200106, high glucuronic acid content ≥60%, viscosity >200mPas, endotoxin ≤100 EU/g, Novamatrix) to a concentration of 3% alginate, vortexed for 3 minutes and placed on an orbital mixer at 10 xg overnight at 4°C.

2.8.2 Preparation of matrix

Two types of matrix were made: Alginate mix and Alginate-Collagen mix. The Alginate mix constituted 50% of the total mix volume. HEPES/NaOH buffer was made by adding 2.5ml 0.05M NaOH (Cat No 52770, Sigma) to 10ml 0.2M HEPES, then adding 37.5ml endotoxin free water to make 50ml NaOH/HEPES solution, of which the appropriate volume was added to the Alginate mix to constitute 4.5% of the total mix volume. $NaHCO_3$ 7.5% (Cat. No 25080-60, Life Technologies) was added to make up 9% of the total mix. The mixture of Alginate, HEPES/NaOH and $NaHCO_3$ was passed through a 0.22µm filter for sterilisation. For the final Alginate mix composition, HBSS was added to make 63.5% Alginate mix and 36.5% HBSS. For the final Alginate-Collagen mix composition, 3mg/ml Collagen was added to make 63.5% Alginate mix and 36.5% Collagen mix. Both matrix mixes were subsequently stored at 4°C.

2.8.3 Peripheral blood mononuclear cell (PBMC) isolation

PBMC isolation from single leucocyte donor cones (National Health Service Blood and Transfusion, Southampton, U.K) was performed. The blood donors are healthy donors who are negative for HIV and Hep B/C but of unknown TB exposure status, though from a low TB endemic area. 6 blood donors were used as biological replicates in the microsphere study. The bloods was diluted with 135ml warmed HBSS (37°C), pipetted on to Ficoll-Paque (GE Healthcare Life Sciences) in Falcon tubes maintaining a clear layer, and centrifuged at 480 xg for 30 minutes (Brake off, 21°C, Acceleration 5, Deceleration 1). The PBMC layer was transferred to a sterile Falcon tube and HBSS added to a total of 50mls. PBMCs were washed at 320 xg for 8 minutes (Brake on, 4°C, Acceleration 5, Deceleration 5) and supernatant discarded. The PBMC pellet was tapped to resuspend, and further HBSS added. A second wash was done at the same centrifugation settings, supernatant discarded, pellet

resuspended and HBSS added. The cell counting was performed before a final repeat wash. After the supernatant was discarded, the PBMC pellet was tapped to resuspend and RPMI (+ Ampicillin + Glutamine) was added for culture.

2.8.4 M.tuberculosis culture

Bioluminescent H37Rv (Andreu et al., 2010) was cultured in Middlebrook 7H9 medium supplemented with 10% ADC, 0.2% glycerol and 0.02% Tween 80 (BD Biosciences, Oxford) with kanamycin 25 μg/ml, as the luminescence was linked to a kanamycin resistant cassette. Cultures at 1 × 10^8 CFU/ml Mtb (OD = 0.6) was used for all experiments at a multiplicity of infection (MOI) of 0.1, except in pHrodo experiments where a MOI of 1 was used for monocyte infection. Luminescence was measured with GloMax® 20/20 single tube luminometer (Promega, UK).

2.8.5 PBMC infection with Mtb

PBMC infection took place in a Class II biosafety cabinet within the containment level 3 suite. After PBMC isolation, the appropriate volume containing 100 x 10^6 PBMCs was taken into each of 2 Falcon tubes to be transferred to two T75 flasks. One flask was incubated without infection and the other inoculated with H37Rv. Both flasks were cultured overnight in an incubator at 37°C and 5% CO_2.

2.8.6 Preparing the matrix for the bioelectrospray model

Preparation was performed under sterile conditions. After overnight incubation, PBMCs were pipetted from the 2 flasks (uninfected and infected) to respective Falcon tubes. Any remaining cells in the flasks were gently removed, first with HBSS, and then with 5ml Versene (Sigma) per flask for 10 minutes in the incubator at 37°C and 5% CO_2 followed by use of a scraper. HBSS was then added to dilute Versene, and flask contents transferred to Falcon tubes again. PBMCs were centrifuged at 700 xg for 8 minutes at 4°C. The supernatant was discarded and pellet volume measured. There were 3 conditions per pellet, so one third of each pellet was resuspended in RPMI supplemented with 10% Human Serum and incubated in T75 flasks at 37°C and 5% CO_2, one third mixed into a bijou containing 6ml of 1.5% Alginate mix, and one third mixed into a bijou containing 6ml of Alginate-Collagen mix. In summary, 6 conditions were created for each donor: 2D uninfected; 3D Alginate uninfected; 3D Alginate-Collagen uninfected; 2D Mtb; 3D Alginate Mtb; and 3D Alginate-Collagen Mtb. The final concentration of PBMCs in matrix was 5 x 10^6 per ml of matrix. The matrix was kept at 4°C until used in the bioelectrospray system.

2.8.7 Preparation of CaCl2 precipitation bath

In order for the matrix to form spheres after passing through the bioelectrospray needle (BES), the droplets of matrix must gel within a liquid bath of $CaCl_2$. To prepare the $CaCl_2$ mixture, 147g of $CaCl_2.2H_20$ was dissolved and 23.8g of HEPES added to 1 litre milli-Q H_20. The pH was adjusted to 5-

6 by adding $NaHCO_3$, and then sterilized using a 0.22µm filter. The $CaCl_2$ mixture was stored at room temperature. On the day of each BES experiment, 5ml $CaCl_2$ mixture was added to 45ml HBSS.

2.8.8 Cell encapsulation

The bioelectrospray (BES) model was used in a Class I biosafety cabinet (Figure 1). The BES needle was sterilised with 70% methylated spirit run through it and allowed to dry before running HBSS without Ca/Mg through to ensure there was good flow through the needle. A magnetic follower was placed in each $CaCl_2$ precipitation bath, and a bath placed beneath the BES needle. A sterile 5ml syringe drew up the matrix containing PBMCs from each bijou, attached to a plastic tube connected to the BES needle, and placed in a syringe driver. A potential difference of 7kV was applied between the needle and precipitation bath before bioelectrospraying began, and then the syringe driver run at 10ml/hr.

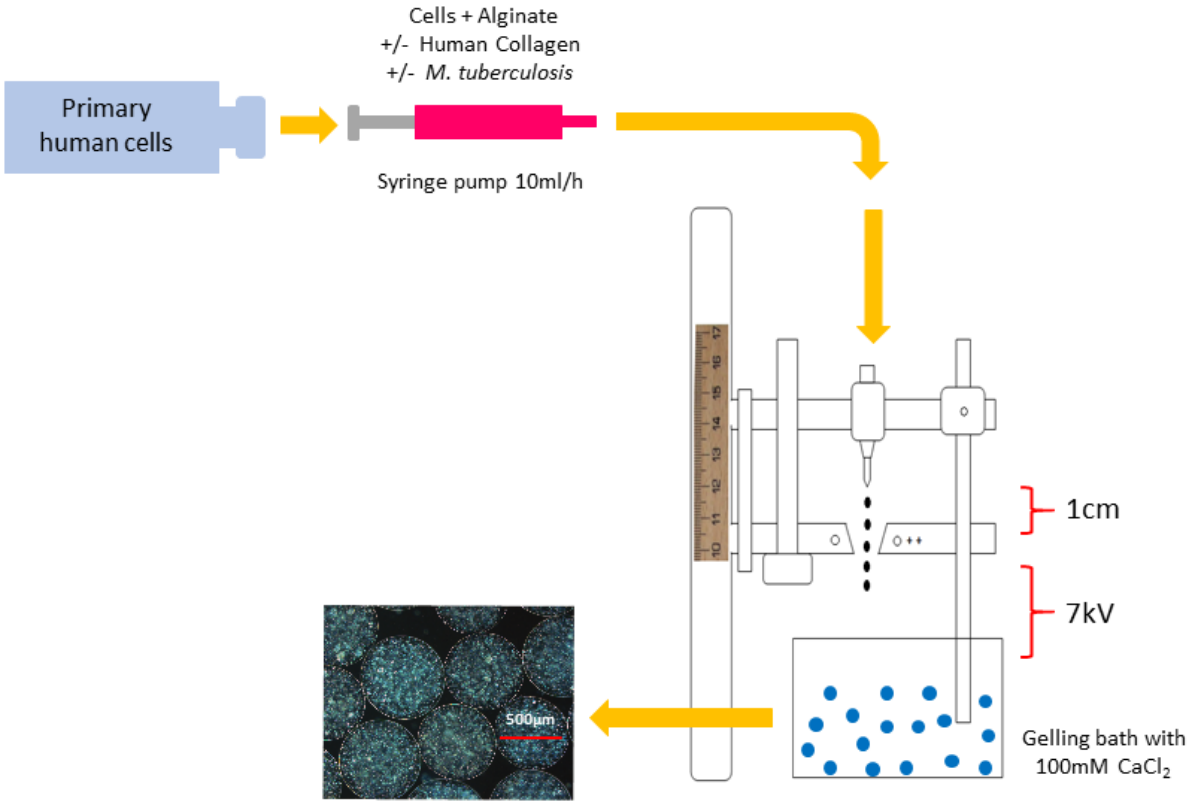


Figure 1: Bioelectrospray model of TB generates microspheres of peripheral blood mononuclear cell aggregates

Microspheres of approximately 500µm formed in the $CaCl_2$ bath, and were transferred to a Falcon tube. In a Class II biosafety cabinet, the microspheres were washed with HBSS with Ca/Mg, then 1ml suspension transferred to Eppendorf tubes for Mtb luminescence measurement (using Glomax® luminometer 20/20, Promega) at regular intervals over 14 days, while remaining microspheres were transferred to T75 flasks. Microspheres were incubated in RPMI 1640 media containing L-glutamine,

ampicillin, kanamycin and supplemented with 10% Human Serum added, at 37°C and 5% CO_2. Time points refer to days post-infection.

2.8.9 Microsphere decapsulation and cell lysis

Decapsulation was performed on day 4 in a Class I biosafety cabinet. Decapsulation solution was made by adding 55mM sodium citrate and 5mM EDTA to HBSS, adjusting the pH to 7.4 by adding 5M HCl. The decapsulation solution was warmed to 37^0C. Microspheres were transferred from the incubated flasks to Falcon tubes, and washed with HBSS. HBSS was removed and 30ml Decapsulation solution added. Once decapsulated, the cells were centrifuged at 2000 xg for 8 minutes at 4^0C. The PBMCs grown in 2D cell culture were treated as similarly to those grown in 3D cell culture (i.e. microspheres) as possible. The cells grown in 2D culture were transferred from the incubating flasks to Falcon tubes, and centrifuged at the same settings as the decapsulated cells. After centrifugation, the supernatant from both 2D cell culture and the decapsulated cells was removed. The cell pellets were resuspended in QIAzol Lysis Reagent (Cat. no. 79306, QIAGEN) to stabilise the cells by RNase inhibition, transferred to sterile Falcon tubes, and stored at -80°C.

2.9 Total RNA extraction from 2D and 3D cell culture

2.9.1 Phase separation

The samples were thawed on ice, and 200µl Chloroform added per 1ml of sample in QIAzol Lysis Reagent (Cat. no. 79306, QIAGEN). The samples were shaken vigorously for 15 seconds, then incubated for 3 minutes at room temperature. The samples were placed in Eppendorf tubes for centrifugation at 12000 xg for 15 minutes at 4^OC. The subsequent upper aqueous layer was transferred to new Falcon tubes. One volume of 70% Ethanol was added to each sample to enhance binding of RNA to the silica membrane of RNeasy Midi columns, shaking vigorously to mix.

2.9.2 RNA isolation

The following materials and reagents (RNeasy Midi columns, Buffer RW1, Buffer RPE) were from RNeasy® Midi Kit, Cat. No. 75144, QIAGEN. All centrifugation for RNA extraction and elution of these samples took place at 3400 xg at 20^OC. For each sample, up to 4ml was pipetted on to an RNeasy Midi column. The columns were centrifuged at 5 minutes, flow-through discarded, and centrifugation repeated for any remaining sample. A volume of 4ml Buffer RW1 was added to each column, centrifuged for 5 minutes and flow-through discarded. Then 2.5ml Buffer RPE was placed on to each column, centrifuged for 2 minutes, and flow-through discarded. A volume of 2.5ml Buffer RPE was again added to each column, then centrifuged for 5 minutes, and flow-through discarded.

2.9.3 Total RNA extraction

For elution of total RNA, 75µl RNase-free H_20 was initially incubated in each column for 5 minutes at room temperature. The columns were then centrifuged for 3 minutes, followed immediately by a final elution with 50µl RNase-free H_20 incubated for 5 minutes at room temperature. A final centrifugation for 3 minutes took place. Eluted samples were transferred into sterile Eppendorf tubes, immediately underwent NanoDrop analysis, and stored at -80°C for subsequent Bioanalyzer analysis and RNA sequencing.

2.10 Microsphere RNA extraction analysis

2.10.1 NanoDrop ND-1000 UV-Vis Spectrophotometer

NanoDrop ND-1000 UV-Vis Spectrophotometer (ThermoFisher Scientific) enabled RNA concentration to be measured. The arm of the spectrophotometer was opened, 1µl of RNA sample pipetted on to the optical pedestal, and the arm closed. On arm closure, the surface tension properties of the sample keeps the sample in place and allows the formation of a liquid column. Two fibre optic cables are brought into contact through the liquid sample column, and the height of the liquid column is controlled to 1mm and 0.2mm. A pulse of xenon is applied, the absorbance of sample measured, and RNA concentration calculated by the spectrophotometer software.

2.10.2 Agilent 2100 Bioanalyzer

Agilent 2100 Bioanalyzer systems measured the RNA Integrity Number (RIN) and RNA concentration using nanochips (RNA 6000 Nano kit, Cat. Number 5067-1511, Agilent Technologies). The method used is the same as applied for picochips described in *2.5.2* Agilent 2100 BioanalyzerSection 2.5.2, except that the RNA conditioning solution is not required when using nanochips.

2.11 RNA sequencing from microspheres

Sequencing was performed by Illumina HiSeq commercially performed by GENEWIZ, 115 Corporate Boulevard, South Plainfield, New Jersey. Sequencing was done in collaboration with the D'Armiento laboratory, Division of Pulmonary, Allergy and Critical Care Medicine, Columbia University, New York. All 36 samples were sequenced using Illumina HiSeq, with 12 samples per lane, and 3 lanes in total used. A configuration of 2 x 150 base pairs per lane was used. The output was 24 to 41 million (mean average of 31 million) paired end reads and provided in FASTQ format.

2.12 Bioinformatics pipeline optimization using microsphere sample RNA sequencing analysis

An emerging bioinformatic challenge is choosing the right bioinformatics pipeline to suit the samples being analysed and, importantly, the hypothesis being tested. The first available RNA sequencing

data, the microsphere study, was mapped first. The multiple pipelines employed for bioinformatic analysis of RNA sequencing data generated from the microsphere samples are illustrated (Figure 2). The development of pipelines was supervised by Dr Marta Polak, lead of the Systems Immunology Group in Medicine at the University of Southampton. Given the wide variation in results from different bioinformatic methodologies, the raw sequenced data was mapped twice using 2 different tools. The first was conventional alignment with STAR (Dobin *et al.*, 2013), a tool which maps the sequenced data to the genome. The second was pseudoalignment with kallisto, which uses kmers and maps them to the transcriptome (Bray *et al.*, 2016), theoretically providing results more reflective of the true biological processes taking place. A third aligner HiSat2 (Kim *et al.*, 2019) was also used to map 2 samples as a quality check.

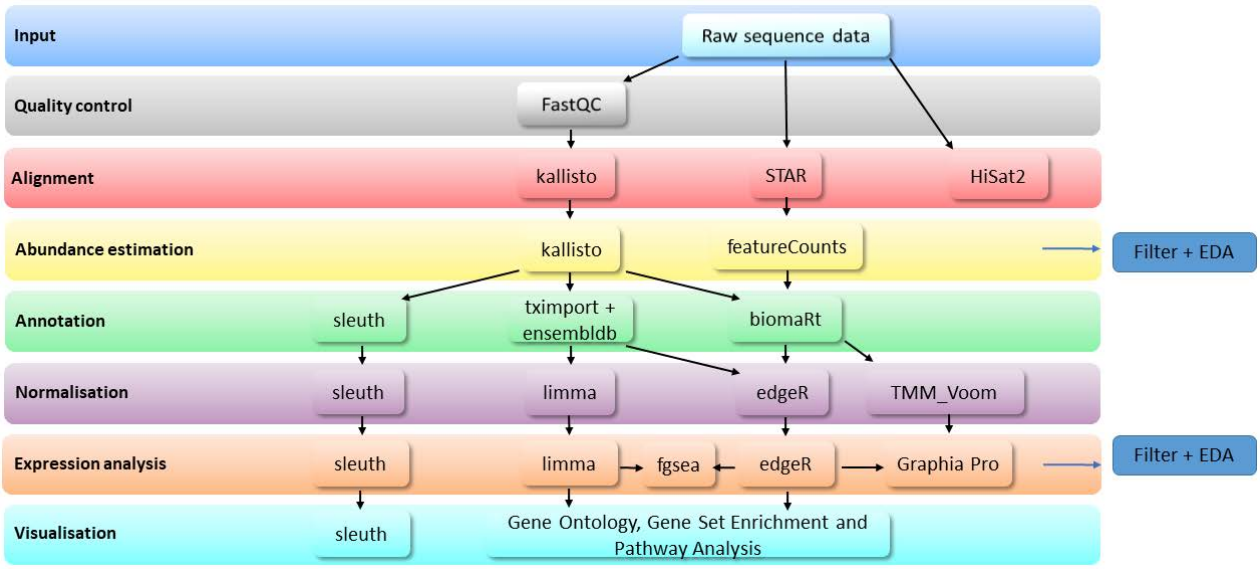


Figure 2: Bioinformatic pipelines used for optimisation of RNA sequencing data from three cell culture models

Kallisto mapping was chosen as a consensus best performing algorithm. One of the two University of Southampton's High Performance Computing facilities, Iridis 4, was utilised for the quality control and alignment of raw FASTQ files. Subsequent analyses were performed using R environment (version 3.5.2). The same optimised bioinformatics pipeline was then used to analyse the clinical sample RNA-seq data. The essential sections of code used are included for reference and provided in the Supplementary methods.

2.12.1 Quality control

Quality control checks of the raw sequenced data was performed using FastQC software. For further details see Section 2.7.1.

2.12.2 Alignment

Alignment was performed using 2 methods. The first was by STAR software, performed by GENEWIZ company. The second method was pseudoalignment using the kallisto program. For completion, both the human transcriptome and the transcriptome of the Mtb strain H37Rv, used for infection of microspheres, were utilised as index transcriptomes to which alignment was performed. Alignment was performed with sequence based bias correction.

2.12.3 Annotation

Annotation is required to label the output of mapped RNA sequencing (RNA-seq) data. Dedicated annotation tools were used to label the data with identifiable gene names: biomaRt (version 2.38.0, R), tximport (version 1.10.1, R), ensembldb (version 2.6.8, R), and sleuth (version 0.30.0, R). BiomaRt is a well established software package that labels aligned data with Ensembl terms (Durinck *et al.*, 2009). A similar program to biomaRt is ensembldb, which employs tximport software to import the mapped reads (Rainer, Gatto and Weichenberger, 2019). For the RNA-seq data aligned by STAR, biomaRt was the first available tool to me and was therefore used to perform annotation. For the RNA-seq data later aligned by kallisto, tximport (Soneson, Love and Robinson, 2015) was found to be quicker and more reliable than biomaRt. Therefore tximport was used to import the reads mapped by kallisto, and ensembldb annotated them to gene level. Since Sleuth is a program specifically developed to analyse RNA-seq data whose transcript abundances have been quantified with kallisto, for comparison, sleuth was also used to annotate the data aligned by kallisto (Pimentel *et al.*, 2017).

2.12.4 Exploratory Data Analysis

Exploratory data analysis (EDA) was undertaken for the AllCounts data (sequencing data aligned using STAR and abundance estimated using featureCounts software, performed by GENEWIZ). Principal component analysis (PCA), and interquartile range (IQR) vs median analysis were performed.

2.12.5 Normalisation

Normalisation is required to account for a number of factors including library size and gene length, and is a critical part of ensuring robust data analysis. TMM (Trimmed mean of M-value) normalisation was used (edgeR, version 3.24.3, R) prior to voom transformation for limma (version 3.38.3, R).

2.12.6 Differential gene analysis

Differential gene analysis was performed using 3 packages: edgeR (version 3.24.3, R) (McCarthy, Chen and Smyth, 2012), limma (version 3.38.3, R) (Ritchie *et al.*, 2015), and sleuth (version 0.30.0, R) (Pimentel *et al.*, 2017). For data aligned by kallisto, the sleuth program incorporates normalisation, expression analysis and visualisation automatically after annotation. Filter values were optimised for

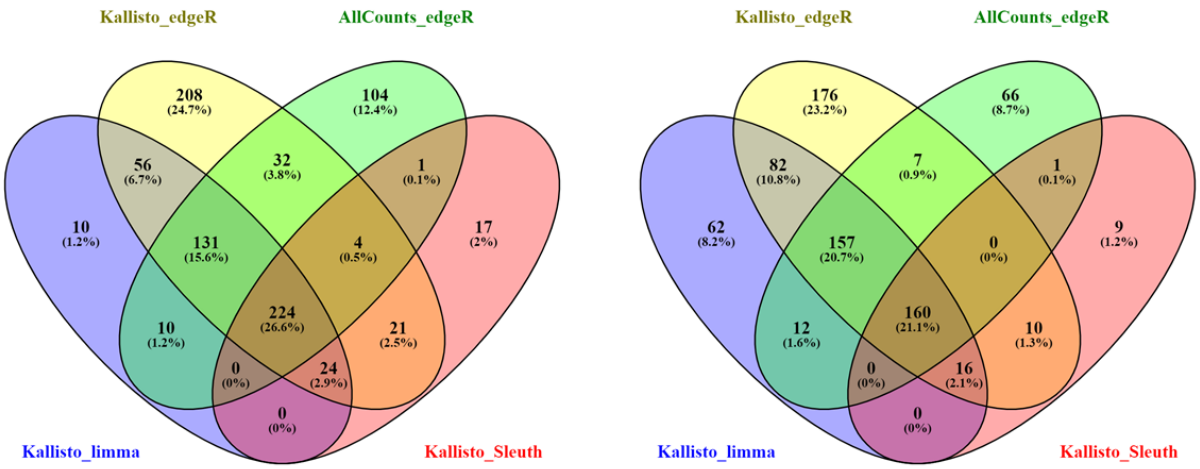
numbers of expressed genes across the study cohort in order to yield the highest number of differentially expressed genes, and the same filter value was applied for all comparisons. Batch effect was not observed on principle component analysis. Furthermore, any batch effect by donor was accounted for by incorporating the donor as a variable in the design matrix analysed by edgeR and limma.

2.12.7 Correlation analysis

TMM normalised voom transformed data was taken forward for gene coexpression analysis. Graphia Professional (version 2.1) used the Markov Cluster Algorithm to perform correlation analysis with Pearson's $r \geq 0.88$, and MCL inflation value of 1.7, pre inflation value of 3, scheme value of 6, and smallest cluster allowed of 5 genes.

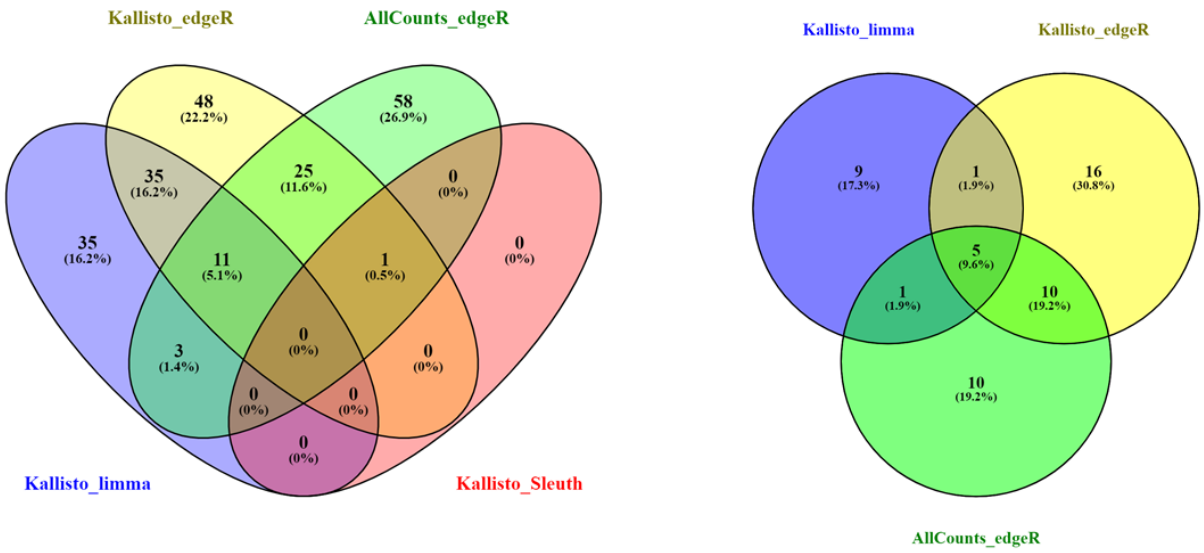
2.12.8 Bioinformatics pipeline comparison and optimisation

Given the diversity of results seen with different bioinformatics pipelines (Costa-Silva, Domingues and Lopes, 2017), I chose to use four bioinformatic pipelines to perform and compare gene expression analysis in each of the cell culture models (Figure 3, Figure 4, Figure 5). Kallisto output was analysed by sleuth ("Kallisto_Sleuth"), edgeR ("Kallisto_edgeR") and limma ("Kallisto_limma) for differential gene expression, while the AllCounts data was analysed by edgeR ("AllCounts_edgeR"). The edgeR package uses an overdispersed Poisson model. In contrast, limma and sleuth use the general linear model. A minimum absolute $\log_2$ fold change (FC) of 1.5 and Benjamini-Hochberg [false discovery rate (FDR)]-adjusted P value less than 0.05 was set for all bioinformatic analysis. The number of differentially expressed genes (DEGs) identified by each pipeline during Mtb infection, compared to uninfected cells, was analysed in each of the 3 experimental conditions: 2D (Figure 3); Alginate (Figure 4); and Alginate-Collagen (Figure 5).



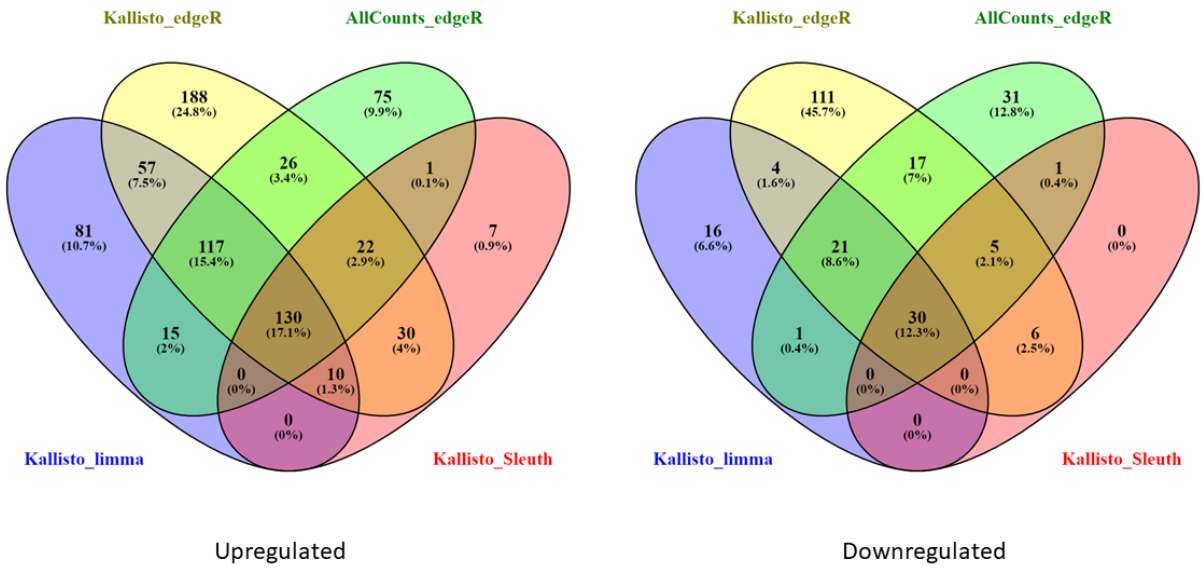
Upregulated Downregulated

Figure 3: Venn diagram showing the number of DEGs identified in the 2D model during Mtb infection, analysed by 4 pipelines (absolute $log_2FC > 1.5$, adjusted P value < 0.05). *The Kallisto_limma pipeline was selected for it's moderate level of stringency and ability to generate sample-specific weighting using voomWithQualityWeights function.*



Upregulated Downregulated

Figure 4: Venn diagram showing the number of DEGs identified in the 3D Alginate model during Mtb infection, analysed by 4 pipelines (absolute $log_2FC > 1.5$, adjusted P value < 0.05). *The Kallisto_limma pipeline was selected for it's moderate level of stringency and ability to generate sample-specific weighting using voomWithQualityWeights function.*



Upregulated Downregulated

Figure 5: Venn diagram showing the number of DEGs identified in the 3D Alginate-Collagen model during Mtb infection, analysed by 4 pipelines (absolute $log_2FC > 1.5$, adjusted P value < 0.05). *The Kallisto_limma pipeline was selected for it's moderate level of stringency and ability to generate sample-specific weighting using voomWithQualityWeights function.*

The most stringent of the 4 bioinformatics pipelines is Kallisto_Sleuth, exemplified by the 3D Alginate model whereby no genes are found to be downregulated (Figure 4). The most lenient of the bioinformatics pipelines is Kallisto_edgeR. In order to adjust the statistical model to the lesser weight of samples identified as possible outliers in exploratory data analysis, limma with its voomWithQualityWeights function was chosen to perform differential gene analysis by generating sample-specific weighting according to the variance of each individual sample within a normal distribution. Therefore samples with more heteroscedasticity are given less weighting.

In summary, the final bioinformatic pipeline chosen for all RNA sequencing data analysis was aligned using kallisto, reads imported by tximport, annotation performed by ensembldb, normalisation carried out using TMM (Trimmed Mean of M-value), transformation implemented with voomWithQualityWeights (limma version 3.38.3, R), and differential gene expression performed by limma.

2.12.9 Gene Ontology

As described in Section 2.7.8, gene ontology of differentially expressed genes was performed

2.12.10 Gene Set Enrichment Analysis

Gene set enrichment analysis was performed as described in *2.7.9 Gene Set Enrichment Analysis*.

2.13 Host-directed therapy experiments in the 3-dimensional Alginate-Collagen model

2.13.1 Microencapsulation of cells and incubation

As described in Section 2.8, PBMCs were infected overnight and then encapsulated using a bioelectrospray (BES) model. Once washed with HBSS with Ca/Mg, 1ml suspension was transferred to Eppendorf tubes for Mtb luminescence measurement (using Glomax® luminometer 20/20, Promega) at regular intervals over 14 days, and 250µl suspension was also transferred to each well of a 96-well plate for cytotoxicity studies on day 7. Microspheres in the Eppendorf tubes and 96-well plate were again incubated in RPMI 1640 media containing L-glutamine, ampicillin, kanamycin and supplemented with 10% Human Serum, at 37°C and 5% CO_2. Time points described refer to days post infection.

2.13.2 Chemical inhibitors and supplementation

Chemical compounds used to inhibit or activate selected targets were first added to the microsphere media just after microsphere generation and transfer to Eppendorf tubes on day 1. On day 7, Mtb luminescence readings were taken, 250µl supernatant removed without microsphere disturbance and frozen at -20°C for subsequent analysis, and replaced with the same chemical compound in 250µl RPMI 1640 containing L-glutamine, ampicillin, kanamycin and supplemented with 10% Human Serum added. The following chemical inhibitors were from ApexBio: Brinzolamide (Cat. No. A4359); 3-Bromopyruvic acid (Cat. No. B7922); GNE-7915 (Cat. No. B3605); KX2-391 (Cat. No. B2282); PF-543 (Cat. No. A3717); and Tipiracil (Cat. No. A3875). The following chemical inhibitors were from Bio-Techne Ltd: CHS 828 (Cat. No. 6753); GSK 583 (Cat. No. 6480); PF 915275 (Cat. No. 3291); SC 26196 (Cat. No. 4189); SID 26681509 (Cat. No. 3625); and SSM3 trifluoroacetate (Cat. No. 5253). K6PC-5 (Cat. No. SML1709) was from Sigma Aldrich. Both 3-Bromopyruvic acid and Tipiracil were reconstituted in water, and all other chemical reconstituted in DMSO. Chemicals were sterilised by filtration through a 0.22µM Durapore membrane (Merck Millipore).

2.13.3 Cell toxicity assays

Two cell toxicity assays were utilised. Both assays were performed in 96 well plates and measured using GloMax Discover (Promega). As described in *2.13.1 Microencapsulation of cells and incubation*, a 96 well plate containing microspheres was incubated for 7 days at 37°C, after which 100µl supernatant was transferred from each well to a second 96 well plate for LDH measurement. The original plate containing microspheres had background luminescence measured before utilising the first assay, CytoTox-Glo Cytotoxicity Assay (Promega), which measures the extracellular activity

of intracellular dead-cell protease using a luminogenic cell-impermeant peptide substrate AAF-Glo™ Substrate (AAF-aminoluciferin). Assay Reagent 50µl was added to each well of sample, shaken for 5 minutes by GloMax Discover, then luminescence read after 25 minutes to measure the initial levels of intracellular dead-cell protease. Lysis Reagent 50µl was then added to each well quickly, shaken for 5 minutes by GloMax Discover, and luminescence measured after 25 minutes for the final cytotoxicity readout.

The lactate dehydrogenase (LDH) assay was also performed on day 7, using the supernatant in the second 96 well plate, utilising a colorimetric assay in Cytotoxicity Detection Kit^{PLUS} (LDH) (Roche). Two repeat standard curves of LDH were made by serial dilution. Reaction mixture was made by adding 250µl Catalyst Solution to 11.25ml Dye Solution, and 100µl added to each well quickly. After leaving the plate in the dark for 15 minutes at room temperature, 50µl Stop solution was added to each well to halt LDH activity, and Glomax Discover used to shake the plate for 10 seconds and measure initial LDH levels. Then digitonin was added to all wells resulting in cell membrane lysis, and Glomax Discover used to again shake the plate and measure final LDH levels.

2.13.4 Monocyte isolation from human blood

Monocytes were isolated from human blood donors as per PBMC isolation described above. Thereafter, the PBMC pellets were resuspended in 15ml RPMI 1640 with ampicillin and L-glutamine. Monocytes were isolated using a hyperosmotic Percoll density gradient (Repnik, Knezevic and Jeras, 2003), where the PBMC suspension was pipetted on to hyperosmotic Percoll solution, then centrifuged at 480 xg for 25 minutes (brake off). Monocytes were pipetted from the interface layer into a Falcon tube, washed by topping up to 50ml HBSS, and centrifuged for a further 10 minutes (brake on). Supernatant was discarded and monocyte pellet resuspended in 10ml RPMI + 2mM EDTA to prevent clumping and adherence to the tube, ready for intracellular pH staining. Monocytes were counted after staining with tryptan blue.

2.13.5 Intracellular pH indicator assays

Fresh monocytes were stained with pHrodo iFL Green STP Ester amine reactive dye (ThermoFisher Scientific) by making 10mM stock solution with DMSO and then diluting with RPMI and 2mM EDTA to give 10µM at pH 7.4. After isolation, monocytes were stained with pHrodo and incubated at 37°C for 30 minutes, before washing with RPMI and 2mM EDTA at 480 xg for 10 minutes (brake high), supernatant discarded, and monocyte pellet resuspended in RPMI 1640 with 10% human AB serum, ampicillin, and L-glutamine. Monocytes were transferred to a 96 well plate for overnight incubation at 37°C. Mtb infection and addition of compounds occurred in quick succession, and fluorescence readings were taken every 5 minutes for 40 minutes using GloMax Discover (Promega). A second

intracellular pH assay, LysoSensor Green DND-189 (ThermoFisher Scientific) was used with frozen monocytes. Monocytes were thawed and transferred to a 96 well plate for incubation at 37°C for 2 hours, then infected with Mtb for a further 1 hour at 37°C before the addition of compounds and LysoSensor 2µM in quick succession. Fluorescence readings were taken every 5 minutes for 40 minutes using GloMax Discover (Promega). Quadruplicate conditions and a minimum of three separate donors were used for each pH assay.

2.13.6 Luminex Analysis

Supernatant taken on day 7 and frozen was later thawed and sterilised by filtration through a 0.22µm Durapore membrane (Merck Millipore) to allow transfer out of the containment level 3 suite. Concentrations of cytokines (ThermoFisher, UK) and MMPs (R&D Systems) were measured using a Bioplex 200 platform (Bio-Rad, UK) according to the manufacturer's protocol and performed by Dr Liku Tezera.

2.13.7 Immunohistochemistry of formalin-fixed paraffin-embedded (FFPE) tissue

Immunohistochemical staining of FFPE lung tissue from TB and control patients was performed using 4µm sections, using a previously optimized anti-SphK1 antibody (ab61148, abcam), and the avidin-biotin-peroxidase technique. Immunohistochemical staining was performed by the Histochemistry Research Unit at the University of Southampton.

2.13.8 Statistics

Experiments were performed on at least three occasions using PBMCs from three separate donors with triplicate conditions. At least three technical replicates on each occasion were studied. Data presented are from a representative donor and include the mean and SD, and are consistent across donors. Analysis was performed in GraphPad Prism 8. Student's t test was used to compare pairs. ANOVA with Dunnett's correction for multiple comparisons was used for groups of 3 or more, where control was DMSO for all chemicals reconstituted in DMSO, and control was Mtb only for those reconstituted in water.

Chapter 3: Clinical study

3.1 Introduction

Tuberculosis (TB) caused by *M.tuberculosis* is a disease exclusive to humans. Therefore, it seems logical that discovering a novel host target to treat TB requires human material to be studied. TB is transmitted through inhalation of aerosol droplets and is primarily a disease of lung and lymph node tissue, resulting in the histological phenomena of caseating granulomas. Given that granuloma formation within lymph nodes is abnormal, two control groups were included in the study, as just comparing to normal lymph nodes might merely identify the gene expression profiles of the predominant cell types. The first comparator was another human granulomatous lung condition, sarcoidosis, known as a chronic inflammatory condition with unknown aetiology. The second clinical group was normal tissue from cancer patients away from the site of disease or healthy patients, but specifically sampling macrophage-rich areas of "reactive change". Consequently, the clinical study comprised of three groups: TB, sarcoidosis and control (Figure 6). Clinical sample selection is fully described in 2.2 Clinical sample selection.

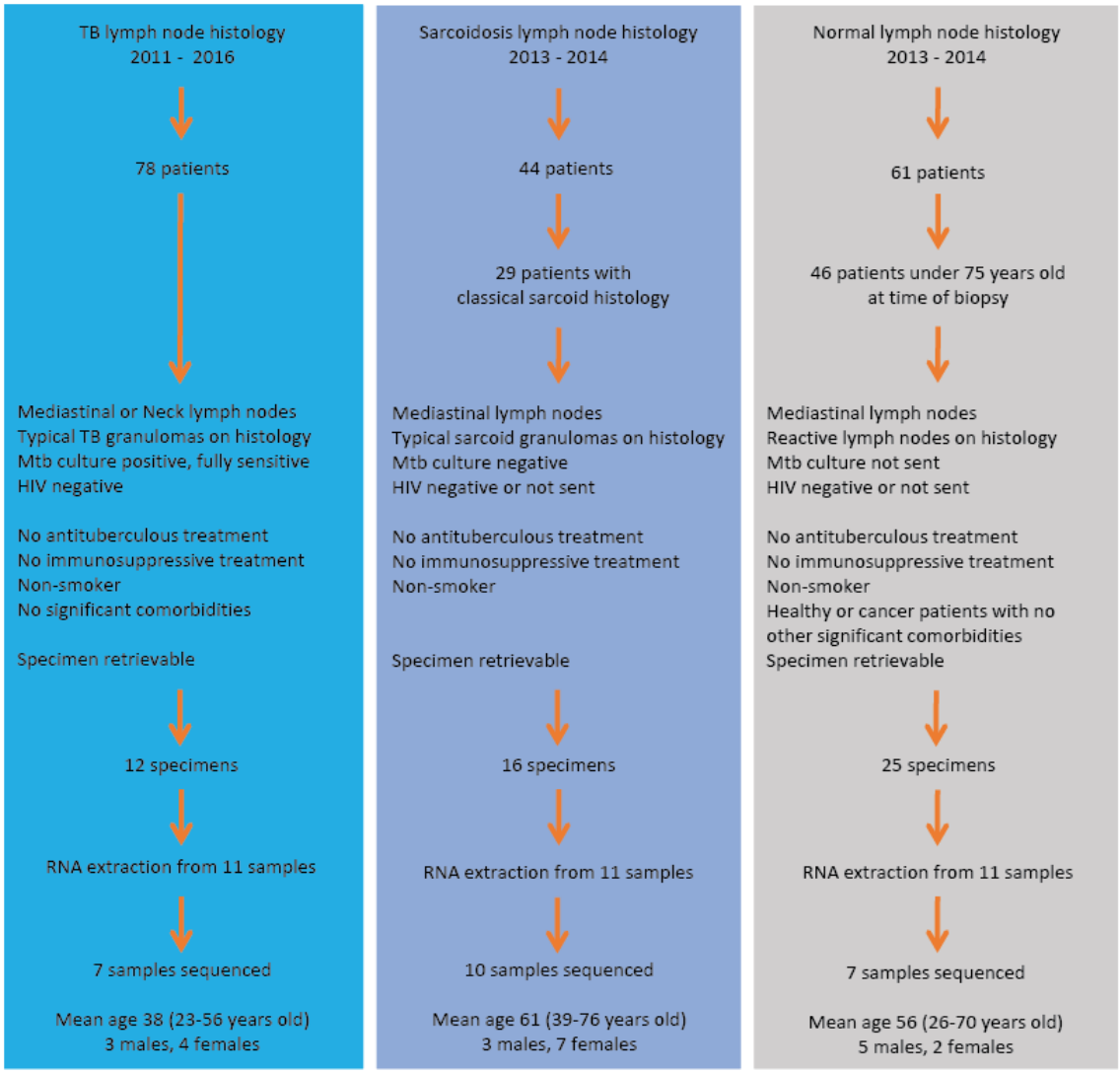


Figure 6: Study design of clinical lymph node samples from 3 groups (TB, sarcoidosis and control)

Formalin-fixed paraffin-embedded (FFPE) lymph node biopsies were identified. Granulomas were acquired using laser capture microdissection, and an equivalent surface area sampled from control lymph nodes, from eleven patient samples from each of the three groups. RNA was isolated from these samples and those with best RNA quality were selected for sequencing, resulting in 7 tuberculosis, 10 sarcoidosis and 7 control samples undergoing bioinformatic analysis.

3.2 Results

3.2.1 Clinical study RNA extraction and sequencing

Laser capture microdissection (LCM) took place on 11 different days, and RNA extraction on 2 different days. Qubit Fluorometer was used to measure RNA concentration, and the Agilent 2100 Bioanalyzer system used to measure the RNA Integrity Number (RIN). Examples of typical Agilent 2100 Bioanalyzer results are displayed below (Figure 7).

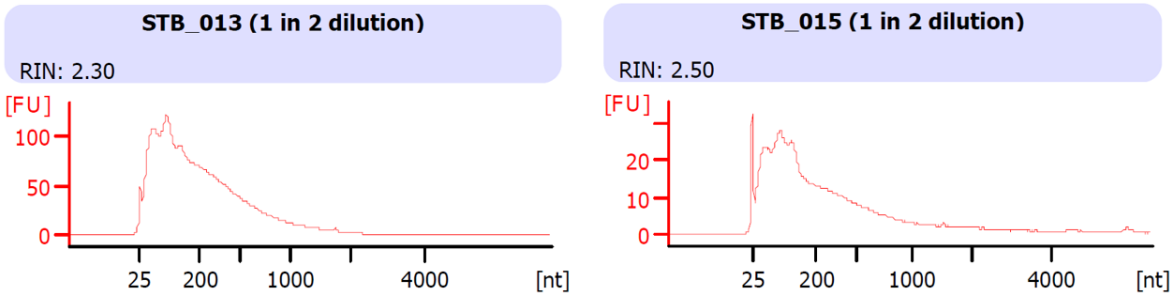


Figure 7: Agilent 2100 Bioanalyzer results for control sample STB_013 and TB sample STB_015. *All clinical samples were formalin-fixed paraffin embedded samples, resulting in low RIN measurements.*

All 33 samples were sent to the Kaminski laboratory for RNA analysis and subsequent sequencing (Table 1).

	Mean	Lowest	Highest
Total area dissected by LCM (mm^2)	8.09	8.00	8.19
Concentration (ng/µl)	15.22	6.20	36.00
RIN	2.28	2.10	2.60
cDNA concentration (pmol/L)	2598	369	7660

Table 1: RIN and concentration of RNA extracted from 33 clinical samples. *Low RINs and wide range of RNA concentration is seen despite similar areas of tissue dissection by LCM.*

Each sequencing chip could hold up to 4 samples, hence the top 24 samples with the highest cDNA concentration were selected for library preparation. Sequencing was then performed using Ion Torrent Sequencing to generate raw sequences in FASTQ format.

3.2.2 Quality control

FastQC analysis of the raw sequenced data demonstrated sequence lengths to be 25 to 391 base pairs, where the most frequent sequence length was found to be 100-119 base pairs in all samples. Sequences were of satisfactory quality with all average Phred scores of 24 to 25. Reads were single end reads, and the mean number of read counts was 18 million per sample (3 to 30 million read counts). The medium quality of Phred scores is a likely to be a reflection of the RNA quality obtainable from formalin-fixed paraffin-embedded (FFPE) tissue.

3.2.3 Kallisto alignment

The fraction of mapped reads was 47-85% (mean of 69%), and number of mapped reads was 2 to 21 million reads (mean of 13 million reads per sample). Similar to the above quality control analysis, the low number and fraction of mapped reads illustrates the lower quality of RNA extracted from formalin-fixed paraffin-embedded specimens. Using the human transcriptome as the reference

transcriptome, with and without the addition of the H37Rv transcriptome, gave exactly the same results, confirming all sequences mapped to the human transcriptome alone.

3.2.4 Transcript abundance and analysis

Within-sample normalisation was done by adjusting for transcript length (transcripts per million, TPM). The transcript abundance of the 24 clinical samples is shown below (Figure 8). The variable number of transcripts accounts for the wide range of imported mapped reads seen here.

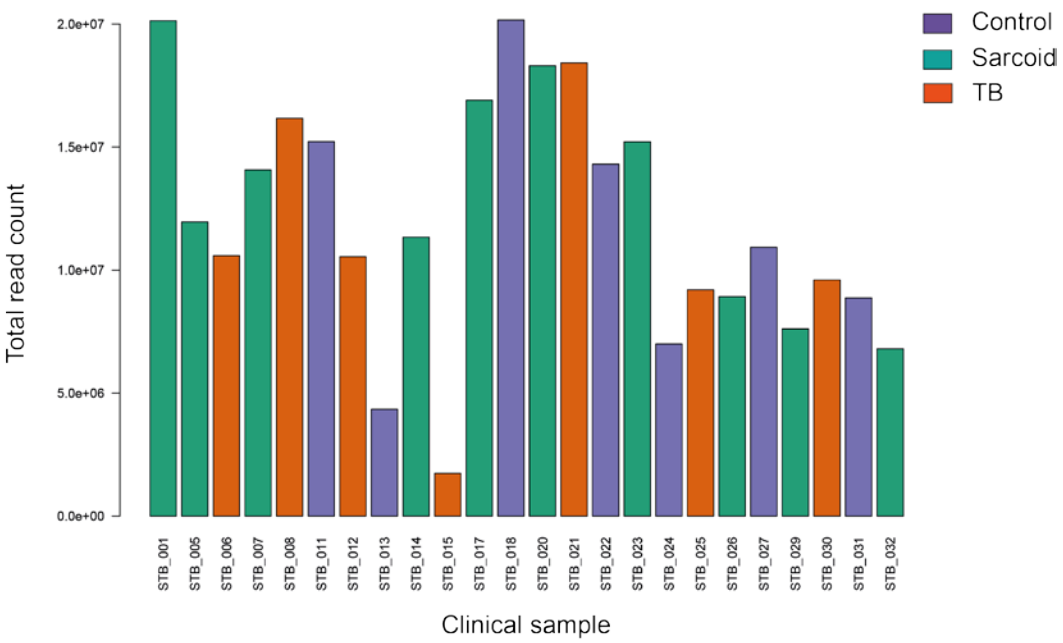


Figure 8: Transcript abundance of the 24 clinical study samples sequenced (transcripts per million). *Low read counts are seen in samples STB_013 and STB_015.*

Of note, the transcript abundance of samples STB_013 and STB_015 are below 5 million (TPM). Initial analysis included observing the data variance, whereby all clinical samples lay within 2 standard deviations of the median values for each sample (Figure 9).

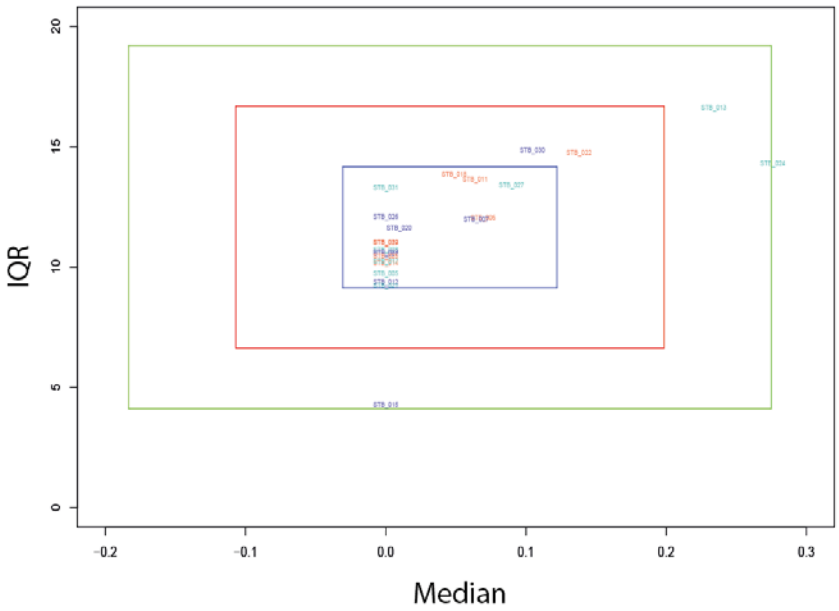


Figure 9: Interquartile range - median plot of clinical study samples. *Standard deviation thresholds of 1 (blue), 2 (red), and 3 (green) are used. No samples are identified as outliers.*

Another method to identify any outlying samples is principle component analysis (PCA) (Figure 10).

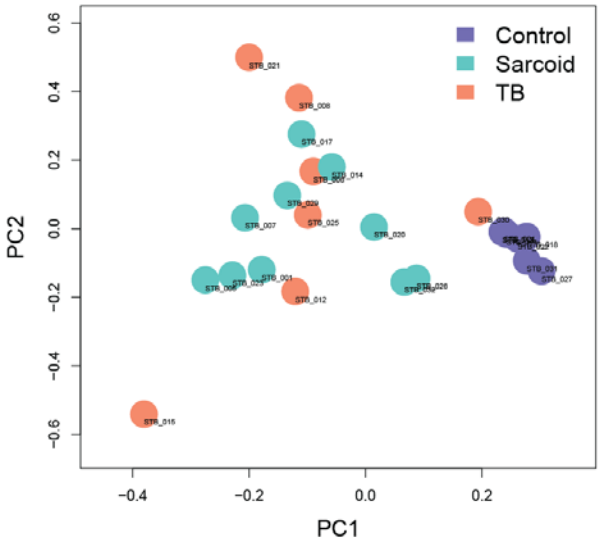


Figure 10: PCA of clinical study samples from TB, sarcoidosis and control patients. *TB and sarcoidosis overlap with each other and both separate from control samples.*

PCA demonstrates both TB and sarcoidosis samples clearly separate from the control group. Moreover, TB and sarcoidosis samples extensively overlap each other. STB_013 behaves similarly to the other samples in its group (control). The TB samples have a wide dispersion on PCA, which may be a reflection of the disease heterogeneity, and here STB_015 behaves no differently to the rest of the TB samples.

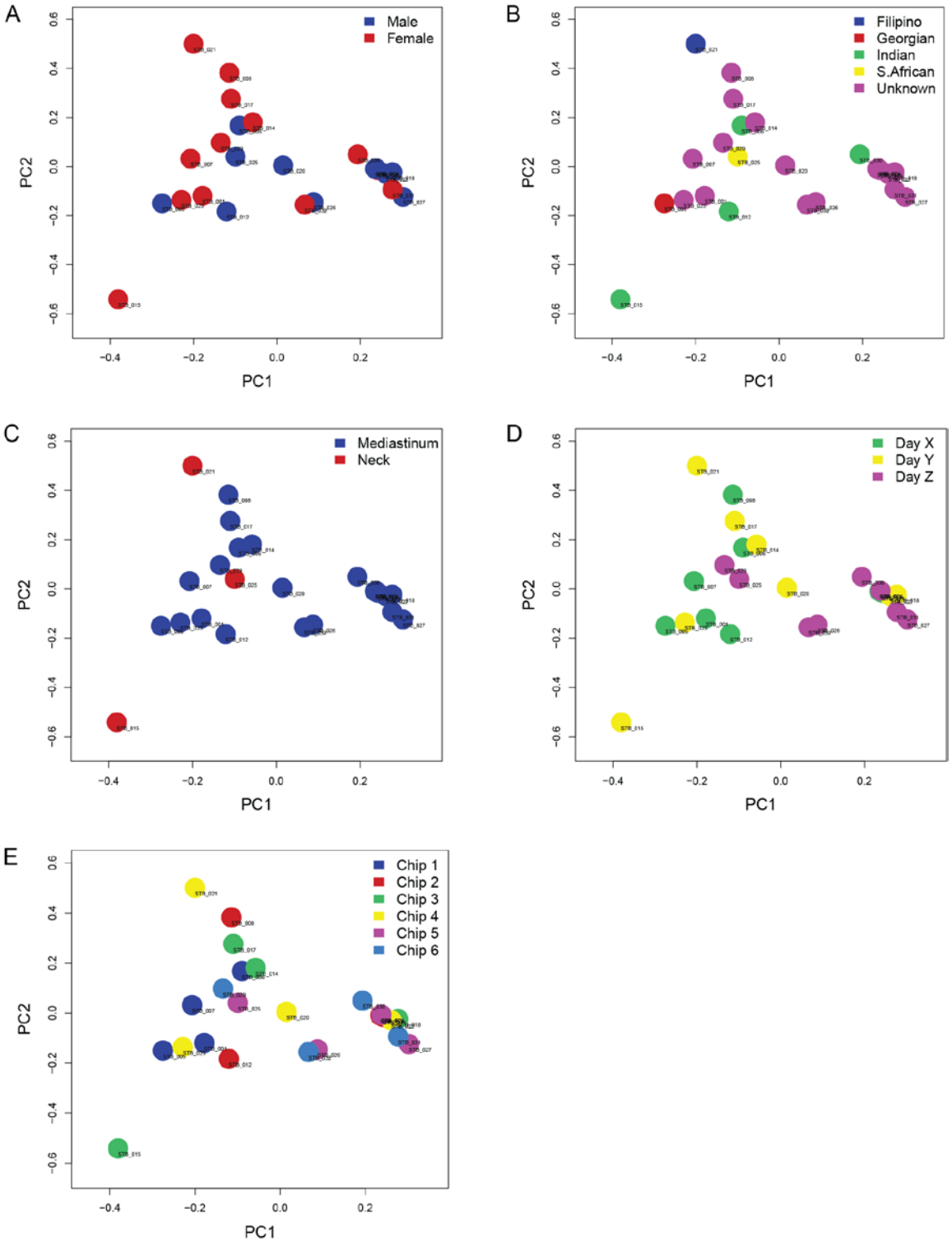


Figure 11: PCA of clinical samples grouped according to variables: Gender (A); Ethnicity (B); Location of biopsy site (C); Day of sequencing (D); and Sequencing chip (E). *Each variable shows an absence of batch effect.*

PCA of other measurable variables such as ethnicity, gender, site of lymph node biopsy, day of RNA sequencing, and RNA sequencing chip show no correlation (Figure 11), demonstrating no batch

effect.

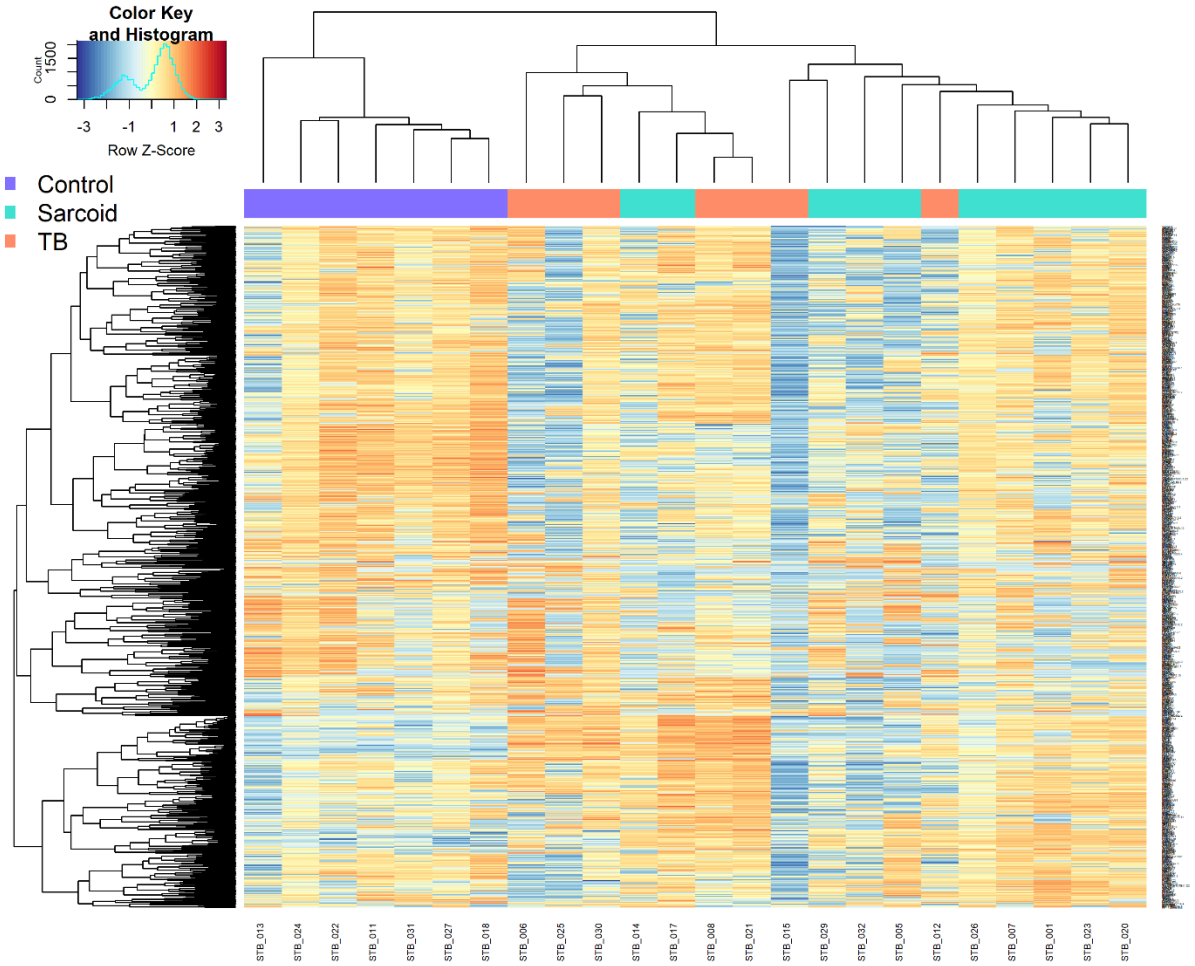


Figure 12: Hierarchical clustering heat map of top 1000 most variable genes across the 24 clinical samples (Spearman correlation and complete linkage). *TB and sarcoidosis cluster separately from control samples.*

Using a heat map with double dendrogram exhibits similar features to that seen on principal component analysis, whereby the control samples clearly cluster together whereas TB and sarcoidosis samples do not (Figure 12). STB_013 and STB_015 appear to behave differently to the rest of their group. However, omission of STB_013 and STB_015 on repeat heatmap analysis still demonstrates control samples to be the only group to cluster together (Figure 13).

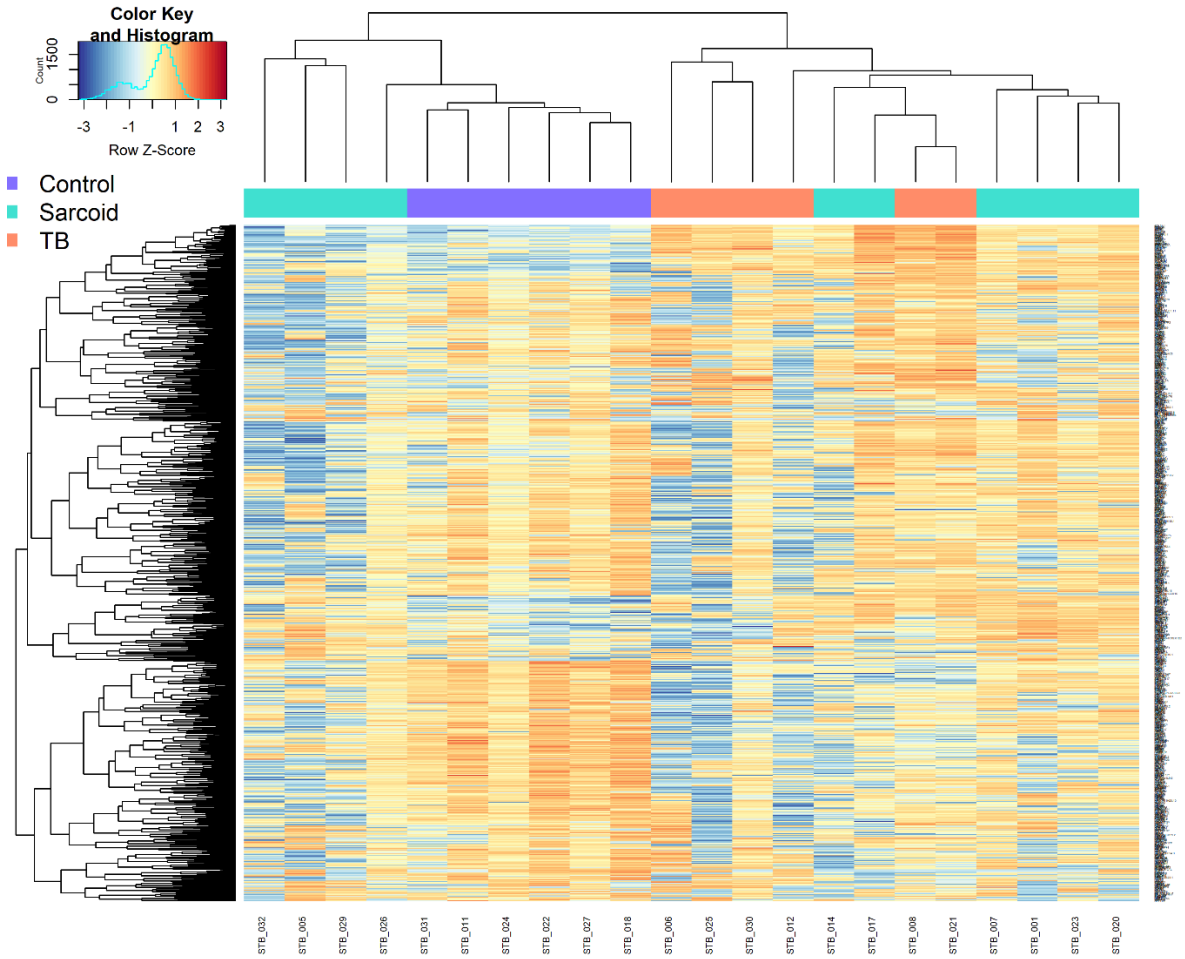


Figure 13: Hierarchical clustering heat map of top 1000 most variable genes after omission of 2 potential outlying clinical samples, STB_013 and STB_015 (Spearman correlation and complete linkage). *TB and sarcoidosis cluster separately from control samples.*

In case a low number of genes is responsible for the biological differences between TB, sarcoidosis and the control samples, the number of genes for heatmap analysis was reduced to 50. Again, the genes with most variance were analysed for all 24 clinical samples (Figure 14). The control samples again clustered together, in contrast to TB and sarcoidosis samples, thereby suggesting that it is not simply a low number of genes causing this difference.

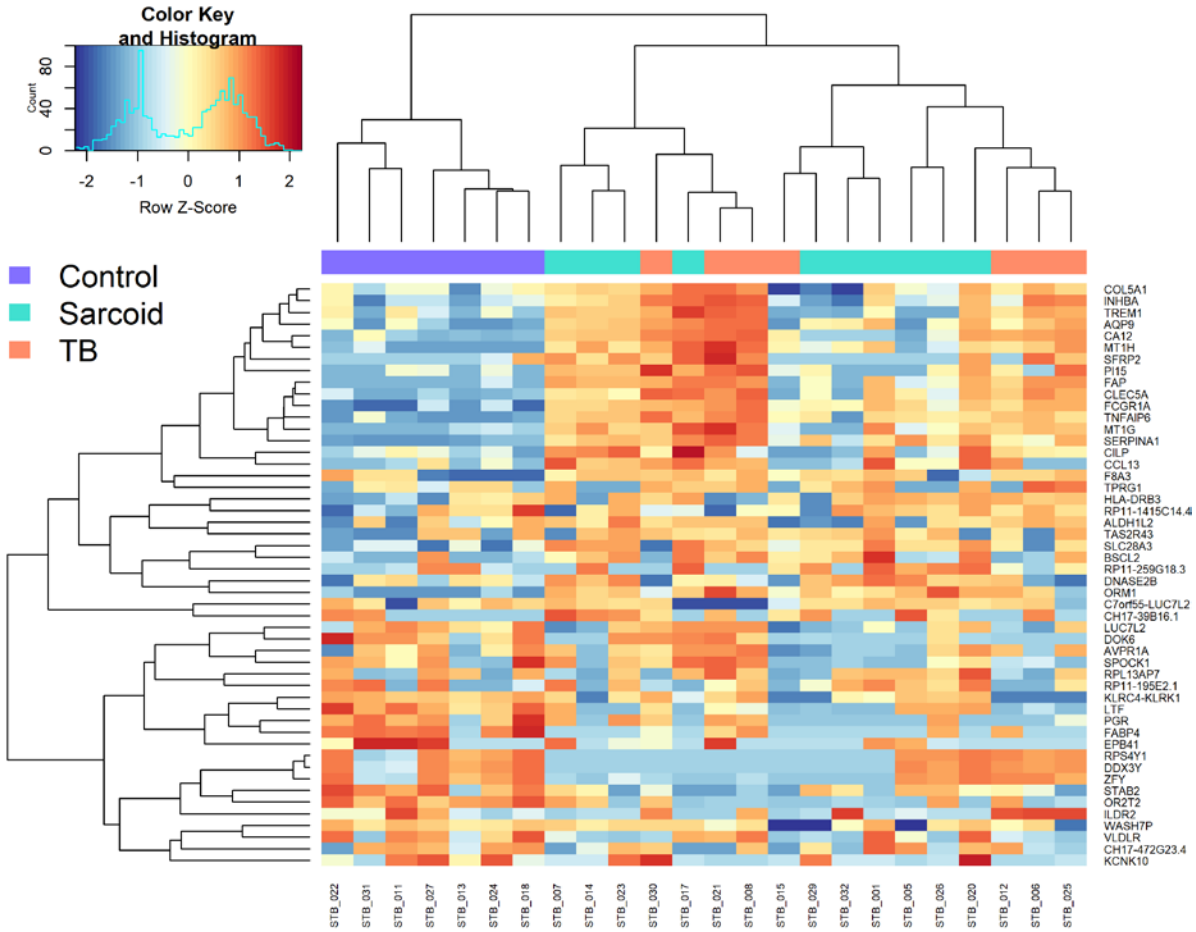


Figure 14: Hierarchical clustering heat map of top 50 most variable genes across the 24 clinical samples (Spearman correlation and complete linkage). *TB and sarcoidosis cluster separately from control samples.*

Finally, to identify any outliers that need removing from analysis, distribution of variance across the means of samples was observed (Figure 15) before normalisation, after between-sample normalisation, and after transformation to a Gaussian distribution (required for differential gene analysis by limma).

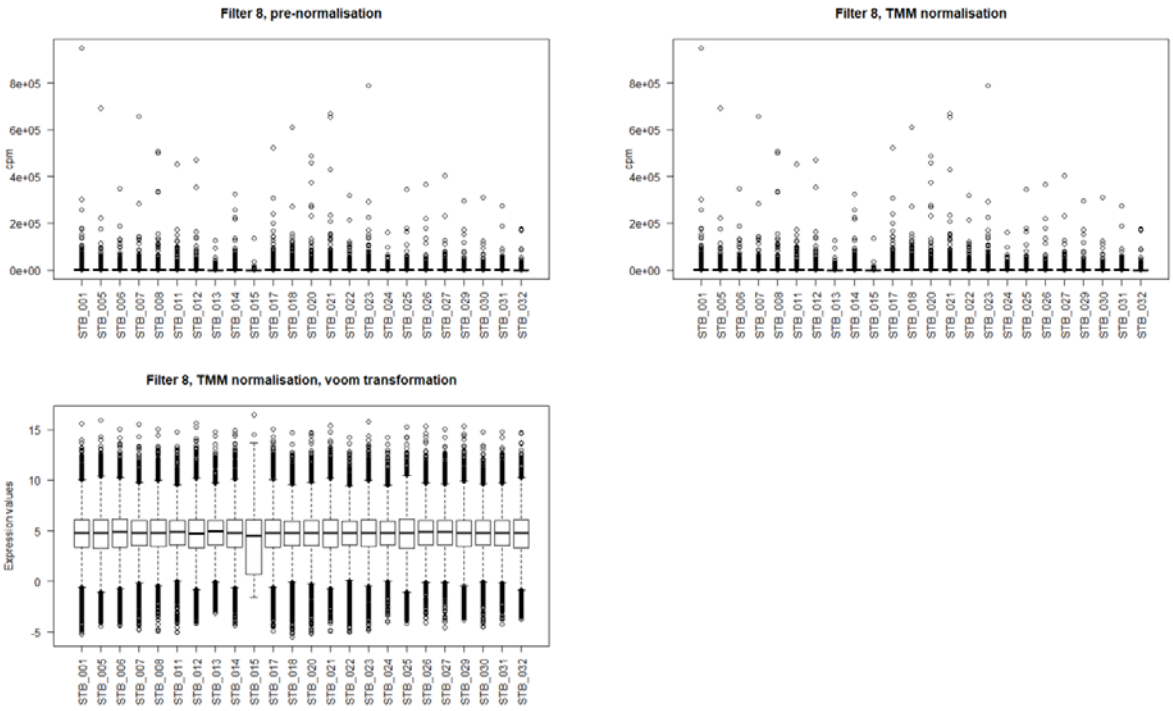


Figure 15: Boxplots of variance distribution across sample means before normalisation, after normalisation, and after voom transformation. *Mean values across samples remain unchanged after normalisation and voom transformation.*

Comparison of the clinical samples shows STB_015 has high variance though similar mean value after between-sample normalisation and voom transformation. In summary, STB_013 and STB_015 both lie within 3 standard deviations of the median, correlate with their respective groups on principle component analysis, do not significantly alter hierarchical clustering on heatmap analysis, and have similar mean values for differential gene analysis as that of the remaining samples. Therefore, STB_013 and STB_015 must be included and cannot be treated as outliers. Consequently, all 24 clinical samples were used for downstream analysis.

3.2.5 Differential gene analysis

Differential gene expression was performed by limma, and comparison made of TB and sarcoidosis samples in 3 steps. Firstly, differentially expressed genes (DEGs) in TB compared to control samples were identified. The same method was repeated for sarcoidosis compared to control samples. Lastly, the DEGs from each analysis were compared to determine which genes are significantly upregulated and downregulated in both TB and sarcoidosis when compared to control, and which DEGs are exclusive to TB or sarcoidosis when compared to control (Figure 16).

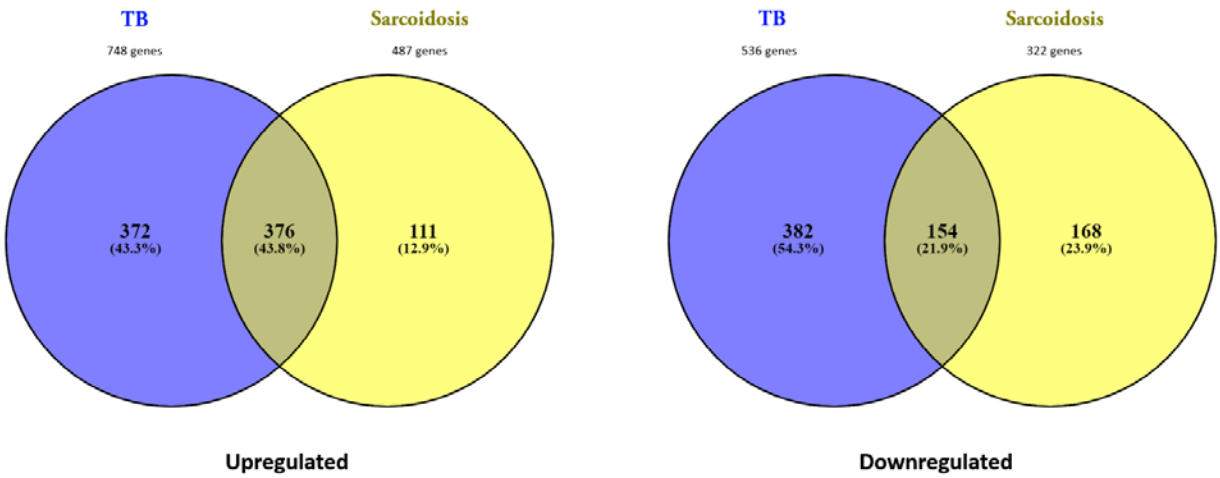


Figure 16: Venn diagram showing the number of DEGs identified on comparison of TB and sarcoidosis with control samples (absolute $log_2FC > 1.5$, adjusted P value < 0.05). *TB and sarcoidosis samples share 530 out of 1563 genes.*

A total of 1563 DEGs were identified compared to control samples. Notably, one third of all DEGs was shared by TB and sarcoidosis when compared to control. A further 754 DEGs were identified unique to TB, and 279 DEGs unique to sarcoidosis. It is important to bear in mind that despite a similar cell composition in the TB granuloma and sarcoidosis granuloma samples, a significant proportion of DEGs are exclusive to each disease, which will be highly important in identifying the biological processes and pathways that define TB and sarcoidosis. Similarly, the 530 DEGs communal to TB and sarcoidosis are crucial in determining the biological mechanisms related to granulomas and thus the role of the granuloma.

The fold changes of all genes is displayed in Figure 17, where TB and sarcoidosis were separately compared to control samples. Clinical TB samples display a higher number of DEGs with higher fold change and significance values than sarcoidosis samples.

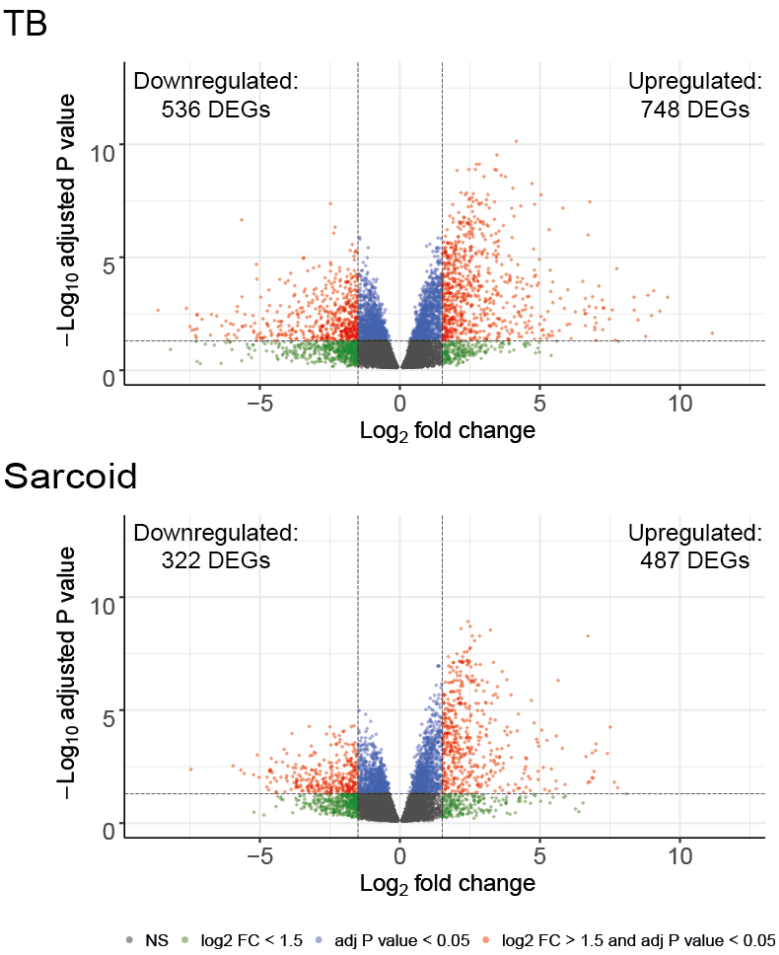


Figure 17: Volcano plots of gene fold changes observed during the comparison of TB and sarcoidosis with control samples. *TB shows more differential gene expression than sarcoidosis samples, demonstrated in higher number of DEGs and higher fold changes.*

Another method of displaying some of the most significant DEGs is in the form of a heat map. Based on fold change values, the top 10 significantly upregulated genes from 3 groups identified by Venn diagram (Figure 16) were taken: genes specific to TB; genes specific to sarcoidosis; and genes communal to TB and sarcoidosis (Figure 18). It is evident that a great proportion of genes are upregulated in both TB and sarcoidosis samples despite being identified as unique to TB. As expected, control samples do not share these upregulated DEGs with TB or sarcoidosis samples. Interestingly, *CCL7* (previously known as *MCP3*) and *MMP1* are the only upregulated genes exclusive to TB in Figure 18. Both genes are highlighted in correlation analysis (*3.2.6 Correlation analysis*).

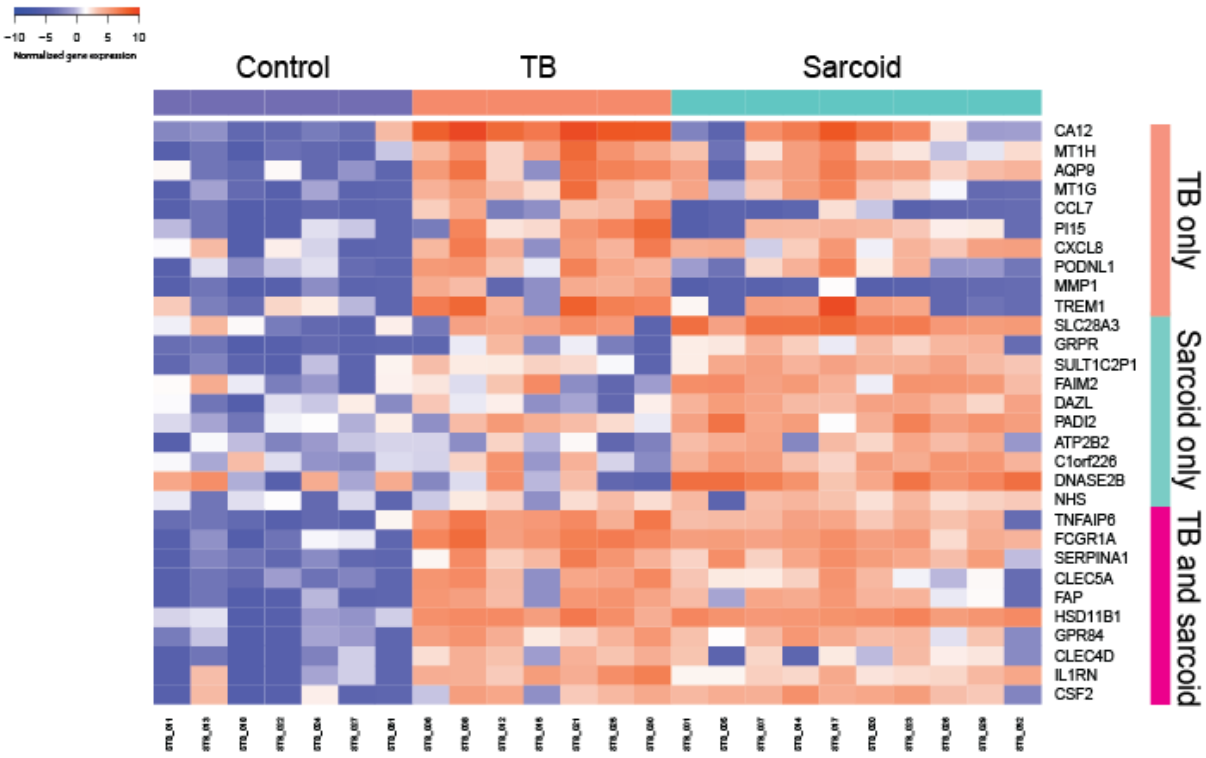


Figure 18: Heat map of clinical samples arranged according to the top 10 disease-specific DEGs based on fold change. *Utilising TB specific (orange), sarcoidosis specific (green) and jointly regulated genes (pink), a clear distinction between control, TB and sarcoidosis is observed, though CCL7 and MMP1 appear exclusive to TB.*

Further exploration of these upregulated genes reveal that the TB-specific DEGs are mainly located in the extracellular environment and are often metal binding, while sarcoidosis only DEGs are located intracellularly or at the plasma membrane and are often associated with neural development and protein synthesis. DEGs shared by both diseases are mainly located extracellularly and relate to the immune response or the extracellular matrix (Table 2).

	Gene symbol	Gene name	Genomic location	Main subcellular compartment	Related to metal binding	Function
TB unique	*CA12*	Carbonic Anhydrase 12	15 q22.2	Plasma membrane	Zinc metalloenzyme	Reversible hydration of carbon dioxide
	MT1H	Metallothionein 1H	16 q13	Extracellular, nucleus	Metallothionein	Metal ion transport
	AQP9	Aquaporin 9	15 q21.3	Plasma membrane	Binds arsenite	Water channel allowing passage of noncharged solutes
	MT1G	Metallothionein 1G	16 q13	Extracellular, nucleus	Metallothionein	Metal ion transport
	CCL7	C-C Motif Chemokine Ligand 7	17 q12	Extracellular	Substrate of MMP2 (zinc-dependent)	Encodes monocyte chemotactic protein 3
	PI15	Peptidase Inhibitor 15	8 q21.13	Extracellular	-	Encodes a trypsin inhibitor
	CXCL8	C-X-C Motif Chemokine Ligand 8	4 q13.3	Extracellular, nucleus	-	Encodes IL-8. Major mediator of the inflammatory response and potent angiogenic factor.
	PODNL1	Podocan Like 1	19 p13.12	Extracellular	-	Unknown function. Paralog of *PODN*, related to collagen binding.
	MMP1	Matrix Metalloproteinase 1	11 q22.2	Extracellular	Zinc-dependent	Interstitial collagenase
	TREM1	Triggering Receptor Expressed On Myeloid Cells 1	6 p21.1	Extracellular, plasma membrane	-	Mediates pro-inflammatory responses triggered by bacterial and fungal infection
Sarcoid unique	*SLC28A3*	Solute Carrier Family 28 Member 3	9 q21.32-q21.33	Endoplasmic reticulum, plasma membrane	Sodium-dependent	Nucleoside transporter involved in neurotransmission, vascular tone, and nucleoside homeostasis.
	GRPR	Gastrin Releasing Peptide Receptor	X p22.2	Plasma membrane	-	Regulates multiple functions of central nervous and gastrointestinal system
	SULT1C2P1	Sulfotransferase Family 1C Member 2 Pseudogene 1	2 q12.3	-	-	Pseudogene, unknown function
	FAIM2	Fas Apoptotic Inhibitory Molecule 2	12 q13.12	Plasma membrane	-	Regulates Fas-mediated apoptosis, particularly in neurons via caspase-8 activation.
	DAZL	Deleted In Azoospermia Like	3 p24.3	Nucleus	-	Encodes RNA-binding proteins essential for gametogenesis
	PADI2	Peptidyl Arginine Deiminase 2	1 p36.13	Cytosol, extracellular, lysosome, nucleus	Calcium-binding	Post-translational protein deimination. Substrates include myelin basic protein, and vimentin in skeletal muscle and macrophages.
	ATP2B2	ATPase Plasma Membrane Ca2+ Transporting 2	3 p25.3	Extracellular, plasma membrane	Magnesium-dependent	Intracellular calcium homeostasis
	C1orf226	Chromosome 1 Open Reading Frame 226	1 q23.3	-	-	Unknown function. Paralog of *NOS1AP*, encoding adaptor protein which binds neuronal nitric oxide synthase.

	DNASE2B	Deoxyribonuclease 2 Beta	1 p31.1-p22.3	Lysosome	-	Nuclear DNA degradation in salivary glands and lungs
	NHS	NHS Actin Remodeling Regulator	X p22.2-p22.13	Plasma membrane	-	Contains conserved nuclear localisation signals. Regulates actin remodelling and cell morphology in CNS development.
Shared by TB and sarcoid	*TNFAIP6*	TNF Alpha Induced Protein 6	2 q23.3	Extracellular	-	Extracellular matrix stability and cell migration. Elevated in synovial fluid of patients with osteoarthritis and rheumatoid arthritis.
	FCGR1A	Fc Fragment of IgG Receptor Ia	1 q21.2	Endosome, plasma membrane	-	High affinity receptor for Fc region of IgG in innate and adaptive immune response
	SERPINA1	Serpin Family A Member 1	14 q32.13	Endoplasmic reticulum, extracellular, golgi apparatus	-	Encodes alpha-1-antitrypsin
	CLEC5A	C-Type Lectin Domain Family 5 Member A	7 q34	Cytosol, plasma membrane	-	Cell adhesion, cell-cell signalling, regulation of inflammatory response.
	FAP	Fibroblast Activation Protein Alpha	2 q24.4	Plasma membrane	-	Extracellular matrix degradation in tissue remodelling, fibrosis, wound healing, inflammation, and tumour growth
	HSD11B1	Hydroxysteroid 11-Beta Dehydrogenase 1	1 q32.2	Endoplasmic reticulum	-	Converts cortisol to inactive cortisone
	GPR84	G Protein-Coupled Receptor 84	12 q13.13	Plasma membrane	-	Receptor for medium-chain free fatty acids with carbon chain lengths C9-14. Paralog of *HRH2*, encoding histamine receptor H2.
	CLEC4D	C-Type Lectin Domain Family 4 Member D	12 p13.31	Extracellular, plasma membrane	Calcium-dependent	Cell adhesion, cell-cell signalling, regulation of inflammatory response.
	IL1RN	Interleukin 1 Receptor Antagonist	2 q14.1	Cytosol, extracellular, plasma membrane	-	Inhibits interleukin-1
	CSF2	Colony Stimulating Factor 2	5 q31.1	Extracellular	-	Encodes Granulocyte-Macrophage Colony-Stimulating Factor

Table 2: Top 10 upregulated DEGs in TB specific, sarcoidosis specific, and jointly regulated genes, from highest to lowest fold change.

Interestingly, chromosomes X and 1 feature highly amongst the sarcoid unique genes. The cytogenetic band Xp22.2 is occupied by *GRPR* (Gastrin Releasing Peptide Receptor) and *NHS* (NHS Actin Remodeling Regulator) genes (Table 2).

3.2.6 Correlation analysis

An alternative method to perform gene correlation is Markov Clustering (MCL), where genes are clustered based on their direction and degree of change. Genes whose behaviour correlate with

each other form a cluster which can be illustrated using tools such as Graphia Pro. The 1563 differentially expressed genes (DEGs) identified by limma on comparison of TB with control and sarcoidosis with control samples (Figure 16), together with 189 DEGs identified using the same method on comparison of TB with sarcoidosis, underwent correlation analysis. In total, 1600 DEGs were analysed by MCL. A Pearson's correlation coefficient value of 0.83 was chosen because it produced the clearest depiction of gene correlation (Figure 19). Graphia Pro generated a total of 130 clusters, 28 of which displayed any gene expression values, and the top 25 of which are depicted below.

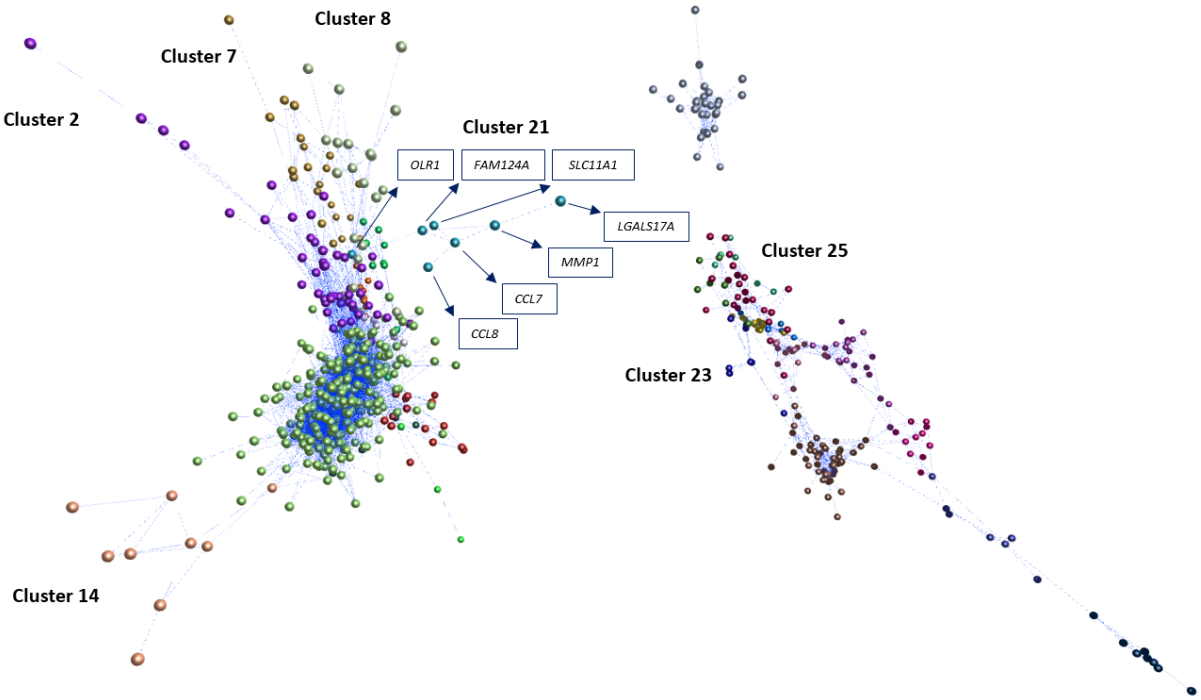


Figure 19: Correlation analysis performed utilising Markov Cluster Algorithm (Pearson's R of 0.83) with genes of absolute log2FC > 1.5 and adjusted P value < 0.05 during the comparison of TB with control samples, sarcoidosis with control samples, and TB with sarcoidosis samples, illustrated by Graphia Pro. *Each node (circle) depicts a transcript/gene, each edge (line) depicts Pearson's correlation value. The left branches display clusters upregulated in TB and sarcoidosis, and the right branches display clusters downregulated. Several co-regulated clusters are observed. Cluster 21 (blue) is the only cluster specific to TB and comprises the 7 annotated genes.*

The clusters seen on the left side of Figure 19 are composed of clusters with higher gene expression in all TB and most of sarcoidosis samples when compared to control. In contrast, clusters on the right side of (Figure 19) have lower gene expression in all TB and sarcoidosis samples when compared to control. In the top left of Figure 19, Cluster 2 (50 genes), Cluster 7 (22 genes) and Cluster 8 (19 genes) can be seen, and were found to be highly expressed to the greatest extent in TB, and to a still significant degree in sarcoidosis, compared to control (Figure 20). Moving down the left side of Figure 19, Cluster 21 (7 genes) is the only cluster to be highly expressed in TB samples alone compared to sarcoidosis and control. The middle left clusters in Figure 19 were highly expressed in both TB and sarcoidosis compared to control. The cluster found at the bottom left of Figure 19 is

Cluster 14 (9 genes), and is the only cluster upregulated more in sarcoidosis samples than TB and control. Conversely, Cluster 23 (7 genes) and Cluster 25 (6 genes) on the right side of Figure 19 were found to be underexpressed to the greatest extent in TB, and to a lesser degree in sarcoidosis, compared to control (Figure 20).

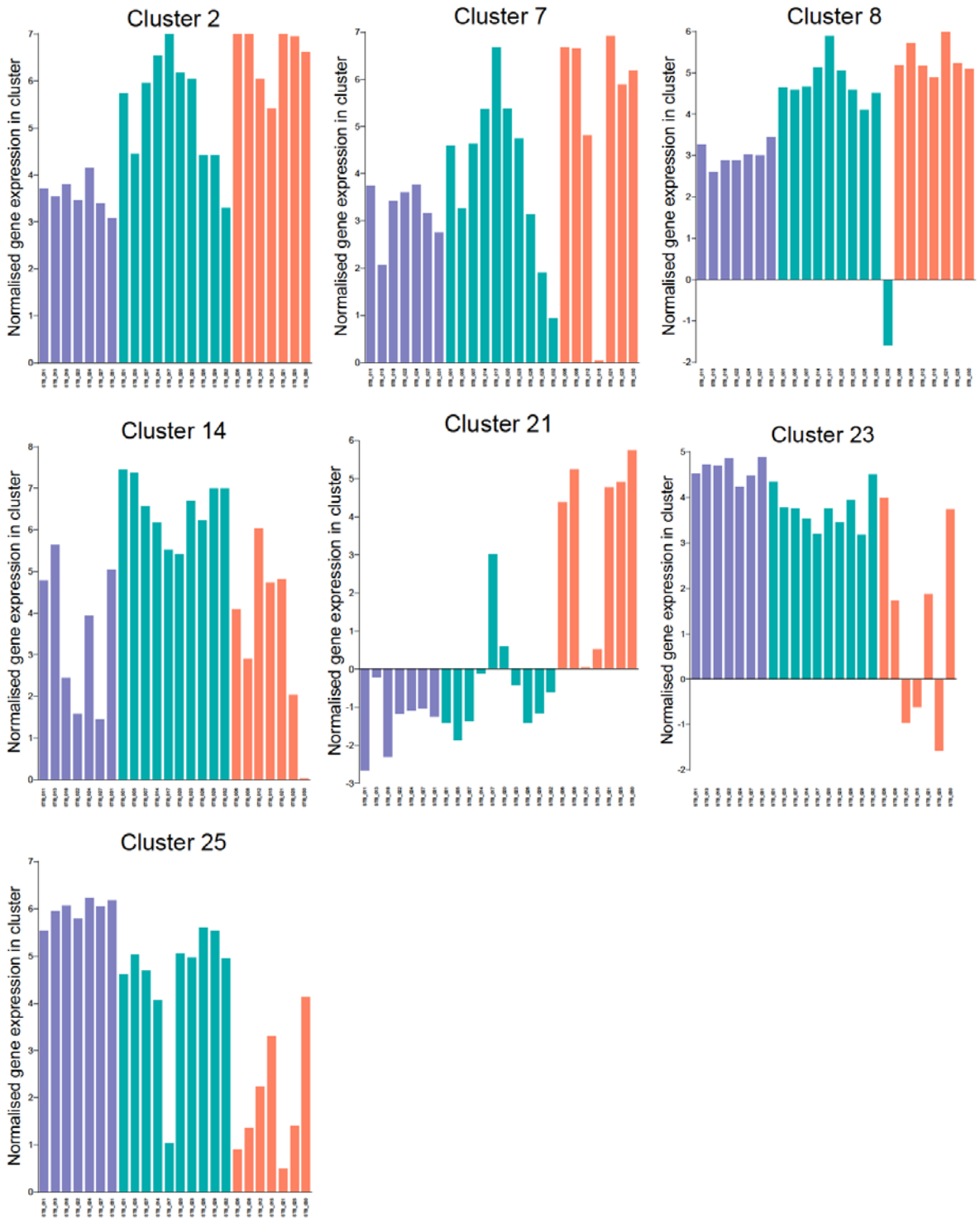


Figure 20: Normalised gene expression of 7 clusters of interest determined by Markov Clustering using Pearson's R of 0.83, analysed with Graphia Pro. *Control (purple), sarcoidosis (green) and TB (orange) samples are shown. Clusters 2, 7, 8, and 14 demonstrate joint upregulation of genes, whereas Clusters 23 and 25 are jointly downregulated. Cluster 21 is the only cluster specific to TB.*

Gene ontology (GO) analysis was carried out on each of the individual 28 clusters identified by Graphia Pro to have any gene expression value. Both ToppFun (false discovery rate cutoff of 0.05)

and g:Profiler (adjusted P value threshold of 0.05) tools were utilised for GO analysis. For each of the 7 clusters of interest (Clusters 2, 7, 8, 14, 21, 23 and 25), the top 10 biological processes taken from GO analysis are observed.

Cluster 2 is summarised in Table 3, with low false discovery rates (FDR) observed. Other biological processes included blood vessel development (GO:0001568, FDR B&H 3.55 x 10^{-4}) and stress response to copper ion (GO:1990169, FDR B&H 4.40 x 10^{-4}). Overall, Cluster 2 demonstrates the most significant processes highly upregulated in TB compared to control, and to a lesser extent in sarcoidosis compared to control, relate to the extracellular matrix and growth.

ID	Name	FDR B&H	Genes from input	Genes in annotation
GO:0030198	extracellular matrix organization	5.92E-09	13	399
GO:0043062	extracellular structure organization	6.48E-08	13	511
GO:0070208	protein heterotrimerization	4.19E-05	4	18
GO:0040007	growth	9.58E-05	14	1182
GO:0001501	skeletal system development	1.53E-04	10	582
GO:0001503	ossification	1.53E-04	9	444
GO:0034097	response to cytokine	1.53E-04	14	1297
GO:0071230	cellular response to amino acid stimulus	1.53E-04	5	75
GO:0009611	response to wounding	1.53E-04	11	756
GO:0010035	response to inorganic substance	2.44E-04	10	638

Table 3: Top 10 biological processes identified on GO analysis of Cluster 2 using ToppFun. *Biological processes overexpressed in TB, and in sarcoidosis to a lesser extent, mainly associate with the extracellular matrix and cell growth.*

Similarly, GO analysis of Cluster 7 was performed (Table 4). Other biological processes included lymphangiogenesis (GO: 0001946, FDR B&H 1.17 x 10^{-2}). Cluster 7 demonstrates the most significant processes highly upregulated in TB compared to control, and to a lesser degree in sarcoidosis compared to control, relate to the extracellular matrix and wound healing.

ID	Name	FDR B&H	Genes from input	Genes in annotation
GO:0030198	extracellular matrix organization	1.19E-10	11	399
GO:0043062	extracellular structure organization	8.80E-10	11	511
GO:0090505	epiboly involved in wound healing	2.21E-03	3	37
GO:0044319	wound healing, spreading of cells	2.21E-03	3	37
GO:0090804	epiboly	2.21E-03	3	38
GO:0060429	epithelium development	2.21E-03	9	1483
GO:0007155	cell adhesion	2.21E-03	9	1509
GO:0022610	biological adhesion	2.21E-03	9	1516
GO:0035987	endodermal cell differentiation	2.21E-03	3	45
GO:0048729	tissue morphogenesis	3.15E-03	7	861

Table 4: Top 10 biological processes identified on GO analysis of Cluster 7 using ToppFun. *Biological processes overexpressed in TB, and in sarcoidosis to a lesser extent, associate with the extracellular matrix and wound healing.*

GO analysis of Cluster 8 is summarised in Table 5, with larger FDR values than those found in Clusters 2 and 7. Other biological processes identified were related to cellular metabolism such as carboxylic acid metabolic process (GO:0019752, FDR B&H 2.67 x 10^{-2}), pyruvate metabolic process (GO:0006090, FDR B&H 2.94 x 10^{-2}), carbohydrate derivative metabolic process (GO:1901135, FDR B&H 3.32 x 10^{-2}), and medium-chain fatty-acyl-CoA catabolic process (GO:0036114, FDR B&H 3.32 x 10^{-2}). Overall, Cluster 8 shows the biological processes highly upregulated in TB compared to control, and to a lesser extent in sarcoidosis compared to control, are mainly metabolic processes, particularly nucleotide metabolism.

ID	Name	FDR B&H	Genes from input	Genes in annotation
GO:0009117	nucleotide metabolic process	1.74E-02	6	694
GO:0006753	nucleoside phosphate metabolic process	1.74E-02	6	702
GO:0055086	nucleobase-containing small molecule metabolic process	1.74E-02	6	789
GO:0006732	coenzyme metabolic process	1.74E-02	4	275
GO:0009150	purine ribonucleotide metabolic process	1.74E-02	5	561
GO:0006637	acyl-CoA metabolic process	1.74E-02	3	113
GO:0035383	thioester metabolic process	1.74E-02	3	113
GO:0009259	ribonucleotide metabolic process	1.74E-02	5	576
GO:0035313	wound healing, spreading of epidermal cells	1.74E-02	2	20
GO:0006163	purine nucleotide metabolic process	1.74E-02	5	5

Table 5: Top 10 biological processes identified on GO analysis of Cluster 8 using ToppFun. *Biological processes overexpressed in TB, and in sarcoidosis to a lesser extent, mainly associate with metabolic processes including nucleotide metabolism.*

Cluster 21 is the only cluster to be highly expressed in TB alone and comprises 7 genes including *MMP1* (Matrix Metalloproteinase 1), divalent transition metal transporter *SLC11A1* (Solute Carrier Family 11 Member 1, formerly known as *NRAMP1*), monocyte chemoattractants *CCL7* and *CCL8* (C-C Motif

Chemokine Ligand 7 and 8), *OLR1* (Oxidized Low Density Lipoprotein Receptor 1, formerly known as *LOX1*), *FAM124A* (Family With Sequence Similarity 124 Member A), and *LGALS17A* (Galectin 14 Pseudogene). Gene ontology performed on these 7 genes alone generated multiple biological processes (Table 6) and pathways (Table 7) with significance after correcting for FDR, implicating a central role for the inflammatory response, vascular wall, and extracellular matrix in TB pathogenesis.

ID	Name	FDR B&H	Genes from input	Genes in annotation
GO:0050900	leukocyte migration	2.13E-03	4	534
GO:0048245	eosinophil chemotaxis	2.13E-03	2	22
GO:0034341	response to interferon-gamma	2.13E-03	3	208
GO:0072677	eosinophil migration	2.13E-03	2	26
GO:0006954	inflammatory response	2.55E-03	4	806
GO:0048247	lymphocyte chemotaxis	8.49E-03	2	66
GO:0002548	monocyte chemotaxis	8.49E-03	2	68
GO:0034097	response to cytokine	8.57E-03	4	1287
GO:0070098	chemokine-mediated signaling pathway	8.57E-03	2	89
GO:0070839	divalent metal ion export	8.57E-03	1	1

Table 6: Top 10 biological processes identified on GO analysis of Cluster 21 using ToppFun. *Biological processes overexpressed in TB alone relate to the inflammatory response.*

ID	Name	FDR B&H	Genes from input	Genes in annotation
M15569	NOD-like receptor signaling	0.0061	2	62
M13088	PPAR signaling	0.0061	2	69
83042	IL-17 signaling	0.0077	2	72
PW:0000518	Atherosclerosis	0.0117	1	3
1269373	Cell surface interactions at the vascular wall	0.0117	2	143
1457779	Ion influx/efflux at host-pathogen interface	0.0117	1	4
M5885	Ensemble of genes encoding ECM-associated proteins including ECM-affiliated proteins, ECM regulators and secreted factors	0.0117	3	751
99051	Chemokine signaling	0.0117	2	182
P00031	Inflammation mediated by chemokine and cytokine signaling	0.0117	2	191
PW:0000244	Angiotensin II signaling	0.0181	1	10

Table 7: Top 10 pathways identified on GO analysis of Cluster 21 using ToppFun. *Pathways overexpressed in TB alone relate to chemokine signalling, vascular wall interaction, and extracellular matrix organisation.*

Conversely, the only cluster to show higher expression values in sarcoidosis than TB is Cluster 14, composed of 9 genes: *ATP6V0D2* (ATPase H+ Transporting V0 Subunit D2); *BEAN1* (Brain Expressed Associated With NEDD4); *CHIT1* (Chitinase 1); *CYP27A1* (Cytochrome P450 Family 27 Subfamily A Member 1); *DNASE2B* (Deoxyribonuclease 2 Beta); *FAIM2* (Fas Apoptotic Inhibitory Molecule 2); *KCNJ5* (Potassium Inwardly Rectifying Channel Subfamily J Member 5); *SLC26A11* (Solute Carrier Family 26

Member 11); and *SLC47A1* (Solute Carrier Family 47 Member 1). GO analysis of Cluster 14 produced no associated biological processes or pathways, though did indicate the lysosome as an important cellular component (Table 8). Four of the nine genes (*ATP6V0D2, CHIT1, DNASE2B* and *SLC26A11)* identify in both lysosome and lytic vacuole activity.

ID	Name	FDR B&H	Genes from input	Genes in annotation
GO:0005764	Lysosome	5.750E-03	4	735
GO:0000323	Lytic vacuole	5.750E-03	4	736
GO:0033181	Plasma membrane proton-transporting V-type ATPase complex	3.787E-02	1	4

Table 8: All cellular components identified on GO analysis of Cluster 14 using ToppFun. *Cellular components overexpressed in sarcoidosis, and in TB to a lesser extent, relate to lysosomal activity.*

In contrast, GO analysis of Cluster 23, a cluster of 7 genes downregulated in TB and sarcoidosis in comparison to control samples, though in TB to a greater extent, demonstrates no associated biological processes or pathways. However, all associated cellular components involve collagen (Table 9), due to the presence of a single gene, *COL4A4*, in the cluster.

ID	Name	FDR B&H	Genes from input	Genes in annotation
GO:0005587	Collagen type IV trimer	2.106E-02	1	6
GO:0098642	Network-forming collagen trimer	2.106E-02	1	7
GO:0098645	Collagen network	2.106E-02	1	7
GO:0098651	Basement membrane collagen trimer	2.106E-02	1	8
GO:0098644	Complex of collagen trimers	3.996E-02	1	19

Table 9: All cellular components identified on GO analysis of Cluster 23 using ToppFun. *Cellular components underexpressed in TB, and in sarcoidosis to a lesser extent, all include collagen.*

The final cluster to be analysed, Cluster 25, is also downregulated in TB relative to control, and downregulated to a lesser degree in sarcoidosis. GO analysis of the 6 genes constituting Cluster 25 shows multiple immune-related processes including regulation of dendritic cell response and T cell activation, and cell-matrix adhesion (Table 10).

ID	Name	FDR B&H	Genes from input	Genes in annotation
GO:0002684	Positive regulation of immune system process	3.326E-02	4	1268
GO:2000529	Positive regulation of myeloid dendritic ell chemotaxis	3.326E-02	1	1
GO:2000527	Regulation of myeloid dendritic cell chemotaxis	3.326E-02	1	1
GO:2000548	Negative regulation of dendritic cell dendrite assembly	3.326E-02	1	1
GO:0042110	T cell activation	3.326E-02	3	537
GO:0035759	Mesangial cell-matrix adhesion	3.568E-02	1	2
GO:0002682	Regulation of immune system process	3.568E-02	4	1775
GO:2000547	Regulation of dendritic cell dendrite assembly	3.568E-02	1	3
GO:0002253	Activation of immune response	3.568E-02	3	752
GO:0038116	Chemokine (C-C motif) ligand 21 signaling pathway	3.568E-02	1	4

Table 10: Top 10 biological processes identified on GO analysis of Cluster 25 using ToppFun. *Biological processes underexpressed in TB, and in sarcoidosis to a lesser extent, associate with the cellular adhesion and the immune response including dendritic cell regulation.*

Following on from correlation analysis, differential gene expression was further analysed using gene ontology and gene set enrichment analysis to corroborate the findings from Markov Clustering.

3.2.7 Gene Ontology: Upregulated genes

To study the upregulated pathways in TB and in sarcoidosis, the differential gene expression performed by limma underwent gene ontology (GO) analysis using the ReactomePA software tool, using Benjamini and Hochberg's false discovery rate (FDR) of less than 0.05. ToppFun (FDR cutoff of 0.05) and g:Profiler (adjusted P value threshold of 0.05) tools were also used to confirm the results produced by ReactomePA software. The top 20 most upregulated biological processes and REACTOME pathways were identified in 3 separate comparisons: TB compared to control; sarcoidosis compared to control; and TB compared to sarcoidosis samples. Comparison of TB with control samples shows upregulated biological processes associated with the immune response, exocytosis, cytokine signalling and response to bacterium (Figure 21). Interestingly, comparison of sarcoidosis with control samples generates many of the same upregulated biological processes identified when comparing TB with control (Figure 22).

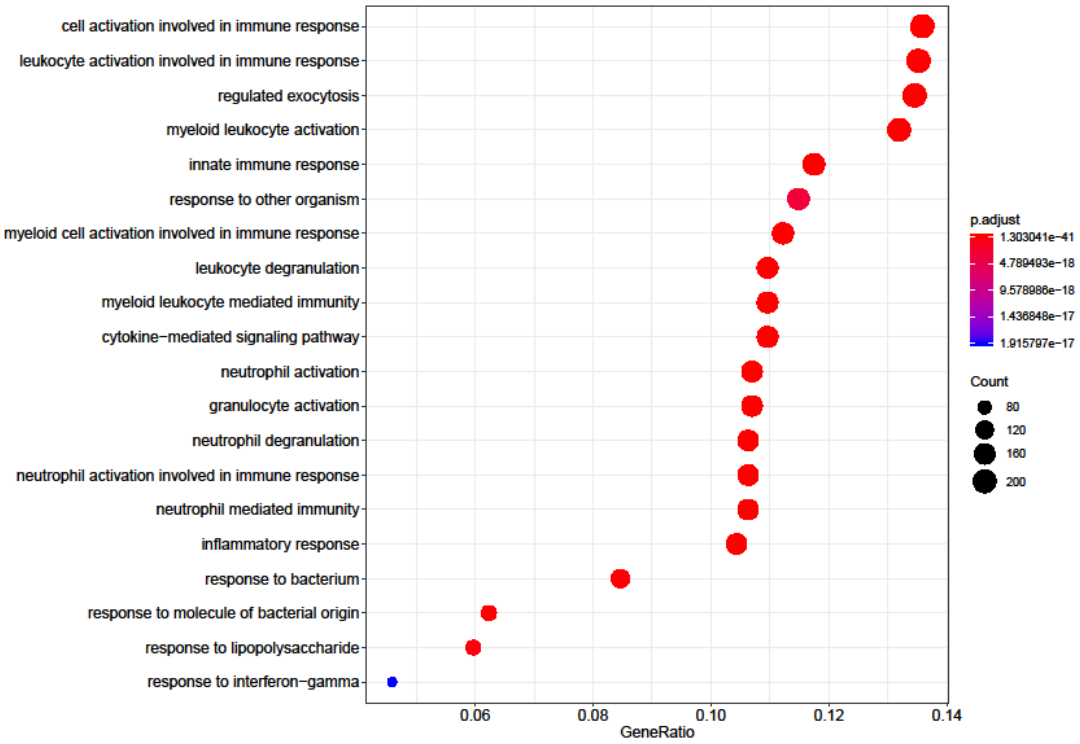


Figure 21: Top 20 biological processes upregulated in the comparison of TB with control samples according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Biological processes relate to immune response, exocytosis, cytokine signalling and response to bacterium.*

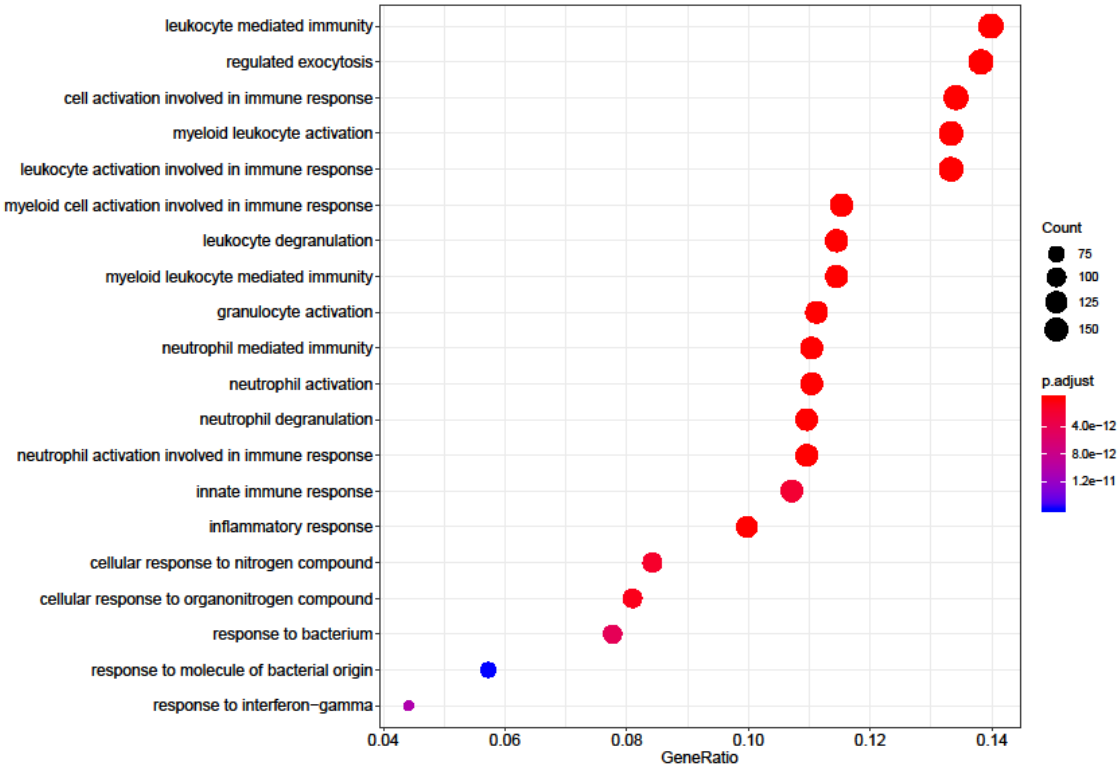


Figure 22: Top 20 biological processes upregulated in the comparison of sarcoidosis with control samples according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Biological processes relate to immune response, exocytosis, cytokine signalling and response to bacterium, similar findings to those found in TB relative to control samples.*

The differences between TB and sarcoidosis were explored with the aim to identify key processes contributing to the development of lung destruction by lung cavitation, a hallmark feature of TB. Biological processes significantly upregulated in TB when directly compared to sarcoidosis samples involve cytokine signalling, angiogenesis, extracellular matrix organisation and nitric oxide biosynthesis (Figure 23).

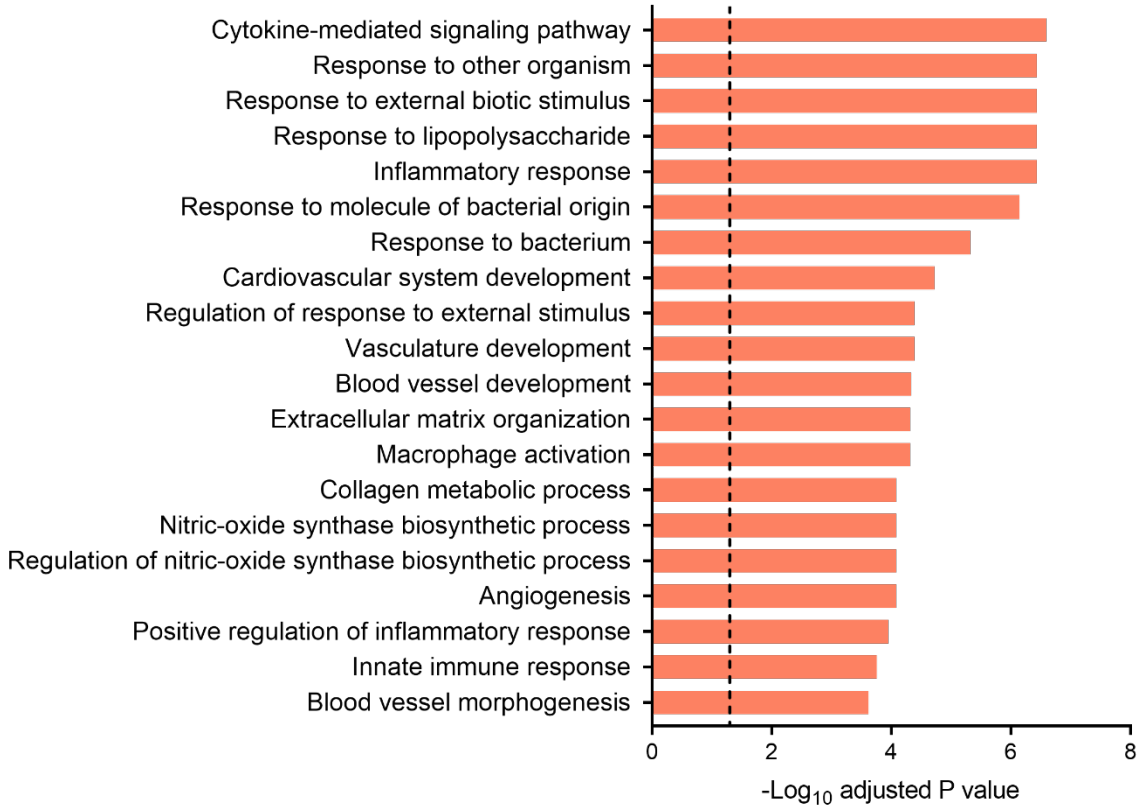


Figure 23: Top 20 biological processes upregulated in clinical TB relative to sarcoidosis samples according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *The cytokine-mediated signalling pathway is the most significantly upregulated biological process in TB. Other biological processes relate to the inflammatory response, angiogenesis and extracellular matrix organisation. Dashed represents adjusted P value of 0.05.*

The most significantly upregulated biological process in clinical TB, compared to sarcoidosis samples, is the cytokine-mediated signalling pathway. For the same comparison, upregulated DEGs in the cytokine-mediated signalling pathway are plotted and demonstrate *MMP-1* (matrix metalloproteinase 1) to have the highest fold change (Figure 24). Both expected genes such as *CCL2* (C-C motif chemokine ligand 2, previously known as *MCP1*, monocyte chemotactic protein 1), *CXCL9* (C-X-C motif chemokine ligand 9), *HIF1A* (hypoxia inducible factor 1 subunit alpha) and *IL1B* (interleukin 1 beta), as well as unexpected genes such as *SPHK1* (sphingosine kinase 1) are upregulated components of the cytokine-mediated signalling pathway in TB.

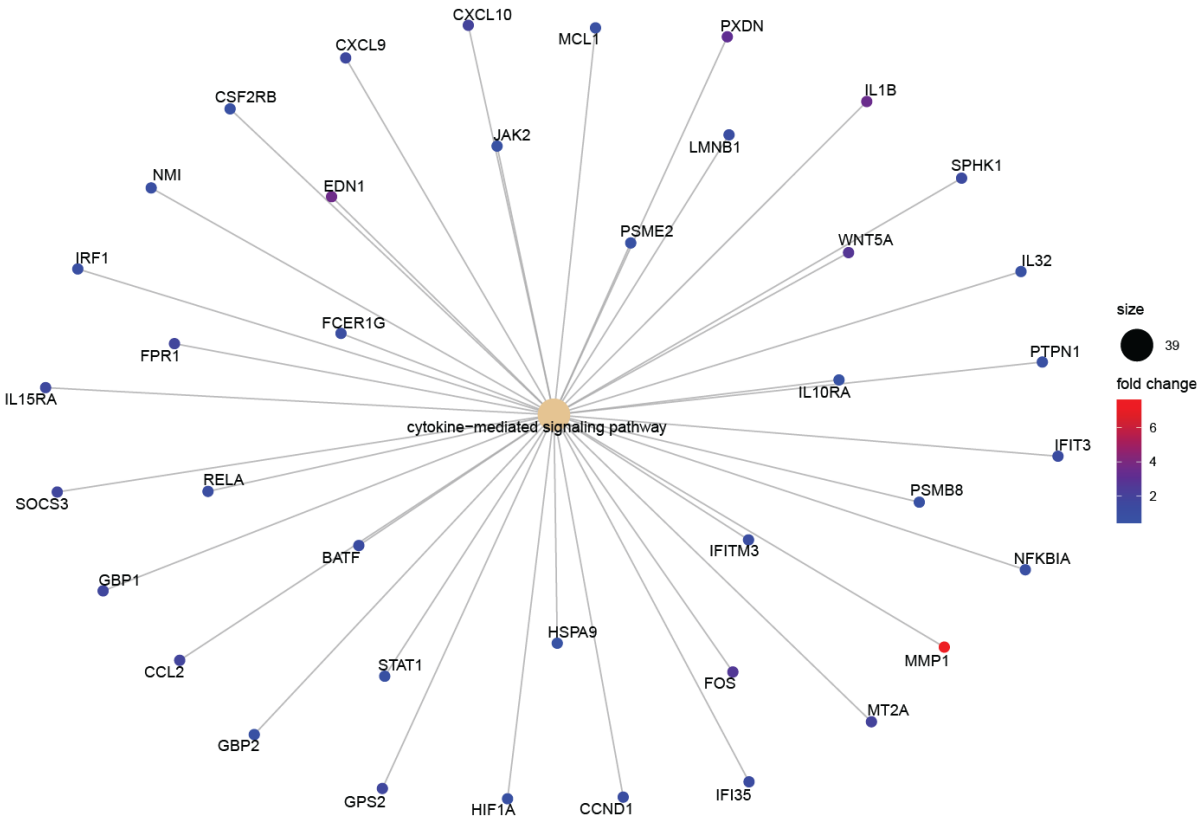


Figure 24: Cnetplot of the cytokine-mediated signalling pathway in TB relative to sarcoidosis, with each upregulated gene and associated log_2 fold change displayed, and adjusted P value < 0.05. *MMP1 was the most highly upregulated gene by fold change.*

The REACTOME pathways significantly upregulated when TB is compared with control samples relate to the immune response, cytokine signalling, extracellular matrix organisation, platelet activation, phagocytosis, and transferrin endocytosis (Figure 25). As observed with identification of upregulated biological processes, comparison of sarcoidosis with control samples generates many of the same upregulated REACTOME pathways identified when comparing TB with control samples (Figure 26). Importantly, comparison of sarcoidosis with control samples shows upregulation of lipoprotein metabolism, a process absent when comparing TB with control samples.

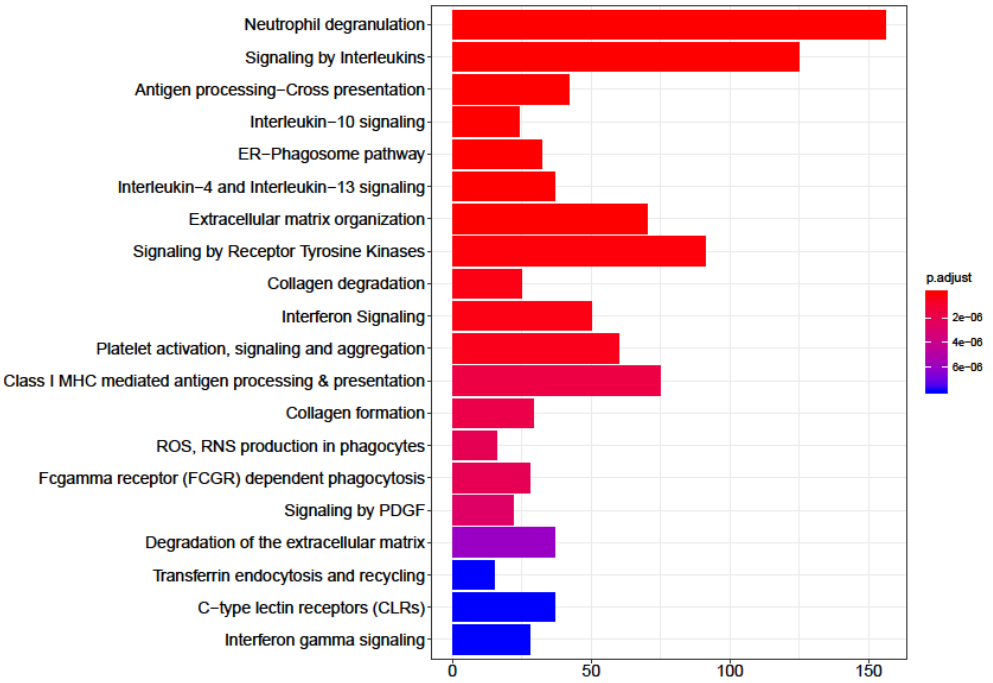


***Figure 25: Top 20 REACTOME pathways upregulated in the comparison of TB with control samples according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment**). Pathways associate with immune response, cytokine signalling, extracellular matrix organisation, platelet activation, phagocytosis, and transferrin endocytosis.*

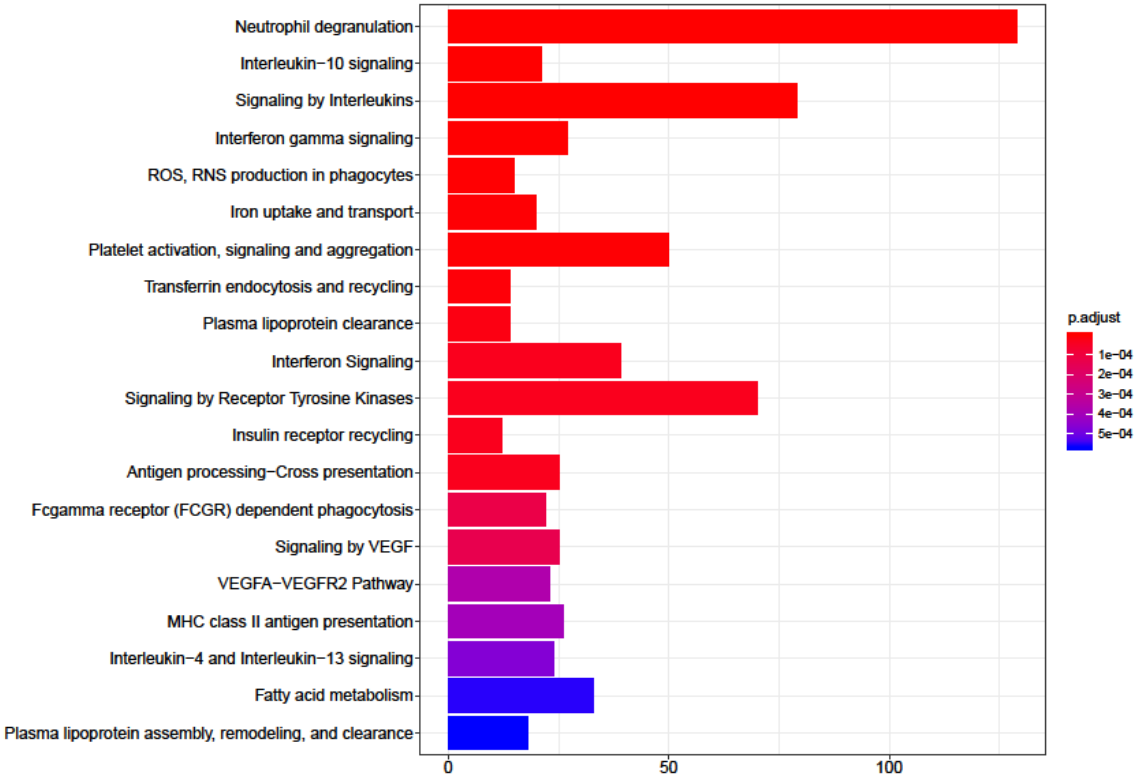


***Figure 26: Top 20 REACTOME pathways upregulated in the comparison of sarcoidosis with control samples according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment).** Lipoprotein metabolism is the only upregulated pathway absent in TB relative to control samples.*

The REACTOME pathways upregulated in TB when directly compared to sarcoidosis samples mainly relate to cytokine signalling and extracellular matrix organisation (Figure 27).

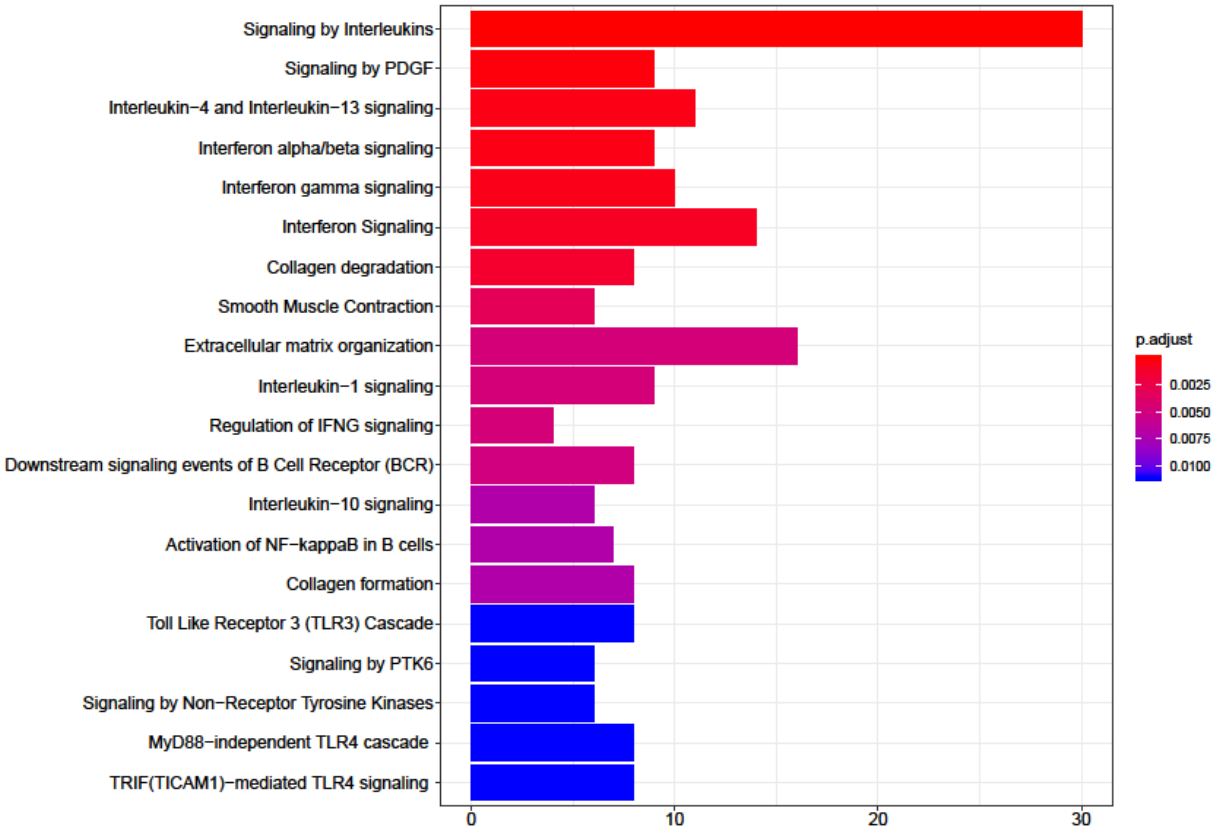


Figure 27: Top 20 REACTOME pathways upregulated in the comparison of TB with sarcoidosis samples according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Pathways associate with cytokine signalling and extracellular matrix organisation.*

Overall, gene ontology identified the main processes and pathways upregulated in TB associate with cytokine signalling, extracellular matrix organisation, and angiogenesis, whereas those in sarcoidosis associate with lipid metabolism and to a lesser degree, cytokine signalling and extracellular matrix organisation.

3.2.8 Gene Ontology: Downregulated genes

To study downregulated pathways in TB and in sarcoidosis, I used the ReactomePA tool with a Benjamini and Hochberg's false discovery rate (FDR) of less than 0.05. The differential gene expression performed by limma underwent gene ontology (GO) analysis. ToppFun (FDR cutoff of 0.05) and g:Profiler (adjusted P value threshold of 0.05) tools were also used to confirm the results generated by ReactomePA. The top 20 most downregulated biological processes and REACTOME pathways were identified in 3 separate comparisons: TB compared to control; sarcoidosis compared to control; and TB compared to sarcoidosis samples. Comparison of TB with control samples shows

most downregulated biological processes are associated with translation and B cell differentiation (Figure 28).

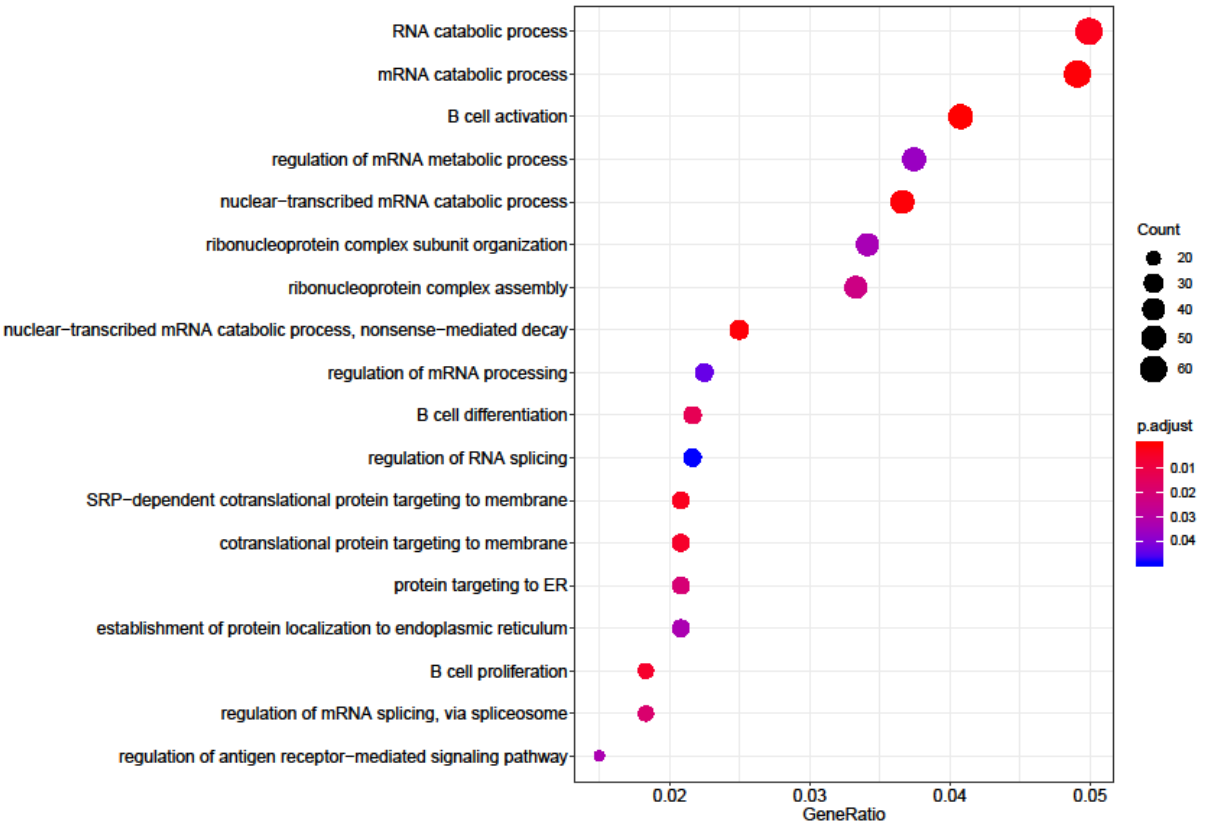


Figure 28: Top 20 biological processes downregulated in the comparison of TB with control samples according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Biological processes associate with RNA metabolism, protein translation and B cell differentiation.*

Similar downregulated biological processes are seen when sarcoidosis is compared to control samples (Figure 29), including RNA metabolism, protein translation and B cell differentiation, demonstrating again overlapping pathology between TB and sarcoidosis.

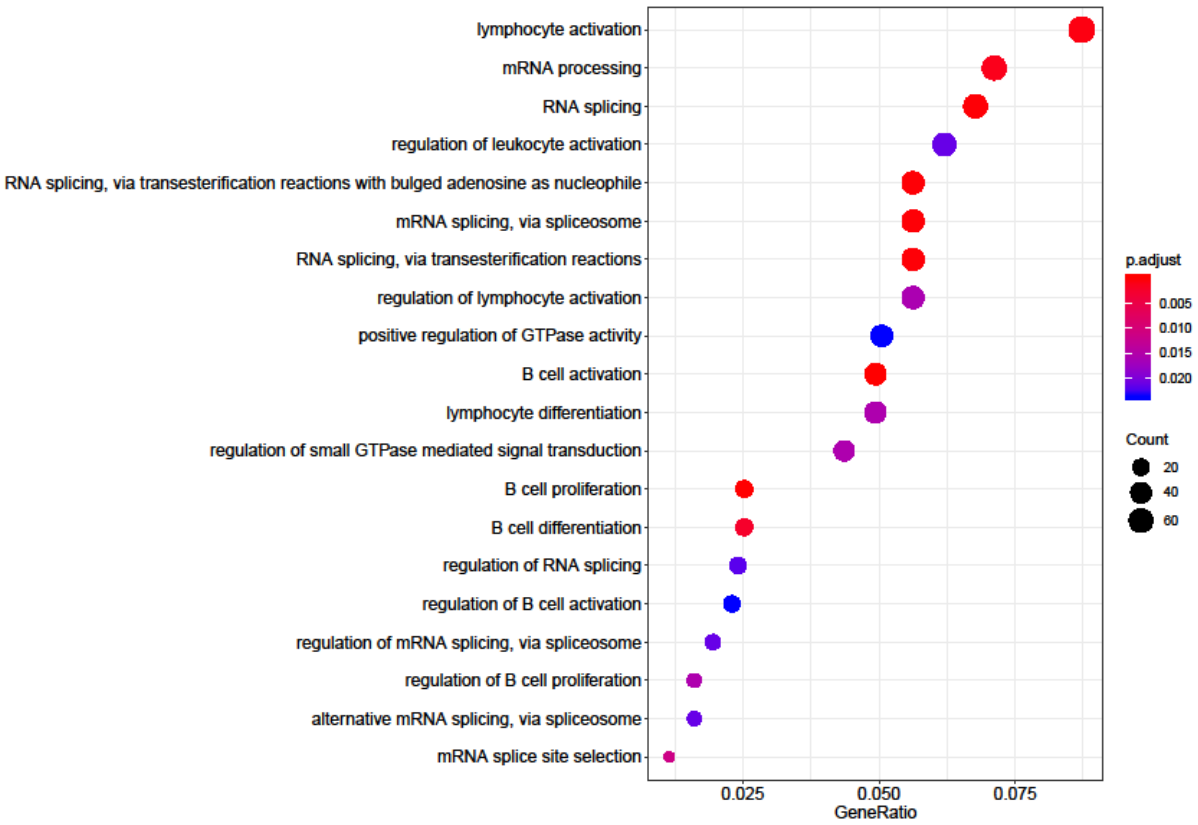


Figure 29: Top 20 biological processes downregulated in the comparison of sarcoidosis with control samples according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Biological processes associate with RNA metabolism, protein translation and B cell differentiation, similar to those seen in TB relative to control samples.*

The inherent difference in cell composition of the three sample groups may partly explain this observation, as the TB and sarcoidosis granulomas consist mainly of macrophages, whereas normal lymph node samples contain more B cells. Of note, no downregulated biological processes are identified when the two granulomatous conditions, TB and sarcoidosis, are compared.

The REACTOME pathways significantly downregulated when TB is compared with control samples are related with transcription, protein translation, cilia assembly and the Influenza virus (Figure 30). Similarly, comparison of sarcoidosis with control samples shows the most significantly downregulated REACTOME pathways involve the cell cycle, transcription, translation, and the Influenza virus (Figure 31), and also demonstrates downregulation in pathways associated with apoptosis.

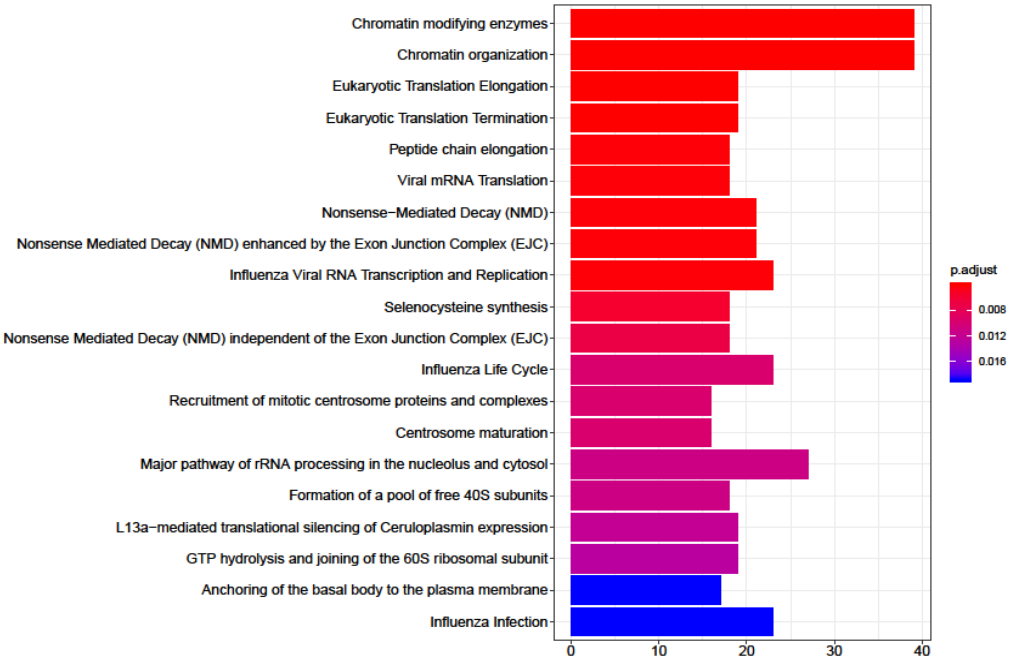


Figure 30: Top 20 REACTOME pathways downregulated in the comparison of TB with control samples according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Pathways associate with transcription, translation, cilia assembly and the Influenza virus.*

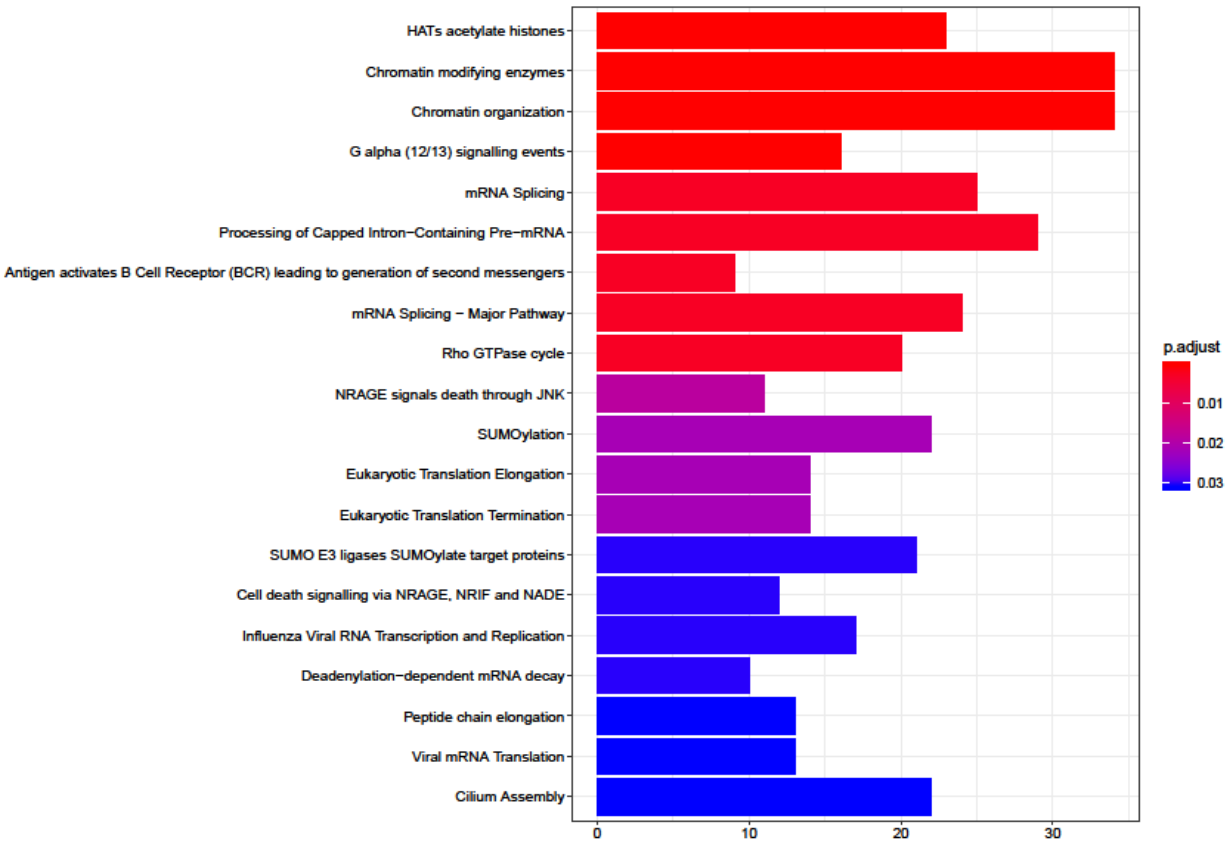


Figure 31: Top 20 REACTOME pathways downregulated in the comparison of sarcoidosis with control samples according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Pathways associate with the cell cycle, transcription, translation, and the Influenza virus, similar to pathways downregulated in TB relative to control samples.*

As observed with biological processes, no significantly downregulated REACTOME pathways are identified when TB is directly compared to sarcoidosis samples.

3.2.9 Gene Set Enrichment Analysis

To further strengthen my analyses, another analytical method was performed. Gene set enrichment analysis (GSEA), also named functional enrichment analysis, was used to determine the most enriched pathways in the comparison of TB and control samples, and sarcoidosis and control samples. The fgsea package analysed the differential gene expression data performed by limma. The canonical pathways gene set was selected for its wide and diverse range of pathways. The top 10 canonical pathways with greatest normalised enrichment score (NES), both positive and negative, were selected for display. The enrichment plot of the pathways with highest and lowest NES are also shown. Comparison of TB with control samples demonstrates the matrisome, otherwise known as the extracellular matrix, plays a significant role in TB (Figure 32). The other important overexpressed canonical pathways relate to cytokine signalling, urokinase-type plasminogen activator (UPA) signalling, and the lysosome. Conversely, the underexpressed canonical pathways associate with olfactory transduction.

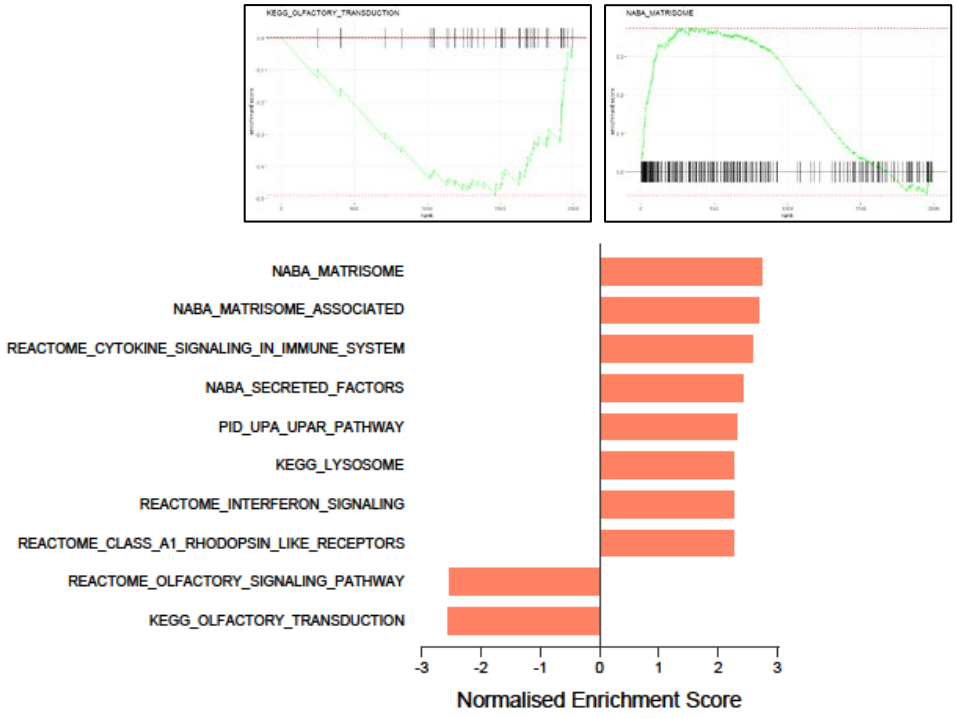


Figure 32: Top 10 canonical pathways with adjusted P value < 0.05 in the comparison of TB with control samples according to most extreme NES, analysed by fgsea (R environment). *Pathways relating to the extracellular matrix, cytokine signalling, urokinase-type plasminogen activator and lysosome are most overexpressed.*

Comparison of sarcoidosis with control samples also demonstrates the significant role of the matrisome in sarcoidosis, as well as urokinase-type plasminogen activator (UPA) signalling and the

lysosome (Figure 33), similar to those seen when TB is compared with control samples. Interestingly, leishmaniasis, another infectious cause of granuloma formation, is also identified as a significant canonical pathway in sarcoidosis.

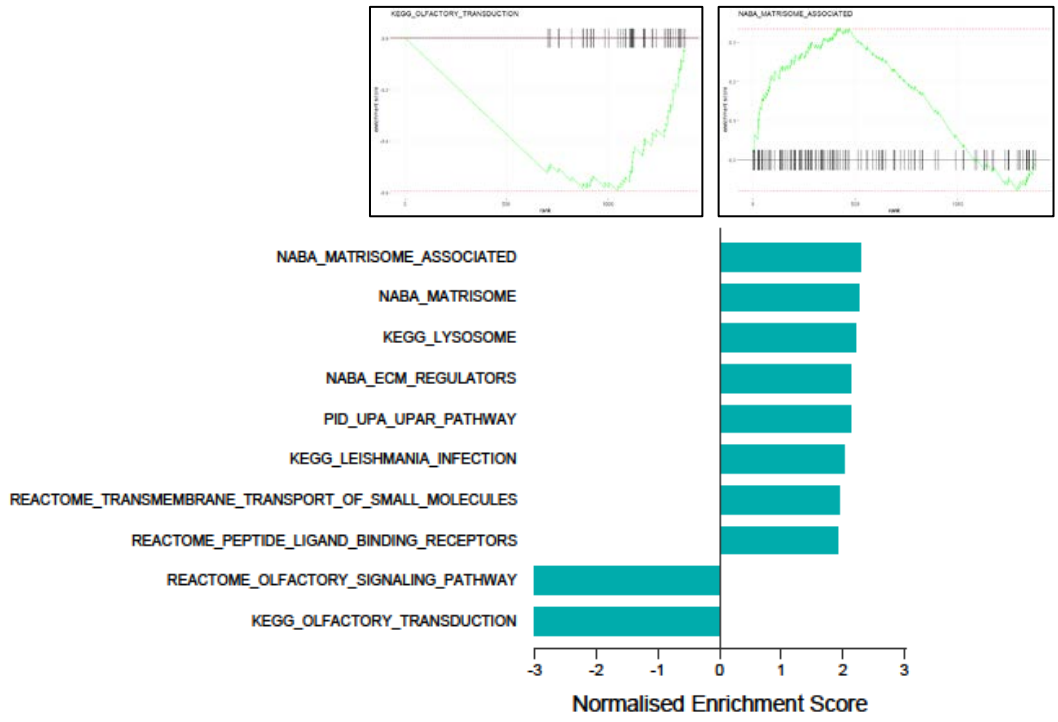


Figure 33: Top 10 canonical pathways with adjusted P value < 0.05 in the comparison of sarcoidosis with control samples according to most extreme NES, analysed by fgsea (R environment). *Pathways relating to the extracellular matrix, lysosome, urokinase-type plasminogen activator and leishmaniasis are most highly expressed.*

The egsea package is a more recently developed GSEA tool, combining 11 established methods for GSEA and enabling a powerful manner of analysis by pathway visualisation. Expression of genes in each KEGG pathway are expected to be upregulated when implicated in disease. Therefore gene expression in clinical TB relative to control samples was overlaid on to the KEGG TB pathway, to see which of the expected genes were upregulated in the clinical TB samples (Figure 34). A majority of the KEGG TB pathway genes are upregulated, demonstrating overlap between existing known TB pathway mechanisms and that found in the clinical TB samples.

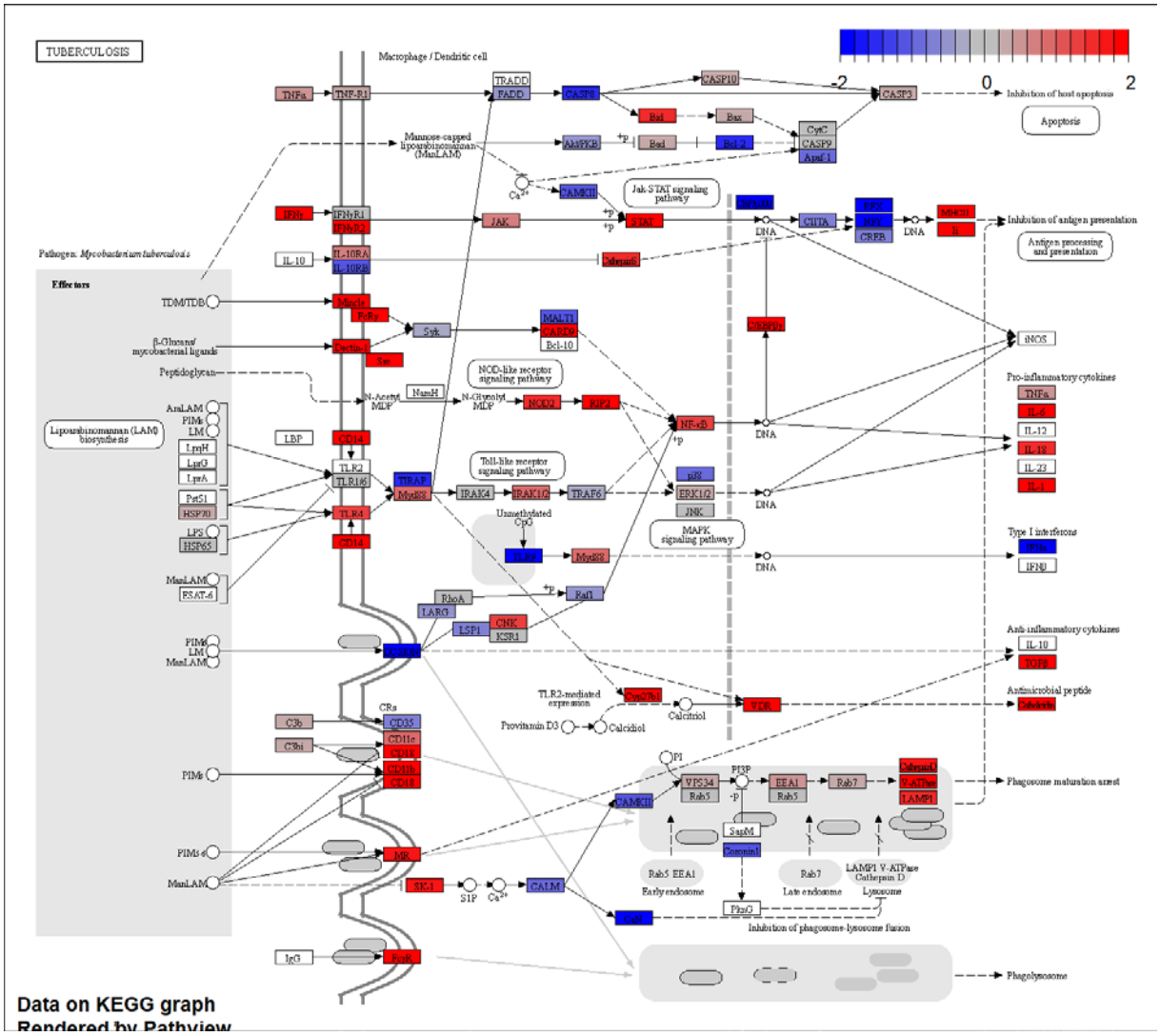


Figure 34: Expression of genes regulated in clinical TB in comparison to control specimens, overlaid on the KEGG TB pathway, analysed by egsea (R environment). *Fold changes have been converted to a Z-score, with upregulated genes in red and downregulated genes in blue. A majority of expected genes in the KEGG TB pathway are upregulated in clinical TB samples.*

Gene set enrichment by egsea was also utilised to find the most significant KEGG pathways associated with TB relative to control, and sarcoidosis relative to control samples (Table 11). A majority of the top pathways identified were communal to TB and sarcoidosis, with the phagosome, lysosome, other infections, and autoimmune conditions such as rheumatoid arthritis highlighted among the pathways.

KEGG pathway identified in TB	Rank	Adj. P value	KEGG pathway in sarcoidosis	Rank	Adj. P value
Rheumatoid arthritis	1	7.47E-12	Lysosome	1	1.02E-15
Phagosome	2	2.74E-11	Phagosome	2	6.09E-13
Staphylococcus aureus infection	3	2.74E-11	Staphylococcus aureus infection	3	1.24E-12
Lysosome	4	6.65E-08	Rheumatoid arthritis	4	2.28E-12
Pertussis	5	1.74E-07	Tuberculosis	5	2.73E-07
Leishmaniasis	6	2.44E-07	Leishmaniasis	6	4.45E-07
Oxidative phosphorylation	7	4.38E-07	Oxidative phosphorylation	7	4.45E-07
Toll-like receptor signaling	8	1.04E-07	Inflammatory bowel disease	8	4.45E-07
Mineral absorption	9	3.25E-07	Systemic lupus erythematosus	9	4.45E-07
Tuberculosis	10	4.38E-07	Mineral absorption	10	8.12E-06

Table 11: Top ten KEGG pathways in TB and sarcoidosis relative to control samples, analysed by egsea (camera method, R environment). *All pathways are upregulated, with 8 out of the 10 pathways communal to TB and sarcoidosis. Shared pathways include autoimmune conditions, TB, leishmaniasis, Staphylococcus aureus infection, and the phagosome and lysosome.*

Further analysis using gene set enrichment by egsea compared the differences in cellular components and KEGG pathways between the 3 clinical groups. Of note, the comparison between TB and sarcoidosis was made in order to identify any differential features which may explain why destructive lung cavitation occurs in TB but rarely occurs in sarcoidosis. As expected, most cellular components and KEGG pathways were upregulated in both TB and sarcoidosis compared to control, with more fold change difference seen in TB. However, the only cellular component downregulated in comparison to sarcoidosis is the lytic vacuole membrane (Figure 35). Similarly the only downregulated KEGG pathway in TB relative to sarcoidosis is the lysosome pathway (Figure 36).

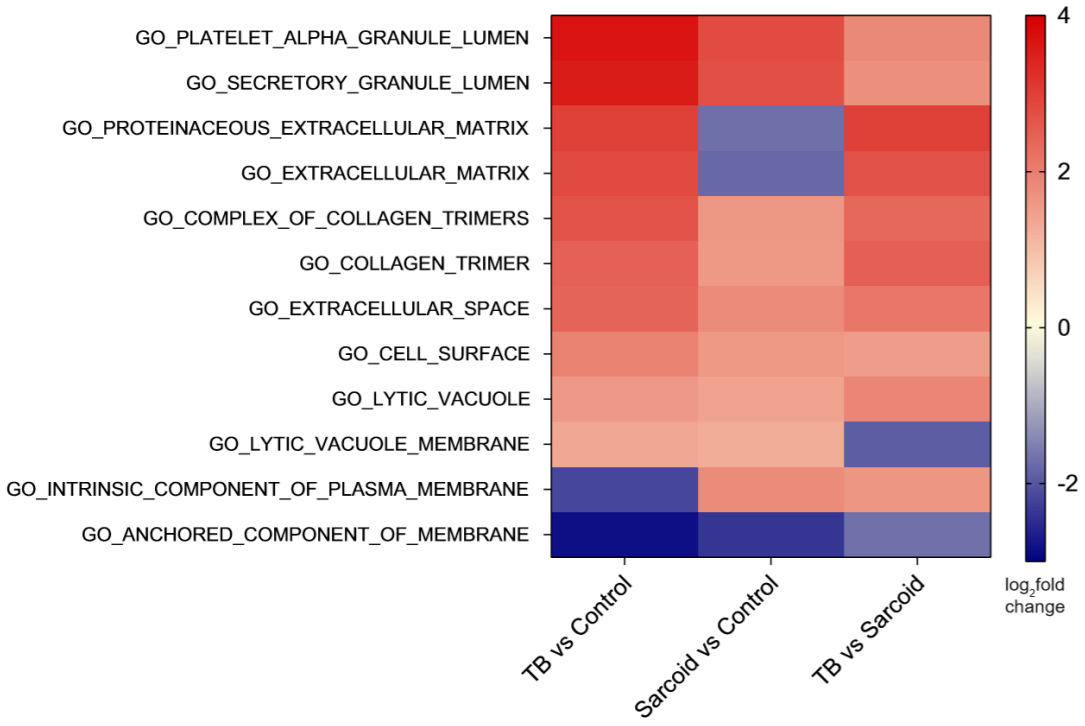


Figure 35: Top cellular components with adjusted P value < 0.05 analysed by egsea (R environment) in 3 comparisons: TB relative to control; sarcoidosis relative to control; and TB relative to sarcoidosis. *The only cellular component to be downregulated in TB relative to sarcoidosis is the lytic vacuole membrane.*

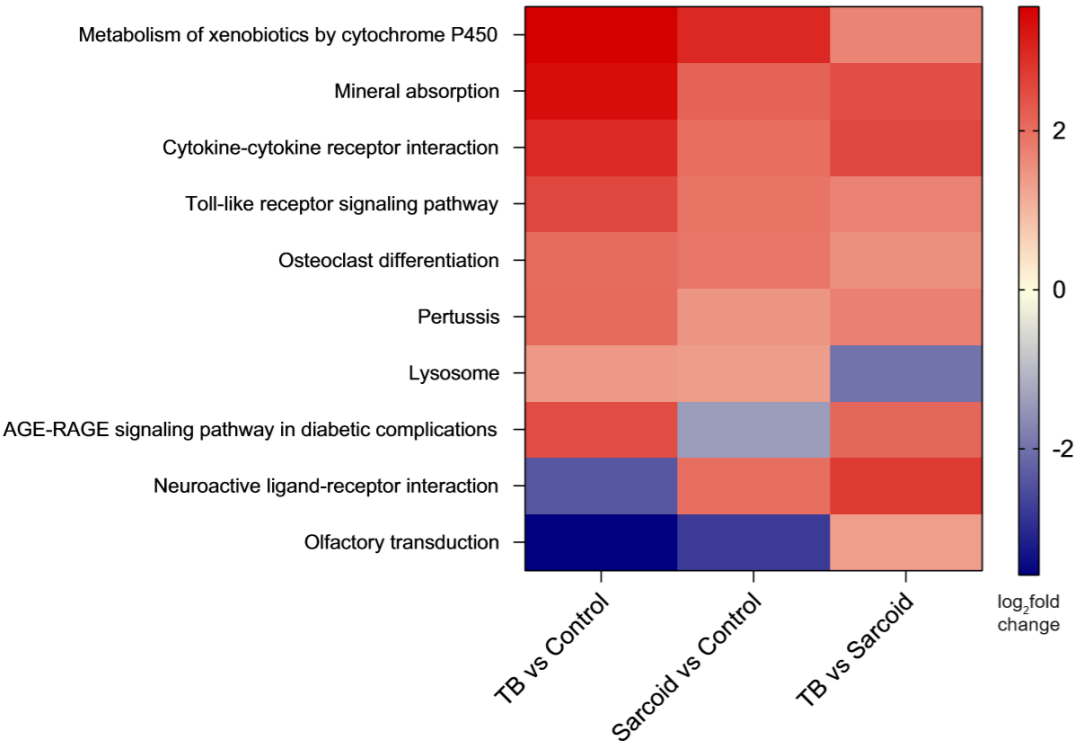


Figure 36: Top KEGG pathways with adjusted P value < 0.05 analysed by egsea (R environment) in 3 comparisons: TB relative to control; sarcoidosis relative to control; and TB relative to sarcoidosis. *The only pathway to be downregulated in TB relative to sarcoidosis is the lysosome pathway.*

3.3 Discussion

The approach of laser capture microdissection and RNA sequencing of granulomas in clinical samples of untreated TB, sarcoidosis and control lymph node, all of whom were non-smokers, has provided further insight into important biological processes and pathways in these granulomatous conditions. Previous genomic studies in TB have used patients who have already received antituberculous therapy (Kim *et al.*, 2010) or whose samples are from a peripheral site of disease such as blood samples (Scriba *et al.*, 2017). Therefore, I have generated the first unbiased analysis of untreated TB lymph node samples.

A significant challenge for the "omic" era is to analyse the huge amount of data generated and to interpret its biological relevance (Pop and Salzberg, 2008). My transcriptomic analysis of the clinical samples employed a variety of methods including correlation analysis, differential gene expression, gene ontology and gene set enrichment analysis to perform a non-hypothesis driven study to understand TB immunopathogenesis and ultimately find potential targets for host-directed therapy. Comparison of TB and sarcoidosis to control tissue demonstrates a significant role for the extracellular matrix and immune pathways, consistent with previous literature (Ong and Friedland, 2014). In contrast, comparison between TB and sarcoidosis shows angiogenesis and lysosomal activity may differentiate these conditions, which has not been described in previous reports (Koth *et al.*, 2011; Agrawal *et al.*, 2016).

Given the histological and clinical similarities between TB and sarcoidosis (Mitchell *et al.*, 1977), it is unsurprising these conditions share many processes and pathways. Initial data exploration finds overlap of TB and sarcoidosis, and separation from control samples, on principle component analysis (Figure 10) and hierarchical clustering (Figure 12, Figure 14). One third of all differentially expressed genes (DEGs) are shared between TB and sarcoidosis (Figure 16), though overall TB has a higher number of DEGs and higher degree of fold change than in sarcoidosis. Correlation analysis identifies clusters which are upregulated in both conditions relative to control, though almost always to a higher degree in TB (Figure 20). Gene ontology and gene set enrichment identifies cytokine signalling and extracellular matrix organisation as significant processes upregulated in TB and sarcoidosis, again to a greater extent in TB. Blood transcriptomes of TB and sarcoidosis have already been shown to overlap in the regulation of a number of inflammatory pathways (Koth *et al.*, 2011). Furthermore, TB and sarcoidosis have been postulated to be on the same disease spectrum (Agrawal

et al., 2016), and my analysis supports this theory, but also identifies significant differences. A proportion of these genes will merely reflect the cell differences, with macrophage gene signatures dominant purely as a reflection of the cell number. However, it is possible a group of communal genes and pathways between TB and sarcoidosis represent the essential components of granuloma formation and stability, and that the granuloma may be the product of host immune tolerance by limiting an excessive and potentially destructive host immune response.

The shared immunopathology of TB and sarcoidosis is not fully understood but likely to be complex and involving many pathways. Bioinformatic analysis of the clinical samples reassuringly demonstrated many of these known immune processes to be upregulated in both conditions. Shared DEGs include *CSF2* (Colony Stimulating Factor 2, previously known as GMCSF, Granulocyte-Macrophage Colony Stimulating Factor) and *IL1RN* (Interleukin 1 Receptor Antagonist) (Table 2), while gene ontology identifies shared pathways include FcγR (Fc gamma Receptor) dependent phagocytosis, and signalling by IFN-γ (Interferon gamma) and T helper 2 (Th2) derived cytokines IL-4, IL-10 and IL-13 (Figure 25, Figure 26). Gene set enrichment finds both TB and sarcoidosis associate with the phagosome, lysosome and autoimmune pathways. However, the cytokine-mediated signalling pathway is also the most significantly upregulated biological process in TB relative to sarcoidosis (Figure 23), comprising of proinflammatory chemokines such as *CCL2* (C-C motif chemokine ligand 2, previously known as *MCP1*, monocyte chemotactic protein 1) and *CXCL9* (C-X-C motif chemokine ligand 9), and immunoregulatory chemokines such as *IL10RA* (Interleukin 10 Receptor Subunit Alpha) (Figure 24). This clinical study has confirmed important immune pathways in TB that have been well documented (O'Garra and Berry, 2013), including overlap between the TB signature and autoimmune diseases such as rheumatoid arthritis, consistent with emerging evidence of shared underlying pathways (Clayton *et al.*, 2017). Importantly, the balance of cytokines is likely to play a part in distinguishing TB immunopathology from sarcoidosis, as opposed to a simple presence or absence.

An essential role for the extracellular matrix is apparent in the TB and sarcoidosis samples. The top ten DEGs, according to fold change, communal to TB and sarcoidosis include *TNFAIP6* (TNF Alpha Induced Protein 6), a protein involved with extracellular matrix stability, and *FAP* (Fibroblast Activation Protein Alpha), a gelatinase involved with extracellular matrix degradation (Table 2). Correlation analysis demonstrates clusters both overexpressing and underexpressing extracellular matrix processes and components, and may also relate to wound healing processes seen in these

clusters (Table 3, Table 4, Table 9, Table 10). Gene set enrichment analysis confirms the extracellular matrix to be the most significant pathway in both TB and sarcoidosis (Figure 32, Figure 33). These findings are consistent with both clinical and experimental data pointing to a central role for the matrix in TB pathogenesis (Elkington and Friedland, 2011b; Al Shammari *et al.*, 2015; Kubler *et al.*, 2015).

Despite the significant overlap between TB and sarcoidosis, the two conditions are often clinically distinct, with more intense inflammation and lung cavitation defining TB. Of the top ten DEGs in TB alone, based on fold change, only two are exclusively upregulated in TB: *CCL7* (C-C Motif Chemokine Ligand 7) and *MMP1* (Matrix Metalloproteinase 1) (Figure 18). Interestingly, these genes are two of the seven DEGs in Cluster 21, the only cluster unique to TB on correlation analysis (**Error! Reference source not found.**). The remaining DEGs comprising Cluster 21 are the divalent transition metal transporter *SLC11A1* (Solute Carrier Family 11 Member 1, formerly known as *NRAMP1*), monocyte chemoattractants *CCL8* (C-C Motif Chemokine Ligand 8), *OLR1* (Oxidized Low Density Lipoprotein Receptor 1, formerly known as *LOX1*), *FAM124A* (Family With Sequence Similarity 124 Member A), and *LGALS17A* (Galectin 14 Pseudogene). Gene ontology performed on these seven genes generates multiple biological processes and pathways with significance after correcting for FDR, implicating a central role for the inflammatory response and extracellular matrix turnover in TB pathogenesis (Table 6, Table 7).

The importance of this seven gene cluster was further considered. Notably, *SLC11A1* (*NRAMP1*) polymorphisms were the first to be linked to TB susceptibility over 20 years ago (Bellamy *et al.*, 1998). *SLC11A1* exerts pleiotropic effects on macrophage function that will increase inflammation, including enhanced expression of MHC Class II, cytokines such as TNF-α and IL-1β, and inducible nitric oxide synthase (Archer, Nassif and O'Brien, 2015). The high expression of chemokines *CCL7* (*MCP3*) and *CCL8* (*MCP2*) supports the emerging concept that excessive monocyte recruitment is harmful in TB (Davis and Ramakrishnan, 2009). *LGALS17A* and *FAM124A* may augment inflammation through NFκB activation (Rual *et al.*, 2005), though their roles are less well defined. *OLR1* can regulate foamy macrophage formation, a typical feature of the human TB granuloma (Russell *et al.*, 2009), and contributes to excessive inflammation in atherosclerosis and myocardial ischaemia (Li *et al.*, 2002; Hansson and Libby, 2006). Moreover, *OLR1* activation induces epigenetic reprogramming in macrophages (Bekkering *et al.*, 2014), suggesting trained immunity may contribute to the hyper-inflammatory state observed in TB. Considering this cluster of seven TB-specific genes together

implies that dysregulated inflammation is a sequential process, whereby excessive monocytes are recruited to the granuloma (*CCL7* and *CCL8)*, reprogrammed to a hyper-inflammatory phenotype (*OLR1)*, further propagating inflammation (*SLC11A1*) and NFκB activation (*FAM124A*), resulting in extracellular matrix degradation (*MMP1*), cavity formation and transmission.

In contrast, correlation analysis does not identify any cluster unique to sarcoidosis, though Cluster 14 is the only sarcoidosis predominant gene cluster (Figure 20). Similarly, the top ten DEGs in sarcoidosis relative to control, based on fold change, are sarcoidosis predominant rather than exclusive to sarcoidosis (Figure 18). These ten DEGs are located intracellularly or at the plasma membrane and are often associated with neural development and protein synthesis (Table 2). Lipid metabolism is another upregulated process identified in sarcoidosis using gene ontology (Figure 26), and further evidenced by the presence of *CYP27A1* (Cytochrome P450 Family 27 Subfamily A Member 1), important in cholesterol homeostasis, in Cluster 14 (Cali *et al.*, 1991). However, the strongest evidence indicates the lysosome plays a key role in sarcoidosis. The lysosome pathway is ranked number one in sarcoidosis relative to control, using gene set enrichment analysis (Table 11). Also, despite a lack of biological processes or pathways on gene ontology analysis of sarcoidosis predominant Cluster 14, four of the nine genes in this cluster contribute to overexpression of the lysosome as a cellular component (Table 8). Moreover, two of the nine genes in Cluster 14 also appear in the top ten sarcoidosis predominant DEGs, based on fold change: *DNASE2B* (Deoxyribonuclease 2 Beta) and *FAIM2* (Fas Apoptotic Inhibitory Molecule 2) (Figure 18). The protein encoded by *DNASE2B* is located in the lysosome, while FAIM2 is a plasma membrane protein for which there is evidence for localisation to the lysosome to facilitate phagolysosomal fusion (Hong *et al.*, 2020). Interestingly, the spindle-shaped bodies known as Hamazaki-Wesenberg bodies often found in lymph nodes of sarcoidosis patients have been shown to be giant lysosomes (Rosen, 2021).

The lysosome may also be a differentiating feature between TB and sarcoidosis. Although both conditions demonstrate overexpression of the lysosome and related pathway, direct comparison of TB and sarcoidosis finds the lytic vacuole membrane and lysosome pathway to be the only significantly underexpressed cellular component and KEGG pathway respectively in TB, using gene set enrichment analysis (Figure 35, Figure 36). It is well known that the lysosome is critical in antigen processing (Nakagawa and Rudensky, 1999), and Mtb can interrupt phagolysosome biogenesis (Vergne *et al.*, 2005). Therefore, the phagolysosome pathway is likely to be a pivotal point in the

host-pathogen interaction in TB, and is known to be manipulated by Mtb (Russell, 2003; Rohde *et al.*, 2007).

A defining characteristic of human TB is lung destruction and cavitation, leading to morbidity, mortality and transmission. Matrix metalloproteinases play a central role in lung matrix turnover (Ong and Friedland, 2014). *MMP1,* also known as interstitial collagenase, is among the top ten upregulated genes according to fold change in TB compared to control samples, of which *MMP1* is also one of the two DEGs exclusive to TB (Figure 18). Furthermore, *MMP1* was the most upregulated gene in TB compared to sarcoidosis, with the highest fold change (log_2 fold change 7.6) of all differentially expressed genes (Figure 24), and one of seven genes in Cluster 21, the only cluster unique to TB on correlation analysis (**Error! Reference source not found.**). MMP-1 degrades lung extracellular matrix in TB (Elkington and Friedland, 2011a), and notably, TB causes much more extensive lung matrix destruction than sarcoidosis.

The study of samples taken from treatment naïve patients has provided further evidence for the significant role of immune pathways and extracellular matrix organisation, particularly MMP-1, in TB pathogenesis. It also brings new insight into the key pathways contributing to the host response to Mtb, with unexpected findings such as sphingosine kinase 1 (Figure 24) and the lysosome pathway (Figure 36). However, the results of bioinformatic analysis are only associative (Byron *et al.*, 2016), and it is impossible to understand the underlying mechanism without either animal studies or advanced cell culture. The opportunity to use a biomimetic microsphere model of TB, incorporating human cells, allowed me to further study the host immune response to Mtb.

Chapter 4: Microsphere study

4.1 Introduction

A number of TB models exist, including the commonly used 2-dimensional (2D) cell culture model and a 3-dimensional (3D) cell culture model developed in the Elkington laboratory at the University of Southampton. Therefore, the microsphere study design consisted of peripheral blood mononuclear cells (PBMCs) from 6 donors, each undergoing 6 conditions (Table 12) to determine which model was most suited to then perform translational studies informed by the lymph node sample analysis.

2D Control	2D Mtb
3D Alginate Control	3D Alginate Mtb
3D Alginate-Collagen Control	3D Alginate-Collagen Mtb

Table 12: Microsphere study design in which PBMCs from each donor were divided into 6 experimental conditions

The 3D cellular model consists of alginate incorporated with collagen. To see the effect of collagen in the 3D model, alginate microspheres were divided into 2 conditions, with or without collagen.

4.2 Results

4.2.1 RNA extraction

A total of 36 samples (6 conditions for each of 6 donors) was obtained. The bioelectrospray system generated microspheres on 3 different days (2 donors per day) on Day 0. To release the cells from the matrix, decapsulation was performed on Day 4 and stored at -80 ^{O}C. Total RNA extraction was then performed, NanoDrop was used to measure the RNA concentration, and the Agilent 2100 Bioanalyzer system used to measure the RNA Integrity Number (RIN), displayed in Table 13.

	Mean	**Lowest**	**Highest**
Concentration (ng/µl)	182	43	461
RIN	8.7	6.1	10

Table 13: RIN and RNA concentration extracted from 36 cell culture samples comprising 6 healthy PBMC donors in 6 different cell culture conditions. *RNA yield was high with mean RIN 8.7.*

Total RNA from the microsphere samples were sent to GENEWIZ company for RNA sequencing, funded through a previous collaborative NIH grant with Jeanine D'Armiento.

4.2.2 Quality control

FastQC analysis of the raw sequenced data demonstrated sequence lengths to be 151 base pairs. All sequences were high quality, with all average Phred scores of 35 or more. Reads were paired-end, and the mean number of read counts was 32 million per sample (25 to 43 million read counts).

4.2.3 Kallisto alignment

The fraction of mapped reads was 74-86% (mean of 80%) and number of mapped reads was 20 to 35 million (mean of 26 million reads per sample). The mean average fragment length was 155 base pairs per sample (131 to 178 base pairs). The kallisto index was constructed in 2 ways: using the human transcriptome alone; and using the human transcriptome and H37Rv transcriptome. Mapping to either index gave exactly the same results, giving evidence that all sequences mapped to the human transcriptome alone.

4.2.4 STAR alignment

The fraction of uniquely mapped reads was 94-96% and mean number of uniquely mapped reads was 29 million per sample (23 to 40 million). Estimation of abundance was performed by featureCounts, resulting in “AllCounts” data whose gene expression analysis was carried out using edgeR.

4.2.5 Transcript abundance and analysis

Within-sample normalisation was performed by adjusting for transcript length (transcripts per million, TPM). The transcript abundance of the 36 different samples is displayed in Figure 37. Most samples have at least 20 million transcripts per sample, which likely reflects the high quality RNA that could be isolated from the cell culture models.

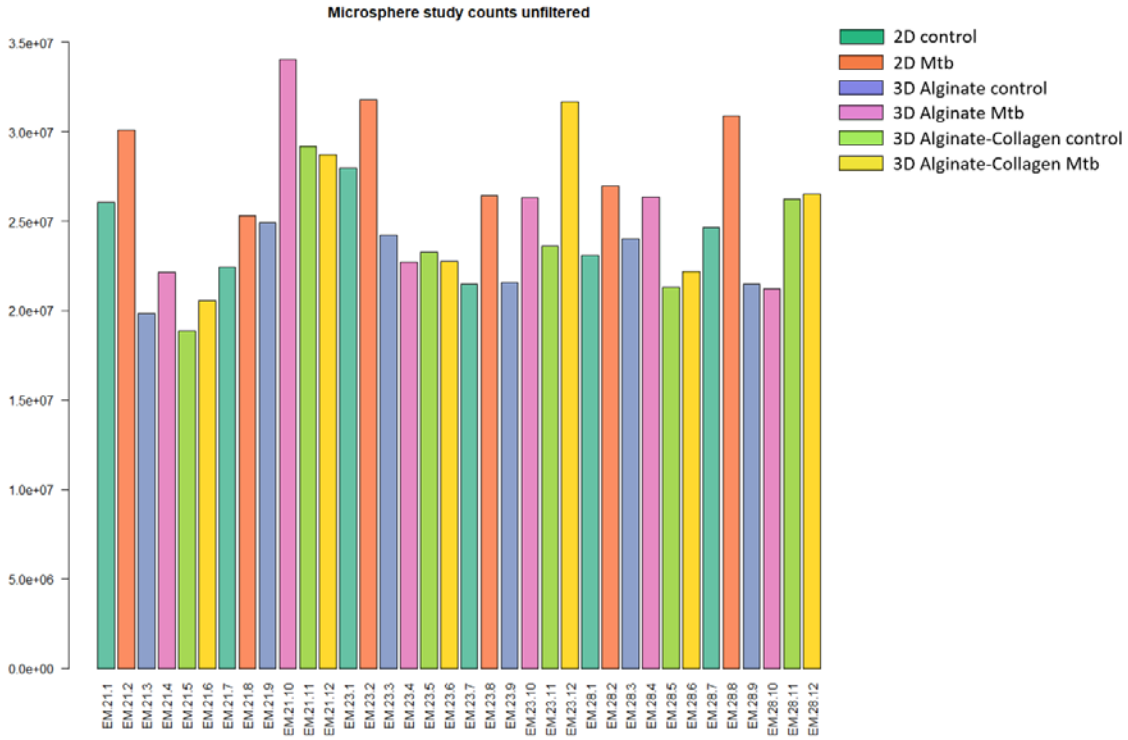


Figure 37: Transcript abundance of the 36 cell culture samples (transcripts per million). *High read counts are observed across all samples.*

Principle component analysis (PCA) was then performed according to the experimental condition (Figure 38), where a tendency for each experimental condition to group together is seen, though there is no clear overall separation between experimental conditions.

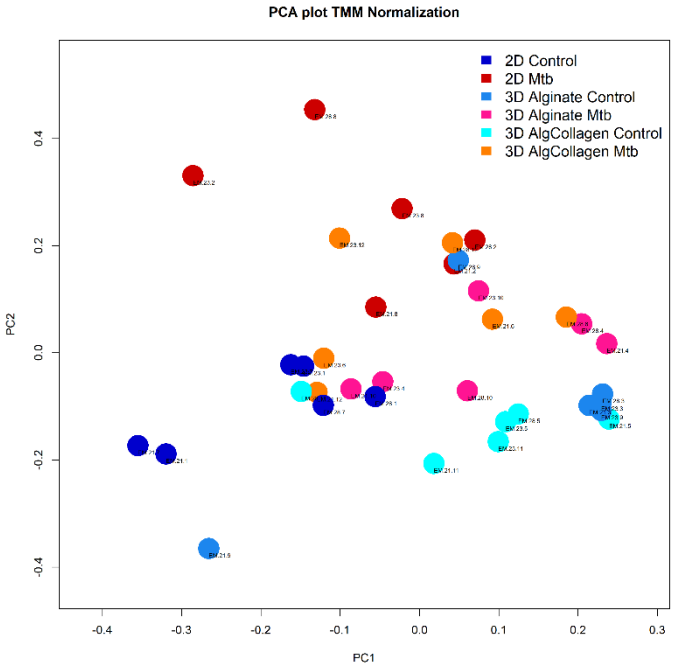


Figure 38: PCA of the cell culture samples from 6 cell culture conditions. *The 6 culture conditions do not clearly separate despite control and Mtb infected samples within each cell culture model showing a tendency to separate from each other.*

Other variables were also observed on PCA (Figure 39). No batch effect due to donor or day of microsphere generation was seen. Of note, the 3 different cell culture models overlap with each other. Moreover, uninfected and infected samples demonstrate a tendency to group separately but there is no clear separation between them. In particular, the control sample EM.28.9 appears to be a possible outlier. Therefore, GENEWIZ company performed a bacterial rRNA QC check to ascertain if EM.28.9 and its infected counterpart EM.28.10 had been accidentally swapped. However, control sample EM.28.9 had 0.04% reads mapping to bacterial rRNA, whereas infected sample EM.28.10 had 0.4% reads mapping to bacterial rRNA, confirming the sample identity.

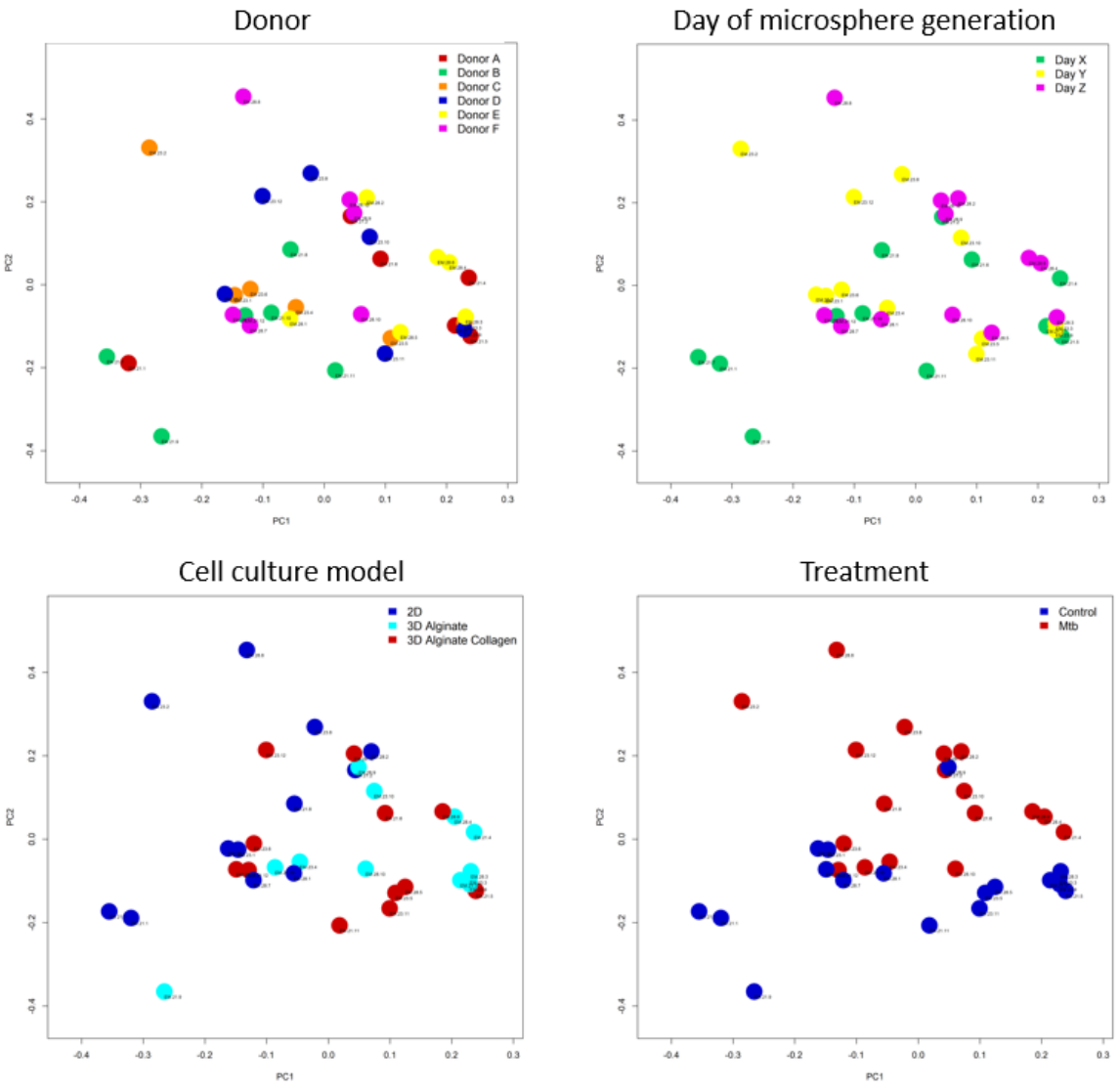


Figure 39: PCA of microsphere study samples grouped according to variables: Donor; Day of microsphere generation; Cell culture model; and treatment with Mtb. *Each variable shows an absence of batch effect.*

Hierarchical clustering also demonstrates a tendency for uninfected and infected cell culture models to group together, though no clear separation is seen with the top 1000 (Figure 40) or top 50 most

variable genes (Figure 41). It is noteworthy to find *MMP1*, a gene appearing to play a significant role in clinical TB sample analysis, is amongst the top 50 most variable genes here too.

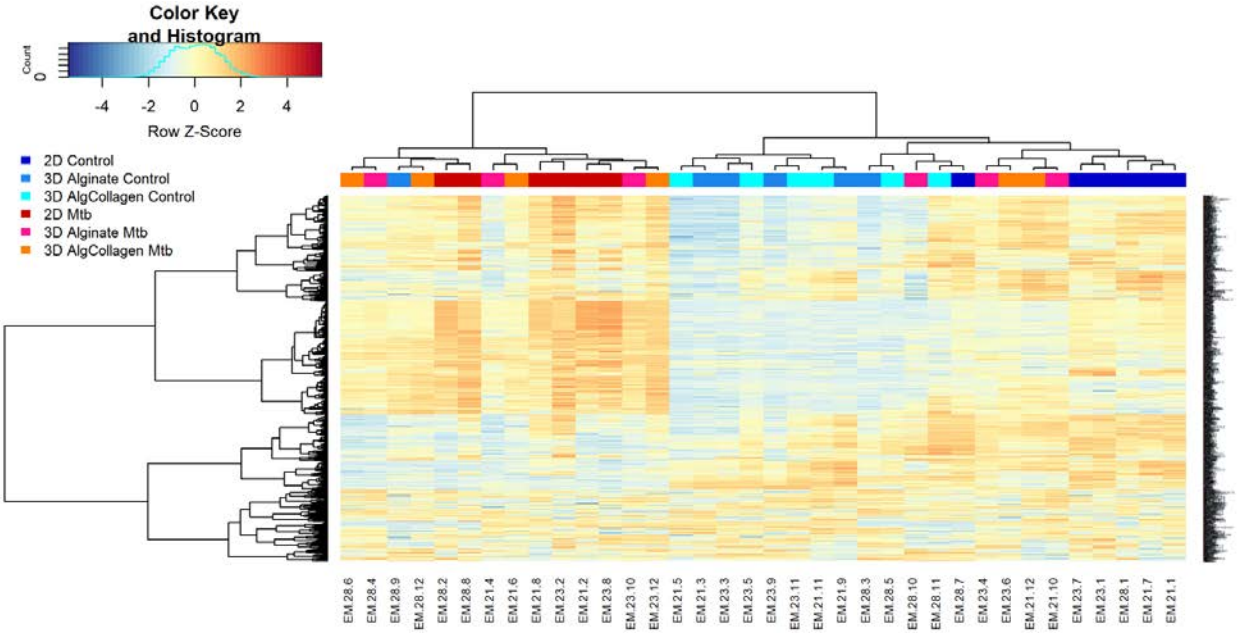


Figure 40: Hierarchical clustering heat map of top 1000 most variable genes across the 36 cell culture samples (Spearman correlation and ward D.2 linkage). *The 6 cell culture conditions do not clearly cluster, though uninfected and infected models tend to group together.*

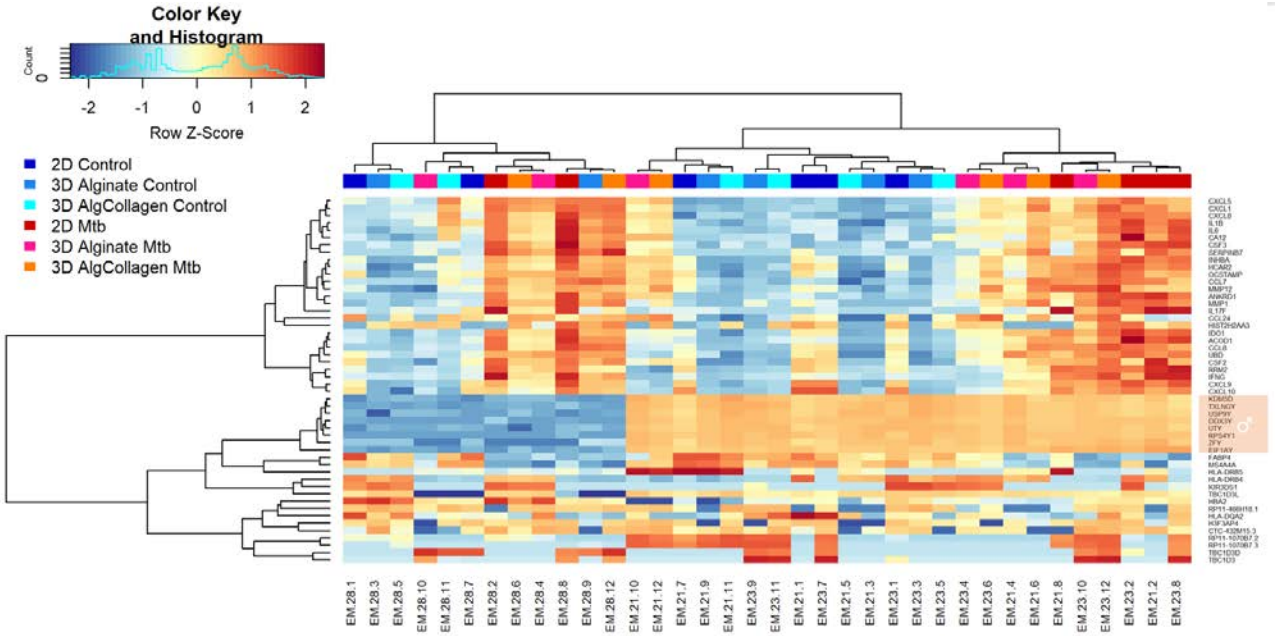


Figure 41: Hierarchical clustering heat map of top 50 most variable genes across the 36 cell culture samples (Spearman correlation and ward D.2 linkage). *The 6 cell culture conditions do not clearly cluster, though uninfected and infected models tend to group together.*

A closer look at the top 50 variable genes also shows those comprising the overexpressed band across the second main cluster to be Y chromosome linked (Figure 41). This finding demonstrates

genetic factors are likely to be important in the host response to Mtb, and interesting that the incidence of TB is twice as high in males than in females worldwide (Jimenez-Corona *et al.*, 2006).

4.2.6 Differential gene analysis

To understand the fundamental level of gene expression in each of the three cell culture models, each model was initially solely compared in its uninfected state. Differential gene expression was performed by limma, with 3 comparisons made between the uninfected models: uninfected 3D Alginate compared to uninfected 2D cell culture; uninfected 3D Alginate-Collagen compared to uninfected 2D cell culture; and uninfected 3D Alginate-Collagen compared to uninfected 3D Alginate. There was no differential gene expression between the two 3D models. The number of differentially expressed genes (DEGs) between the uninfected microsphere models and uninfected 2D model are shown below (Figure 42).

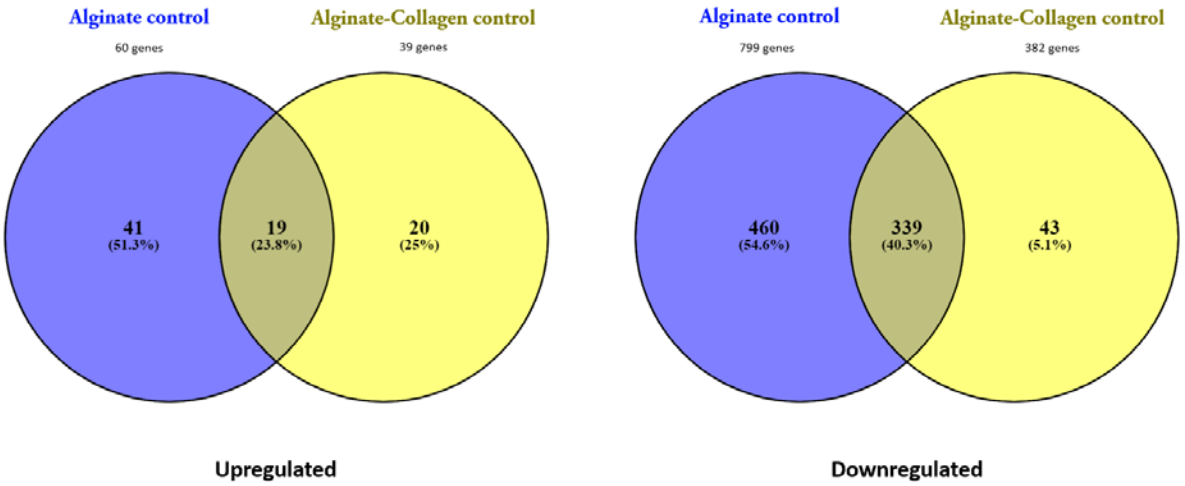


Figure 42: Venn diagram showing the number of DEGs identified on comparison of uninfected cells in 3D Alginate and 3D Alginate-Collagen with uninfected 2D cell culture model (absolute $log_2FC \geq 1.5$, adjusted P value < 0.05). *A majority of genes are downregulated, demonstrating higher gene expression in uninfected 2D cell culture.*

As expected, a large proportion of DEGs, 358 of the 922 genes, are shared between the two microsphere models. However, a far higher number of downregulated DEGs is seen in uninfected microspheres than upregulated, relative to uninfected 2D model, indicating a higher baseline gene expression in 2D culture. 3D Alginate-Collagen culture also displays fewer DEGs than 3D Alginate, showing 3D Alginate-Collagen is more similar than 3D Alginate to the 2D cell culture model.

The effect of Mtb infection on each of the three cell culture models was next analysed. Three separate comparisons were made: Mtb infected 2-dimensional (2D) cell culture compared to uninfected 2D cell culture; Mtb infected 3-dimensional (3D) Alginate microspheres compared to uninfected 3D Alginate microsphere cell culture; and Mtb infected 3-dimensional (3D) Alginate-

Collagen microspheres compared to uninfected 3D Alginate-Collagen microsphere cell culture. The number of differentially expressed genes (DEGs) is illustrated in Figure 43.

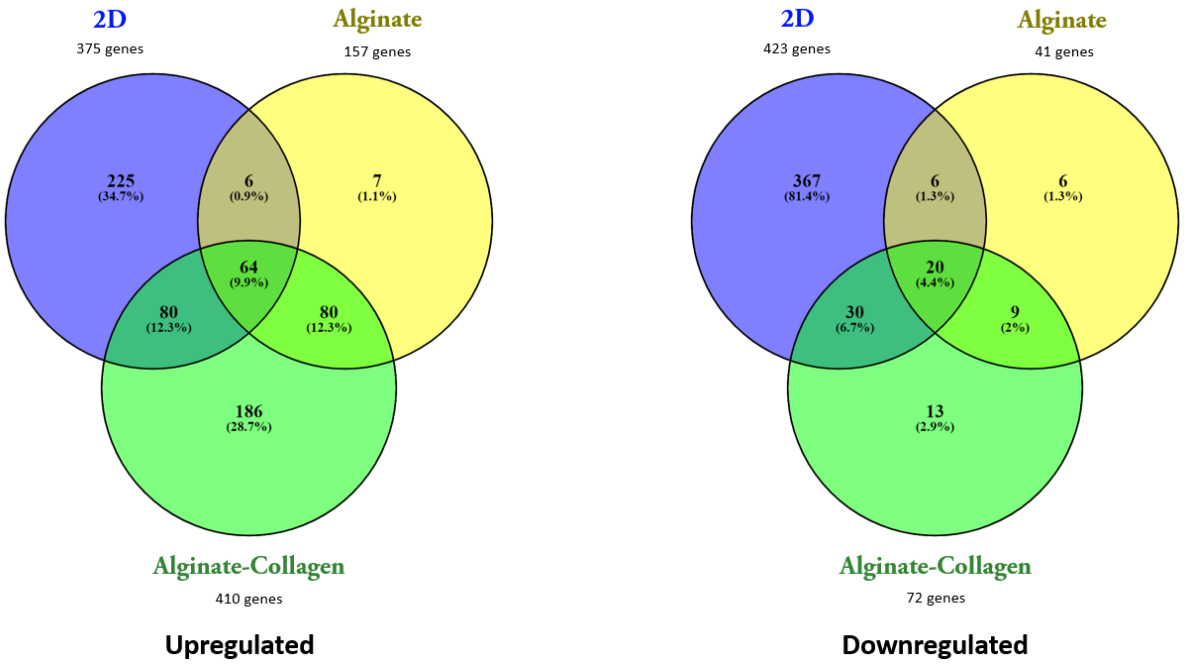


Figure 43: Venn diagram demonstrating the number of DEGs identified on comparison of Mtb infected and uninfected cells in 2D, 3D Alginate and 3D Alginate-Collagen cell culture models (absolute $log_2FC \geq 1.5$, adjusted P value < 0.05). *A majority of genes are seen in 2D model, of which more are shared with 3D Alginate-Collagen than with 3D Alginate model. The 3D Alginate model shows the fewest DEGs, most of which are shared with 3D Alginate-Collagen model.*

A total of 1226 DEGs were found on differential gene analysis. The 2D cell culture model has 944 DEGs, the highest number of DEGs out of the 3 different cell culture models. The 3D Alginate model demonstrates the lowest number of DEGs, of which 185 of its 198 DEGs are shared with the 2D and 3D Alginate-Collagen models. In contrast, 737 DEGs are found to be unique to the 2D model and 226 DEGs unique to the 3D Alginate-Collagen model. Direct comparisons between the infected cell culture models demonstrated no differential gene expression between infected 3D Alginate and infected 3D Alginate-Collagen.

The remaining comparisons are illustrated below, where infected 3D Alginate is compared to infected 2D cell culture (Figure 44, blue), and infected 3D Alginate-Collagen is compared to infected 2D cell culture (Figure 44, yellow). Both infected 3D Alginate and infected 3D Alginate-Collagen exhibit many more upregulated than downregulated DEGs when compared to the infected 2D culture model, of which 297 of the 356 DEGs (83%) seen in infected 3D Alginate are shared with infected 3D Alginate-Collagen. In contrast, infected 3D Alginate-Collagen demonstrates 210 of its 507 (41%) upregulated DEGs to be unique to itself.

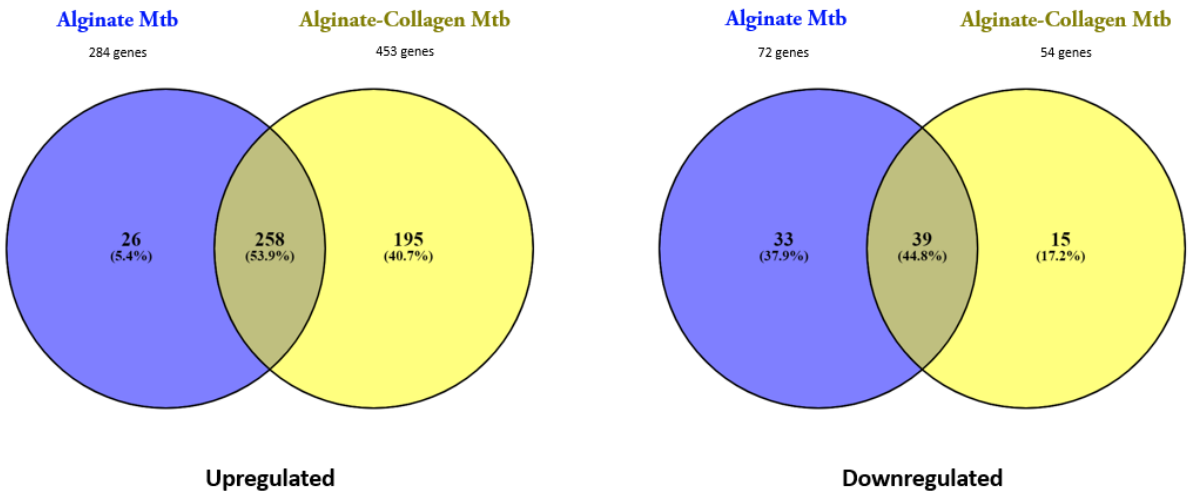


Figure 44: Venn diagram showing the number of DEGs identified on comparison of Mtb infected microsphere models with Mtb infected 2D model (absolute $log_2FC \geq 1.5$, adjusted P value < 0.05). *Of the 3 cell culture models during Mtb infection, the 3D Alginate-Collagen model displays the highest number of upregulated genes.* Collagen is a major constituent of the lung extracellular matrix. The addition of collagen to the microspheres demonstrates more biological variation and subsequent differential gene expression when compared to 2D cell culture, which is likely to have biological consequences. Therefore, further analysis was performed to understand the biological processes and pathways prominent in each cell culture model.

4.2.7 Correlation analysis

An alternative method to perform gene correlation is Markov Clustering (MCL), where genes are grouped together based on their direction and degree of change. Genes whose behaviour correlate with each other form a cluster which can be illustrated using tools such as Graphia Pro. All differentially expressed genes (DEGs) identified by limma on comparison of each infected cell culture model with their uninfected state underwent correlation analysis. In total, 1209 DEGs were analysed by MCL. A Pearson's correlation coefficient value of 0.88 was chosen because it produced the clearest depiction of gene correlation, where 16 clusters displayed any gene expression values (**Figure 45**).

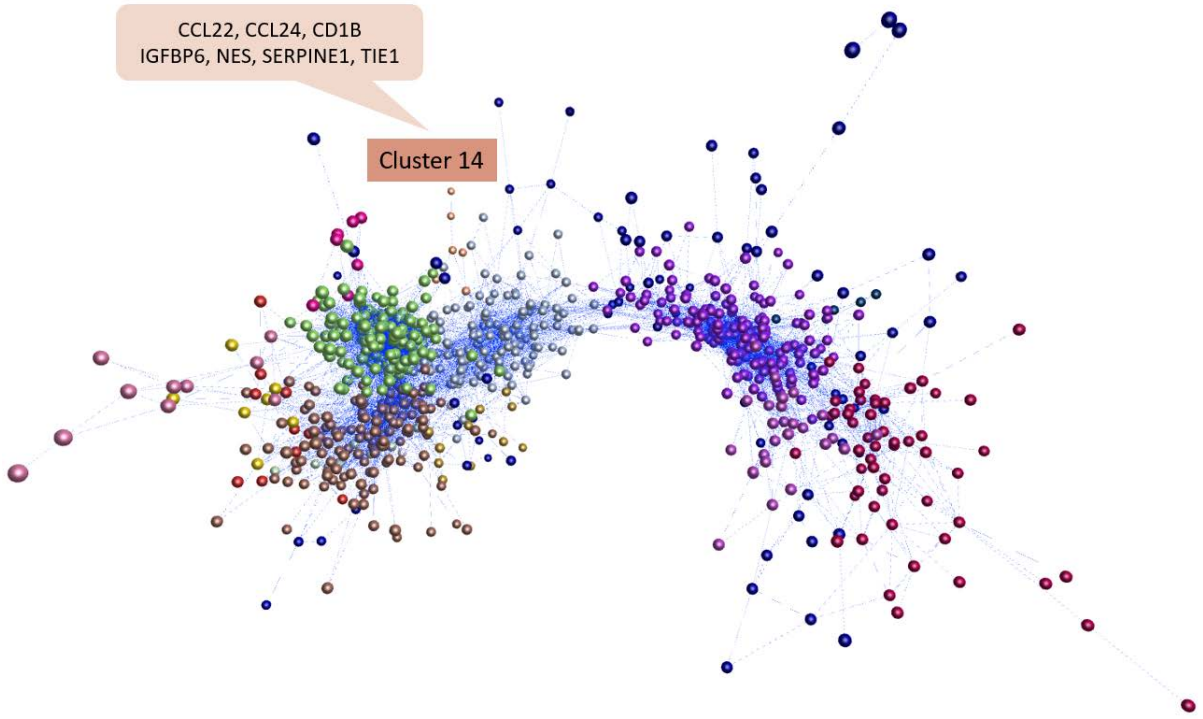


Figure 45: Correlation analysis performed utilising Markov Cluster Algorithm (Pearson's R of 0.88) with genes of absolute $log_2FC \geq 1.5$ and adjusted P value < 0.05 during the comparison of each of the 3 cell culture models during Mtb infection relative to their uninfected state, illustrated by Graphia Pro. *Cluster 14 is the only cluster underexpressed in uninfected and overexpressed during Mtb infection in the microsphere models.*

In Figure 45, upregulated DEGs are seen in the left half and downregulated DEGs seen in the right half of the illustration. The main cluster of interest is Cluster 14, the only cluster to be underexpressed in uninfected microspheres and overexpressed with Mtb infection, particularly in the 3D Alginate-Collagen model (Figure 46). This cluster also demonstrates moderate expression in the 2D model though is unaffected by Mtb infection. Cluster 14 comprises 7 genes: *CCL22* (C-C motif chemokine ligand 22); *CCL24* (C-C motif chemokine ligand 24); *CD1B* (CD1b Molecule); *IGFBP6* (Insulin like growth factor binding protein 6); *NES* (Nestin); *SERPINE1* (Serpin Family E Member 1); and *TIE1* (Tyrosine Kinase With Immunoglobulin Like And EGF Like Domains 1).

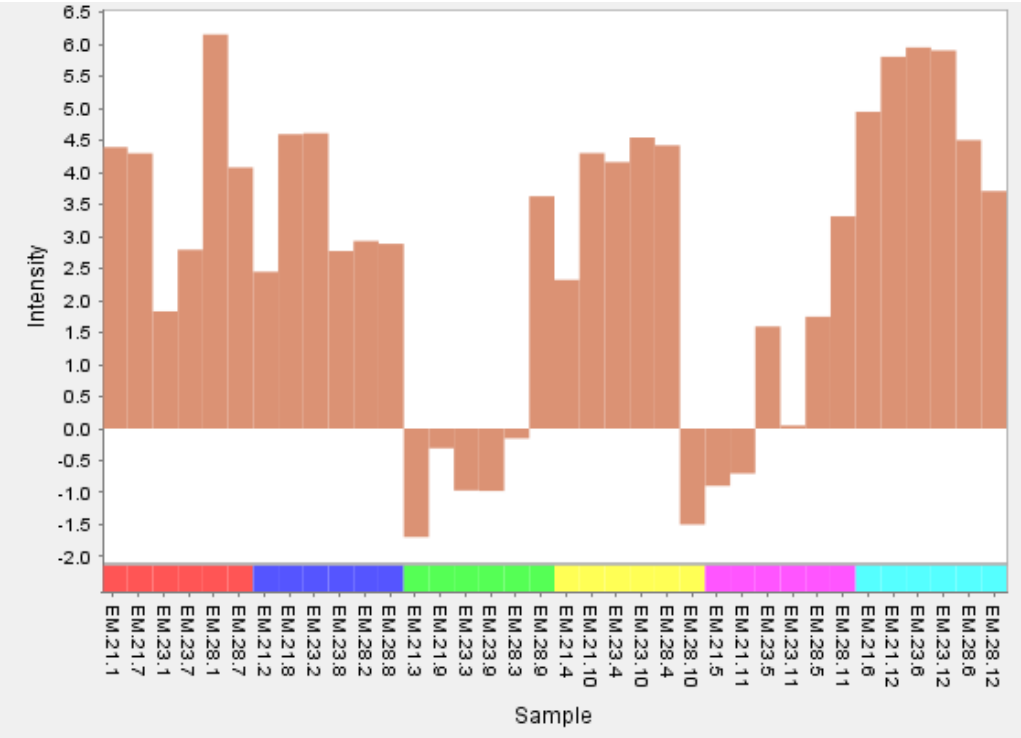


Figure 46: Normalised gene expression of Cluster 14 determined by Markov Clustering, using Pearson's R of 0.88, analysed with Graphia Pro. *Uninfected 2D (red), infected 2D (blue), uninfected 3D Alginate (green), infected 3D Alginate (yellow), uninfected 3D Alginate-Collagen (pink) and infected 3D Alginate-Collagen (turquoise) are shown. Cluster 14 is underexpressed in uninfected microspheres and overexpressed in Mtb infected microsphere models.*

Gene ontology performed on these seven genes generated multiple biological processes (Table 14) and pathways (Table 15) with significance after correcting for FDR.

ID	Name	FDR B&H	Genes from input	Genes in annotation
GO:0002548	monocyte chemotaxis	6.53E-04	3	68
GO:0071674	mononuclear cell migration	9.86E-04	3	98
GO:0048245	eosinophil chemotaxis	3.27E-03	2	22
GO:0071347	cellular response to interleukin-1	3.27E-03	3	184
GO:0072677	eosinophil migration	3.33E-03	2	26
GO:0070555	response to interleukin-1	3.33E-03	3	212
GO:0045766	positive regulation of angiogenesis	3.33E-03	3	226
GO:0097529	myeloid leukocyte migration	3.33E-03	3	235
GO:0030595	leukocyte chemotaxis	3.33E-03	3	243
GO:1904018	Positive regulation of vasculature development	3.38E-03	3	253

Table 14: Top 10 biological processes identified on GO analysis of Cluster 14 using ToppFun. *Biological processes overexpressed in Mtb infected microspheres alone associate with immune cell chemotaxis, response to IL-1 and angiogenesis.*

ID	Name	FDR B&H	Genes from input	Genes in annotation
M5889	Ensemble of genes encoding extracellular matrix and extracellular matrix-associated proteins	1.30E-02	4	1026
PW:0000356	Oxygen homeostasis	1.30E-02	1	1
M5885	Ensemble of genes encoding ECM-associated proteins including ECM-affiliated proteins, ECM regulators and secreted factors	2.86E-02	3	751
M4844	Chemokine signaling pathway	2.86E-02	2	189
M9809	Cytokine-cytokine receptor interaction	3.02E-02	2	265
M2842	Fibrinolysis Pathway	3.02E-02	1	12
PW:0000238	Insulin-like growth factor signaling	3.02E-02	1	12
M22016	Ghrelin: Regulation of food intake and energy homeostasis	3.02E-02	1	13
1269372	Dissolution of fibrin clot	3.02E-02	1	13
M6487	Platelet amyloid precursor protein pathway	3.02E-02	1	14

Table 15: Top 10 pathways identified on GO analysis of Cluster 14 using ToppFun. *Pathways overexpressed in Mtb infected microspheres alone mainly associate with extracellular matrix organisation, oxygen homeostasis, chemokine signalling, and fibrinolysis.*

Correlation analysis suggests important roles for regulation of the inflammatory response including immune cell chemotaxis and response to IL-1, angiogenesis, extracellular matrix organisation, oxygen homeostasis and fibrinolysis in response to Mtb infection in 3D cell culture. I returned to differential gene expression to corroborate these findings.

4.2.8 Gene Ontology: Uninfected cell culture model comparison

To understand the fundamental gene expression in each of the three cell culture models, each model was compared in its uninfected state. The differential gene expression performed by limma underwent gene ontology (GO) analysis using the ReactomePA software tool, with Benjamini and Hochberg's false discovery rate (FDR) of less than 0.05. ToppGene (FDR cutoff of 0.05) and g:Profiler (adjusted P value threshold of 0.05) tools were also used to confirm the results produced by ReactomePA software. The top 20 most upregulated biological processes and top 20 most upregulated REACTOME pathways were identified in two separate comparisons: uninfected 3D Alginate compared to uninfected 2D cell culture; and uninfected 3D Alginate-Collagen compared to uninfected 2D cell culture (Table 16).

	Uninfected 3D Alginate	Uninfected 3D Alginate-Collagen
Up	RNA catabolism	RNA catabolism
	Protein translation	Protein translation
	Protein localisation to ER	Protein localisation to ER
	Selenocysteine synthesis	Selenocysteine synthesis
	Influenza viral RNA transcription and replication	Influenza viral RNA transcription and replication
	rRNA processing in the nucleus and cytosol	rRNA processing in the nucleus and cytosol
		Regulation of expression of SLITs and ROBOs
Down	Neutrophil degranulation	Neutrophil degranulation
	Ag processing-Cross presentation	Ag processing-Cross presentation
	TNFR2 non-canonical NF-κB pathway	TNFR2 non-canonical NF-κB pathway
	ER-phagosome pathway	ER-phagosome pathway
	Cell cycle	Cell cycle
	Apoptosis	Apoptosis
	Signaling by IFNα/β	Signaling by IFNα/β
	Signaling by interleukins	Exocytosis
	Response to IFNγ	
	Mitochondrion organisation	
	Nucleoside triphosphate metabolism	
	Defective CFTR causes cystic fibrosis	
	Vif-mediated degradation of APOBEC3G	

Table 16: Top biological processes and REACTOME pathways in the comparison of uninfected 3D models with uninfected 2D model according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *In their uninfected states, 3D cell culture overall demonstrates lower levels of gene expression than 2D cell culture, including lower expression of immune-related pathways.*

Both uninfected microsphere models demonstrate higher levels of activity related to transcription than the uninfected 2D model, particularly the 3D Alginate-Collagen model. Interestingly, the presence of collagen in the 3D Alginate-Collagen model may also permit more cellular development, as shown by the regulation of expression of SLITS and ROBOs, a REACTOME pathway involved in developmental biology. However, the baseline activity of uninfected 2D culture is higher than in uninfected 3D culture, exhibited by a diverse range of processes and pathways downregulated in the microsphere model, particularly in the 3D Alginate model.

4.2.9 Gene Ontology: Uninfected and infected cell culture model comparison

Gene ontology was utilised to study the differences in each cell culture model during Mtb infection. The same methods were used as described earlier in *4.2.8 Gene Ontology: Uninfected cell culture*

model comparison. Each infected model was compared to its uninfected state, and is summarised below (Table 17).

	Mtb in 2D culture	Mtb in 3D culture	Mtb in 3D culture with collagen
Up	Cell cycle	Cell cycle	Cell cycle
	Regulation of cytoskeleton	ER-phagosome pathway	ER-phagosome pathway
		Class I MHC Ag presentation	Ag processing and presentation
		CLEC7A (Dectin-1) signaling pathway	Transcription including that activated by Dectin-1
		Signaling by and response to IFNγ	Signaling by and response to IFNγ
		Signaling by and response to IL-1	Signaling by and response to IL-1
		Apoptosis	Apoptosis
		Host interaction of HIV factors	HIV infection of primary cell populations
		Protein metabolism	Protein catabolism
		TNFR2 non-canonical NF-κB pathway	TNFR2 non-canonical NF-κB pathway
		Response to IL-12	Neutrophil degranulation
Down	Neutrophil degranulation	-	ROBO receptors which regulate cell migration
	Vesicle-mediated transport		RNA metabolism
	Class II MHC Ag presentation		Protein translocation
	Signaling by TLR 2/4		Selenocysteine synthesis
	Signaling by PD-1		Influenza transcription and replication
	Signaling by Receptor tyrosine kinase		
	Angiogenesis		
	Platelet activation		
	IP3 and IP4 metabolism		
	Monocarboxylic acid metabolism		
	Lipid catabolism		
	Carbohydrate catabolism		

Table 17: Top biological processes and REACTOME pathways in the comparison of Mtb infected cell culture models with their uninfected state according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Most pathways are upregulated in 3D models and downregulated in 2D models, including immune and phagosome-related pathways.*

Mtb infection upregulates the cell cycle processes and pathways in all three models. Both microsphere models upregulate many processes and pathways in the presence of Mtb, most of which are shared such as the ER-phagosome pathway, signalling involving IFNγ and IL-1, and apoptosis. Similarly, no processes or pathways are downregulated in the 3D Alginate model during Mtb infection, while only a few pathways are downregulated in 3D Alginate-Collagen. In stark contrast, infected 2D cell culture mostly downregulate processes and pathways, including vesicle-mediated transport, PD-1 signaling, and angiogenesis, all of which are likely to play an important role in human TB pathogenesis.

4.2.10 Gene Ontology: Infected cell culture model comparison

Lastly, gene ontology was used to study the effect of collagen in Mtb infection by directly comparing the three infected cell culture models. The same methods were used as described earlier in *4.2.8 Gene Ontology: Uninfected cell culture model comparison*. The top 20 most upregulated biological processes and top 20 most upregulated REACTOME pathways were identified in 2 separate comparisons: infected 3D Alginate compared to infected 2D cell culture (Figure 47); and infected 3D Alginate-Collagen compared to infected 2D cell culture (Figure 48). Comparisons between each cell culture model in its infected state were made to help determine which cell culture model of TB is most likely to reflect *in vivo* conditions during Mtb infection. Each result is detailed in graph format, and summarised in Table 18. Both comparisons show significantly upregulated biological processes relating to the immune response and angiogenesis. In addition, infected 3D Alginate-Collagen demonstrates upregulation in cell motility and phagocytosis when compared to infected 2D cell culture.

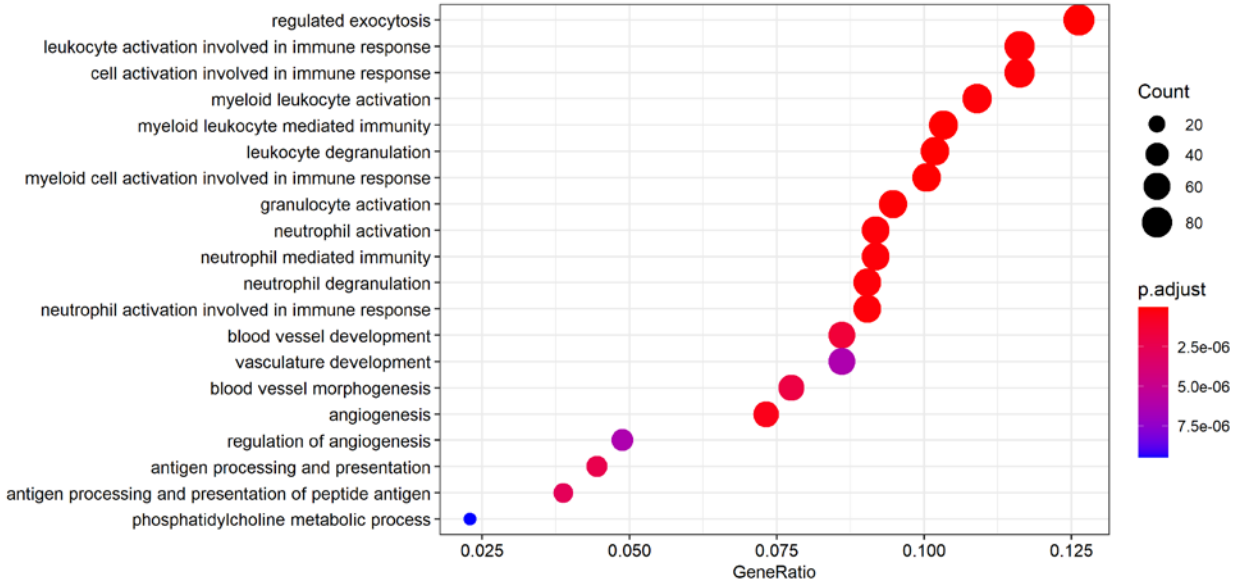


Figure 47: Top 20 biological processes upregulated in the comparison of infected 3D Alginate with infected 2D cell culture according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Biological processes relate to exocytosis, the immune response including antigen presentation, and angiogenesis.*

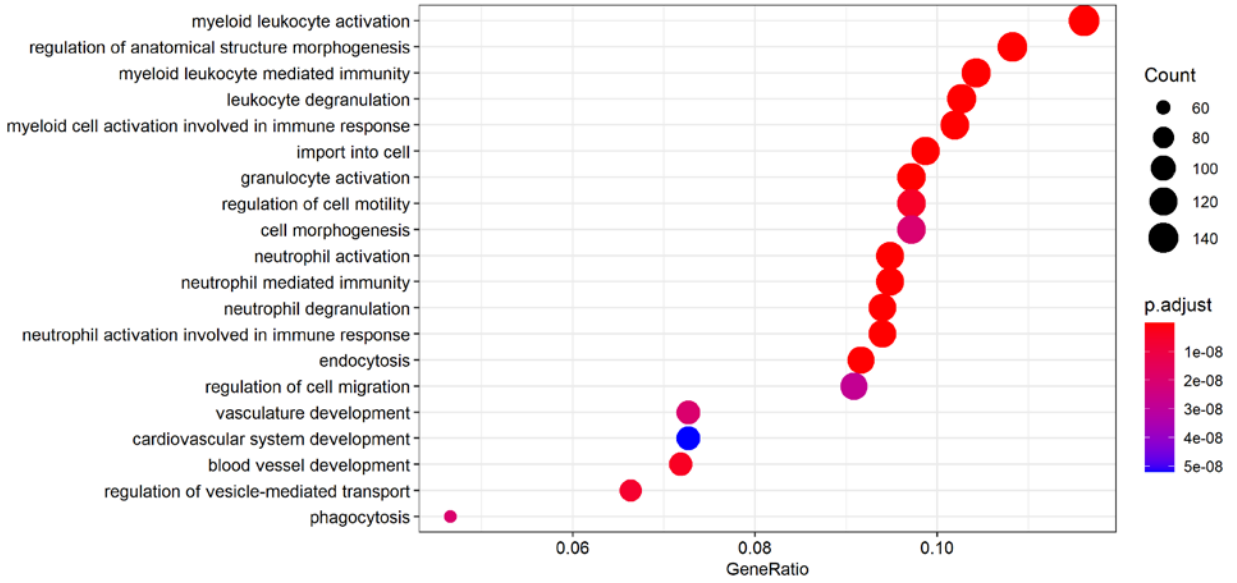


Figure 48: Top 20 biological processes upregulated in the comparison of infected 3D Alginate-Collagen with infected 2D cell culture according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Biological processes relate to the immune response, phagocytosis and angiogenesis.*

More difference is evident between the culture model comparisons when observing upregulated REACTOME pathways. Comparison of infected 3D Alginate with the infected 2D model shows upregulated REACTOME pathways associated with the immune response, platelet activation, glycerophospholipid synthesis (Figure 49), and cellular respiration.

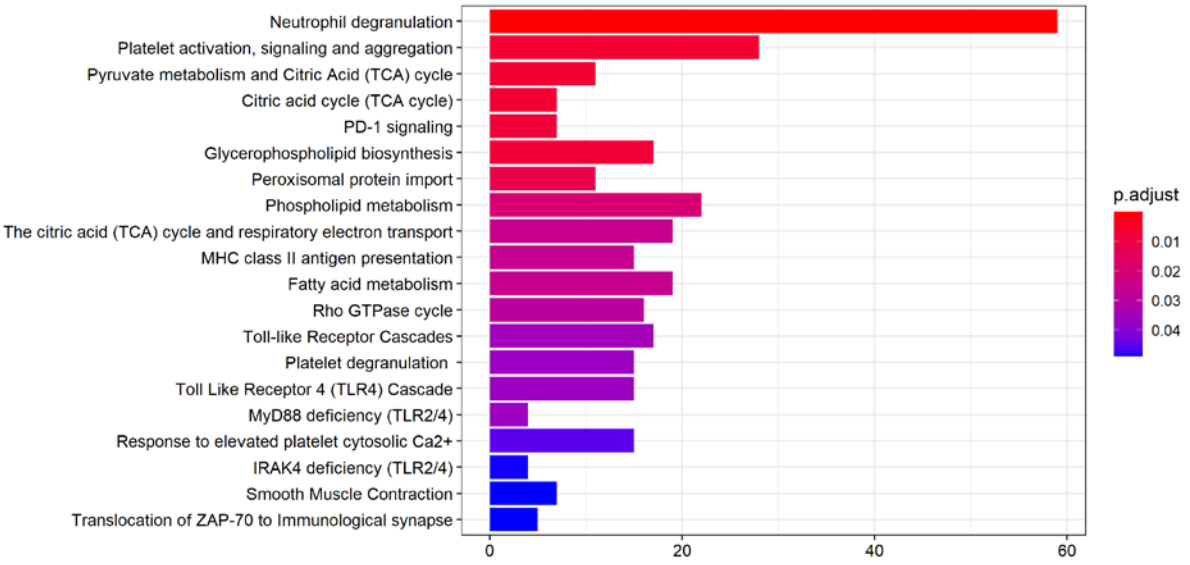


Figure 49: Top 20 REACTOME pathways upregulated in the comparison of infected 3D Alginate with infected 2D cell culture according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Pathways relate to the immune response, plate activation, glycerophospholipid turnover and cellular respiration.*

However, comparison of infected 3D Alginate-Collagen with infected 2D cell culture shows more diversity in upregulated REACTOME pathways and are associated with the immune response, platelet activation, receptor tyrosine kinases, vascular endothelial growth factor (VEGF) and semaphorin signalling, cellular respiration and insulin receptor recycling (Figure 50).

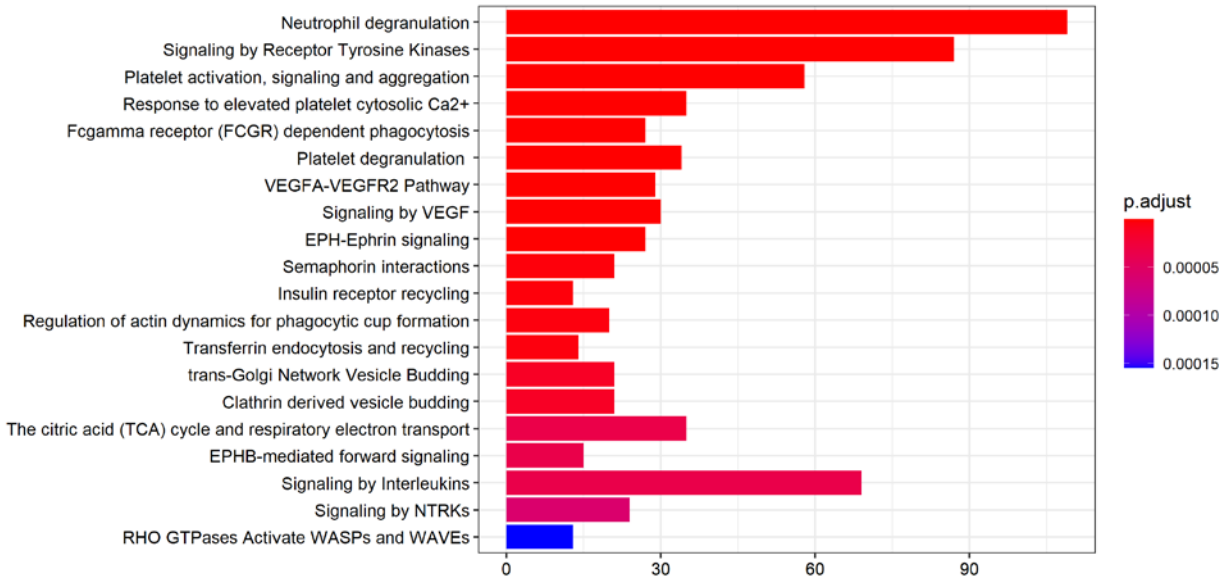


Figure 50: Top 20 REACTOME pathways upregulated in the comparison of infected 3D Alginate-Collagen with infected 2D cell culture according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Pathways relate to the immune response,platelet activation and cellular respiration, as well as phagocytosis, and signalling by receptor tyrosine kinases, VEGF and semaphorins.*

The same method was applied to study the biological processes and pathways downregulated in the infected microsphere models relative to infected 2D model. The top 20 most downregulated biological processes and top 20 most downregulated REACTOME pathways were identified in 2 separate comparisons: infected 3D Alginate compared to infected 2D cell culture (Figure 51); and infected 3D Alginate-Collagen compared to infected 2D cell culture (Figure 52). Less diversity in biological processes is seen downregulated than upregulated when the infected cell culture comparisons are made. Downregulation in biological processes associated with immune diversification, lymphocyte activation and cell cycling are seen when infected 3D Alginate is compared to infected 2D cell culture (Figure 51), whereas biological processes related to transcription and translation are downregulated when infected 3D Alginate-Collagen is compared to the infected 2D model (Figure 52). However, both comparisons clearly demonstrate downregulation in cell cycle associated processes.

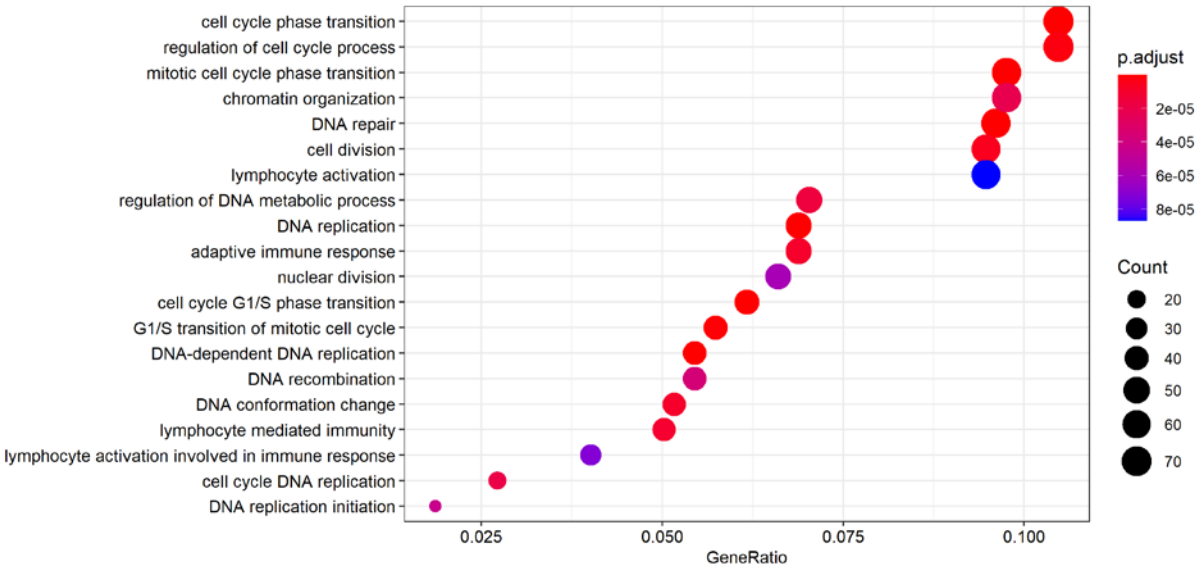


Figure 51: Top 20 biological processes downregulated in the comparison of infected 3D Alginate with infected 2D cell culture according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Biological processes relate to lymphocyte activation and cell division.*

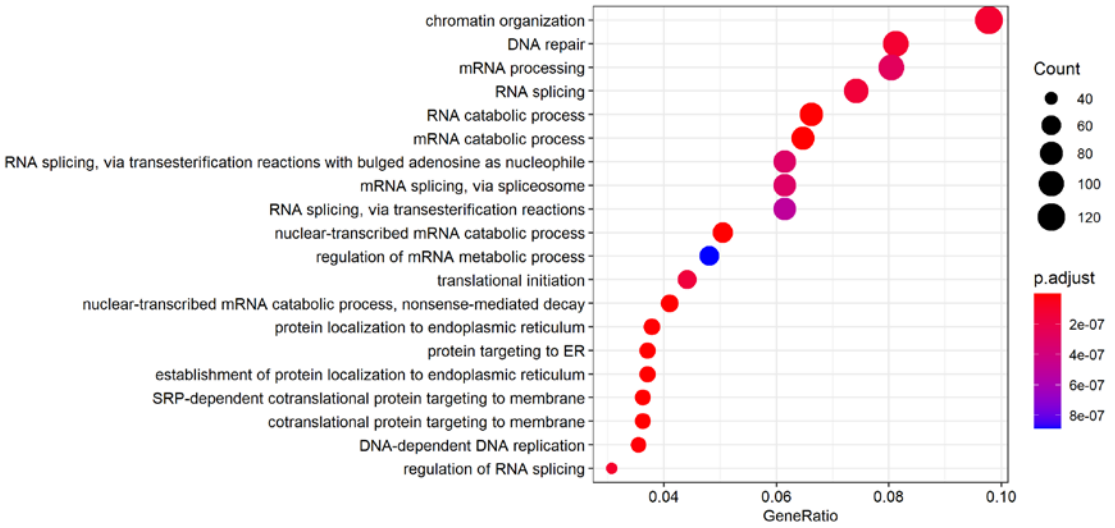


Figure 52: Top 20 biological processes downregulated in the comparison of infected 3D Alginate-Collagen with infected 2D cell culture according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Biological processes relate to transcription and protein translation.*

Similarly, all downregulated REACTOME pathways involve the cell cycle and related checkpoints, transcription and cellular senescence when infected 3D Alginate is compared to infected 2D cell culture (Figure 53).

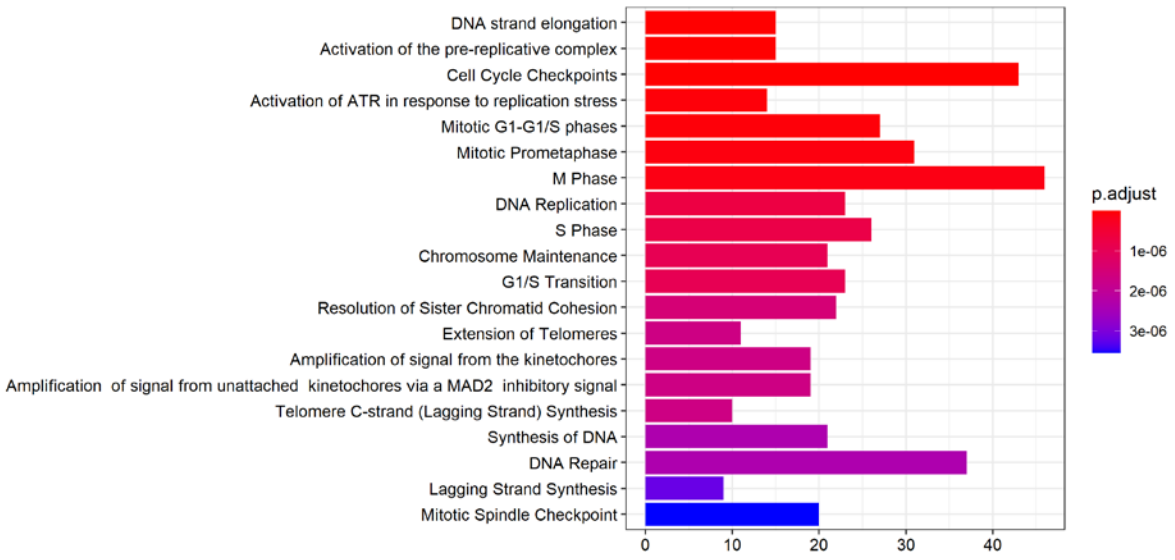


Figure 53: Top 20 REACTOME pathways downregulated in the comparison of infected 3D Alginate with infected 2D cell culture according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *All pathways relate to cell cycling.*

In addition, almost all the downregulated REACTOME pathways in the comparison of infected 3D Alginate-Collagen with the infected 2D culture model relate to the cell cycle, as well as transcription and protein translation (Figure 54). A summary of the top dysregulated biological processes and REACTOME pathways in Mtb infected 3D models relative to the infected 2D model is also provided (Table 18).

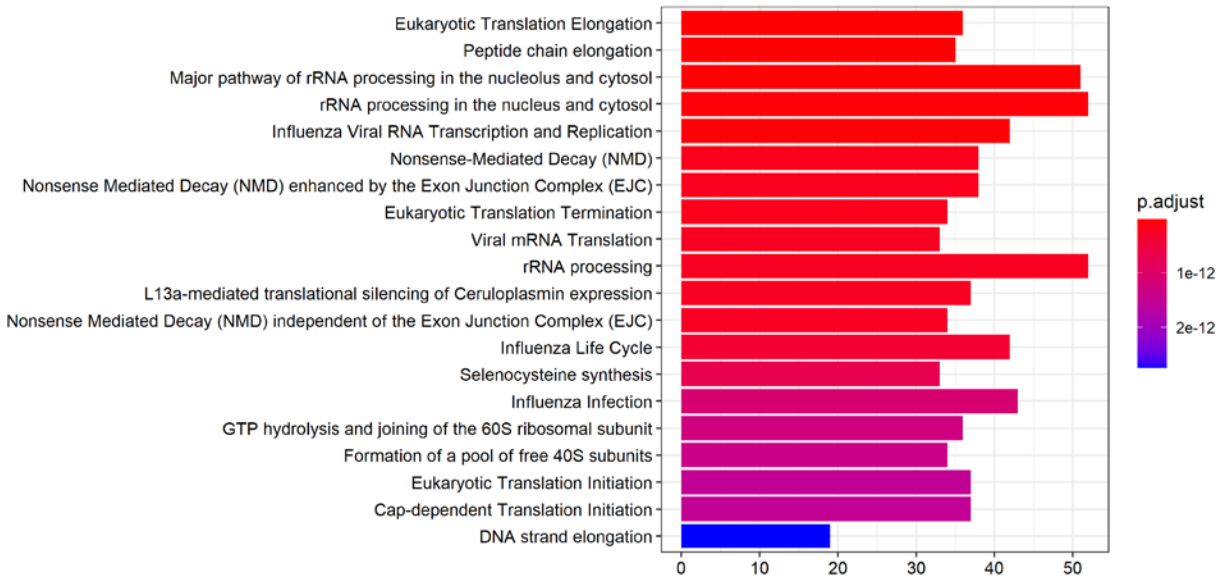


Figure 54: Top 20 REACTOME pathways downregulated in the comparison of infected 3D Alginate-Collagen with infected 2D cell culture according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Pathways relate to cell cycling, transcription and translation.*

	Infected 3D Alginate	Infected 3D Alginate-Collagen
Up	Angiogenesis	Angiogenesis
	Neutrophil degranulation	Neutrophil degranulation
	Platelet activation	Platelet activation
	TCA cycle	TCA cycle
	Exocytosis	FCγ-dependent phagocytosis
	MHC Class II Ag presentation	Transferrin endocytosis
	Glycerophospholipid biosynthesis	Actin polymerisation
	Fatty acid metabolism	Cell motility
	Signaling by PD-1	Cell migration
	Signaling by TLR2/4	Signaling by EPH, leading to cell migration
		Signaling by VEGF
		Signaling by interleukins
		Signaling by Receptor Tyrosine Kinases
Down	Cell cycle	Cell cycle
	Lymphocyte activation	RNA metabolism
		Protein translocation
		Protein translation
		Selenocysteine synthesis
		Viral mRNA translation

Table 18: Top biological processes and REACTOME pathways in the comparison of infected 3D cell culture models with infected 2D model according to adjusted P value, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *During Mtb infection, 3D models upregulate many pathways associated with immune response, cell signalling and angiogenesis, with more pathways seen in the 3D Alginate-Collagen model.*

In summary, direct comparison of Mtb infected cell culture models using gene ontology analysis demonstrates higher expression of more diverse process and pathways in 3D culture, particularly in the presence of collagen. These pathways include angiogenesis, phagocytosis, cell motility and cell signalling by interleukins and tyrosine kinases, which are likely to be significant processes in modelling human TB in a cellular system.

4.2.11 Gene Set Enrichment Analysis

A final method of bioinformatic analysis, gene set enrichment analysis (GSEA), was utilised to study the effect of Mtb infection in each cell culture model. The fgsea package was used to analyse the differential gene expression data performed by limma. The canonical pathways gene set was selected for its wide and diverse range of pathways. The top 10 canonical pathways with greatest normalised enrichment score (NES), both positive and negative, are illustrated. During Mtb infection of the 2D model, 9 of the 10 canonical pathways are overexpressed and relate to the cell cycle and extracellular matrix (Figure 55).

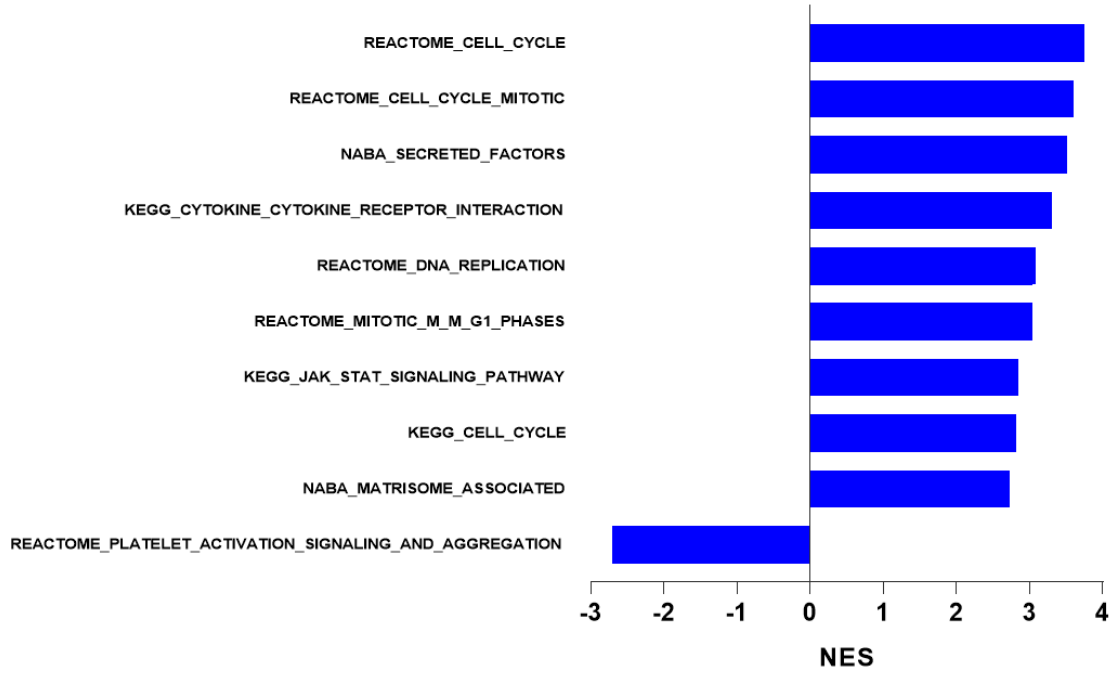


Figure 55: Top 10 canonical pathways with adjusted P value < 0.05 in the Mtb infected 2D model compared to it's uninfected state according to most extreme NES, analysed by fgsea (R environment). *All but 1 pathway is overexpressed and relate to cell cycling and the extracellular matrix.*

In contrast, Mtb infection of the 3D Alginate model results in only 5 canonical pathways with adjusted P values less than 0.05. All 5 pathways are associated with the extracellular matrix and cytokine signalling (Figure 56).

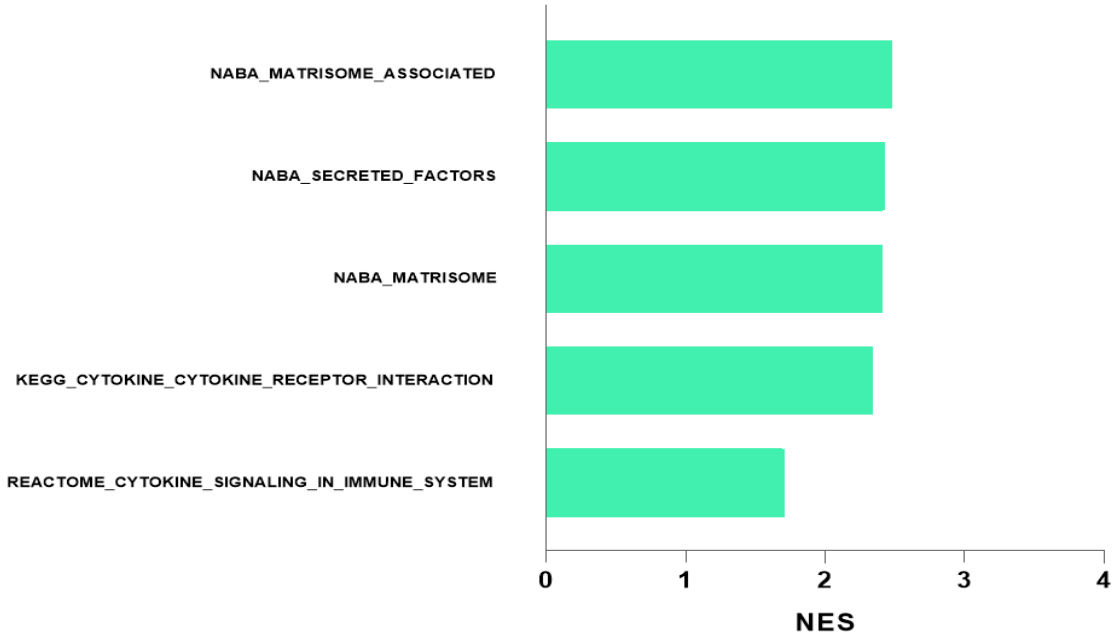


Figure 56: Top 10 canonical pathways with adjusted P value < 0.05 in Mtb infected 3D Alginate model compared to it's uninfected state according to most extreme NES, analysed by fgsea (R environment). *Only 5 pathways are seen, all overexpressed and relating to the extracellular matrix and signalling by cytokines.*

Similarly, the 3D Alginate-Collagen model demonstrates all top 10 canonical pathways to be overexpressed and to again associate with the extracellular matrix and cytokine signalling. However, many more significant pathways are overexpressed in the 3D Alginate-Collagen model and more specifically include chemokine and G protein-coupled receptor (GPCR) signalling (Figure 57).

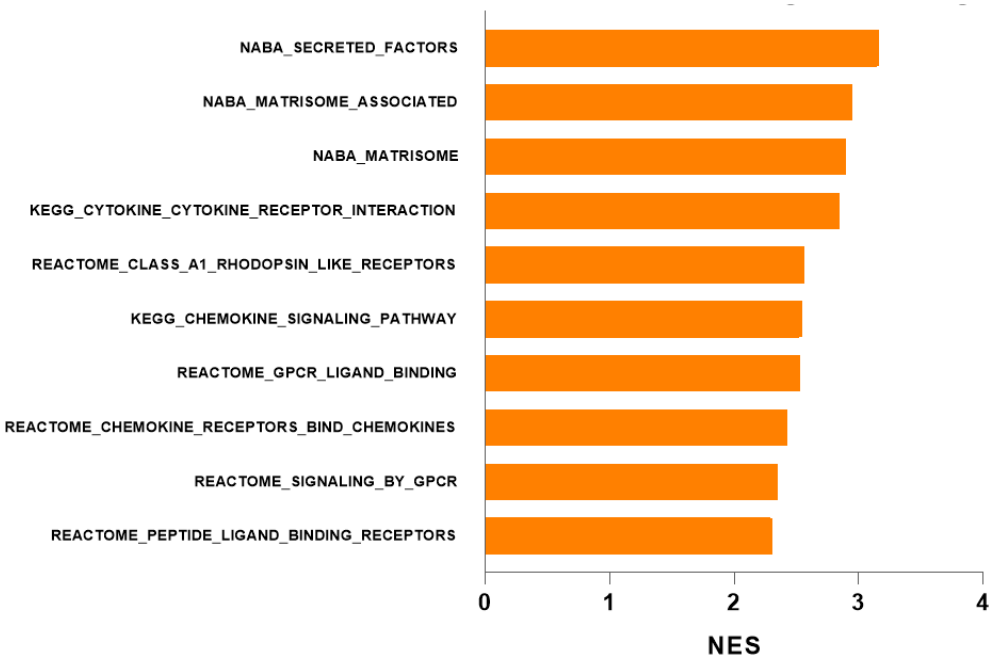


Figure 57: Top 10 canonical pathways with adjusted P value < 0.05 in Mtb infected 3D Alginate-Collagen model compared to its uninfected state according to most extreme NES, analysed by fgsea (R environment). *All 10 pathways are overexpressed and relate to the extracellular matrix and cytokine-receptor signalling.*

Another GSEA tool is the egsea package, which combines 11 established methods for GSEA, and was utilised to compare the difference in cellular component activity in each of the three cell culture models in their infected relative to uninfected state (Figure 58).

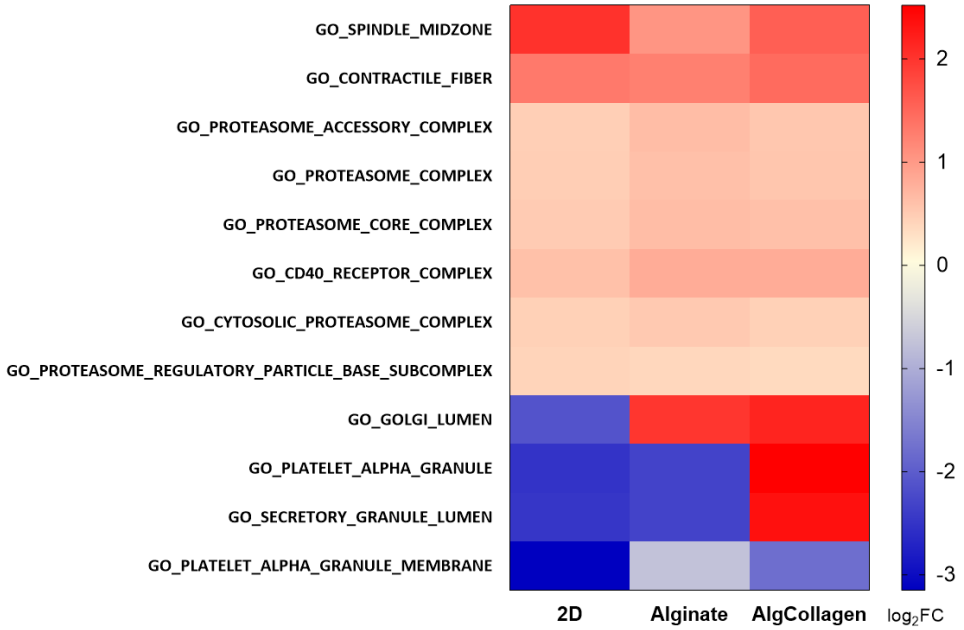


Figure 58: Top cellular components affected by Mtb infection of each of the 3 cell culture models compared to their uninfected states with adjusted P value < 0.05, analysed by egsea (R environment). *The 3D Alginate-Collagen model demonstrates overexpression of components relating to secretory granules, not seen in the other cell culture models.*

All three cell culture models share upregulation of the proteasome and CD40 receptor. The most upregulation in the 2D model is seen in the spindle midzone component, correlating with upregulation of cell cycling observed using the other GSEA (Figure 55) as well as gene ontology (Table 17). Conversely, the highest upregulated cellular component in the microsphere models is the golgi lumen. The 3D Alginate-Collagen model also displays significant upregulation of components relating to secretory granules within the cell.

4.2.12 Comparison of gene expression in infected cell culture models and clinical TB

An overriding goal of my approach is to identify pathways that favour Mtb and those that protect the host, and apply host-directed therapy to shift the balance to minimize lung destruction during Mtb infection without compromising the host's ability to control Mtb. Following on from the bioinformatic analysis of both the clinical study and microsphere study, the next step was to compare how differential gene expression in the Mtb infected cell culture models compares with that of the clinical TB samples. DEGs in clinical TB relative to control samples were identified as illustrated in Figure 16. DEGs in infected cell culture models relative to their uninfected states were identified using a filter optimized to each cell culture model comparison, resulting in more DEGs than presented in Figure 43. All comparisons used the same bioinformatic pipeline and parameters

as used in the clinical analysis, utilizing limma to identify differentially expressed genes (DEGs) with log fold changes of at least 1.5 and adjusted P value maintained less than 0.05. Comparison of the number of DEGs observed in clinical TB and infected 2D model (Figure 59), 3D Alginate model (Figure 60), and 3D Alginate-Collagen model (Figure 61).

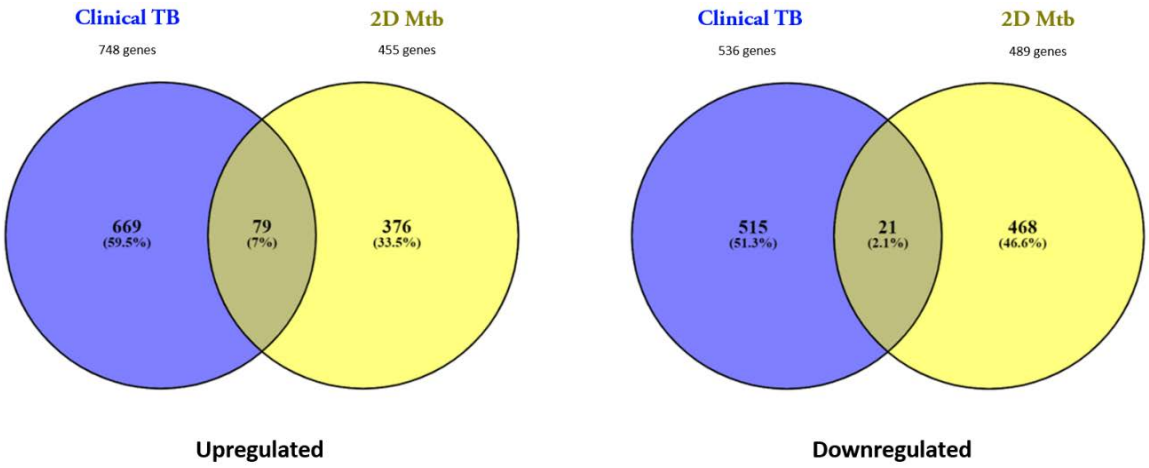


Figure 59: Venn diagram showing the number of DEGs identified on comparison of clinical TB with control samples, and infected 2D model with its uninfected state (absolute $log_2FC \geq 1.5$, adjusted P value < 0.05). *Clinical TB and infected 2D model share 100 DEGs, equal to 8% of those identified in clinical TB samples.*

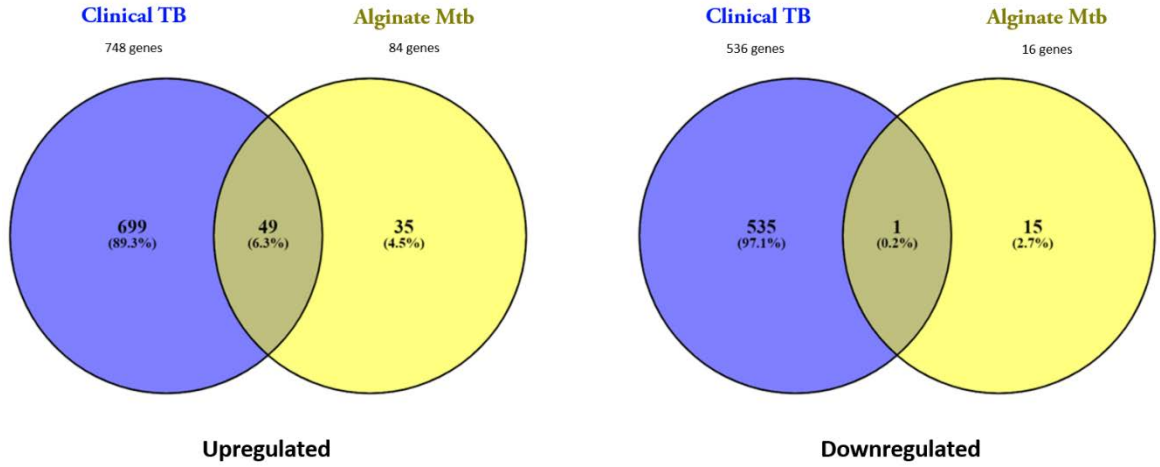


Figure 60: Venn diagram showing the number of DEGs identified on comparison of clinical TB with control samples, and infected 3D Alginate model with its uninfected state (absolute $log_2FC \geq 1.5$, adjusted P value < 0.05). *Clinical TB and infected 3D Alginate model share 50 DEGs, equal to 4% of those identified in clinical TB samples.*

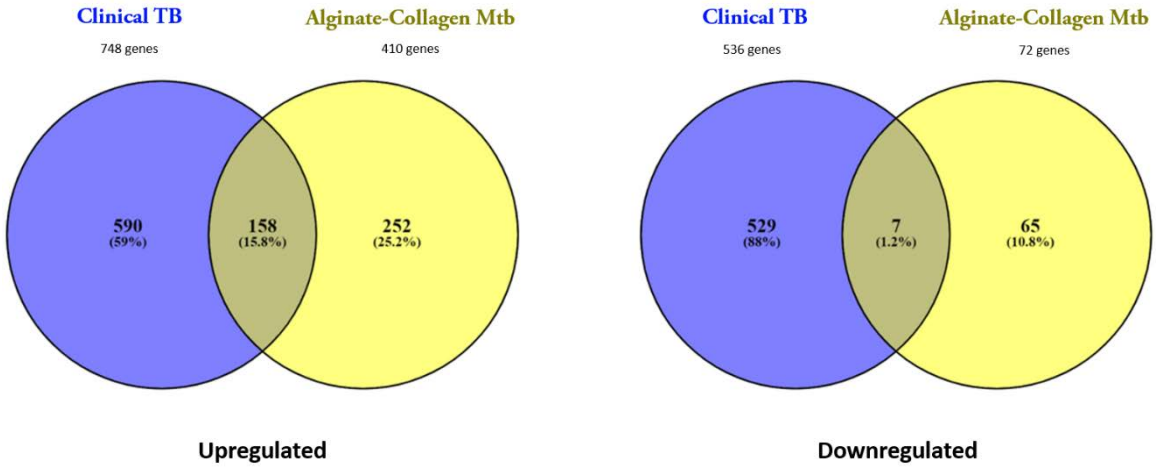


Figure 61: Venn diagram showing the number of DEGs identified on comparison of clinical TB with control samples, and infected 3D Alginate-Collagen model with its uninfected state (absolute $log_2FC \geq 1.5$, adjusted P value < 0.05). *Clinical TB and infected 3D Alginate-Collagen model share 165 DEGs, equal to 13% of those identified in clinical TB samples.*

The 3D Alginate-Collagen model showed the highest proportion (13%) and number of genes (165) overlapping with clinical TB samples, with 158 of the 165 genes upregulated. However, this is still a relatively low degree of overlap. Therefore, to ensure biological relevance, comparison of REACTOME pathways identified on gene ontology analysis was performed between the cell culture models and clinical TB samples was performed. The top 20 most upregulated and downregulated REACTOME pathways in clinical TB relative to control samples, according to adjusted P value, were selected. Utilising the same gene ontology analytical method in comparison of each Mtb infected cell culture model to its uninfected state, the 40 REACTOME pathways identified in clinical TB samples were observed in the cell culture models. Those pathways with adjusted P values below a threshold of 0.05 are illustrated with their direction of fold change (Figure 62).

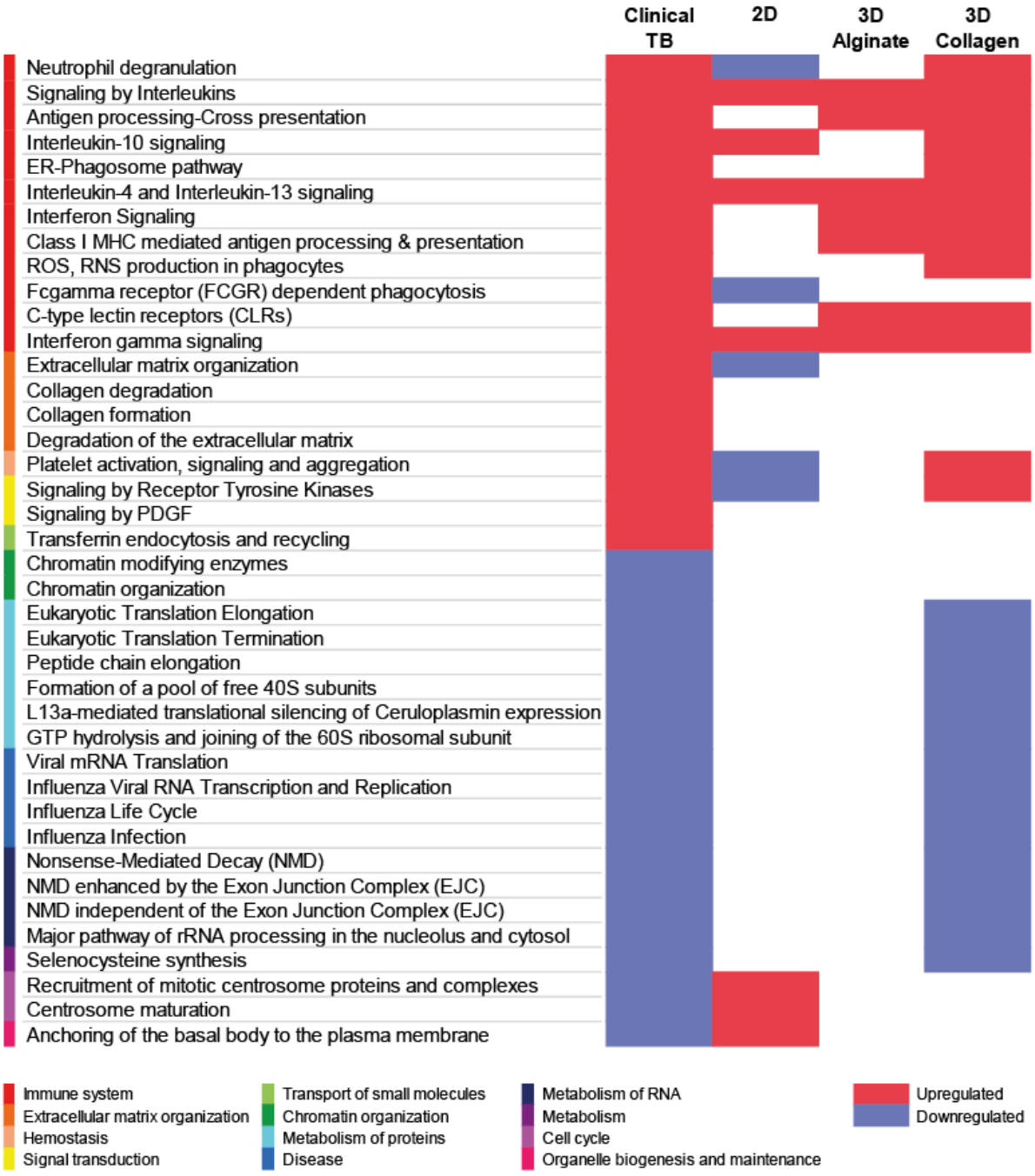


Figure 62: Top 20 REACTOME pathways upregulated and downregulated in clinical TB relative to control samples according to adjusted P value, and comparison to infected models relative to their uninfected state, analysed by ReactomePA using genes with adjusted P value < 0.05 (R environment). *Pathways are ordered within in each gene ontology category according to adjusted P value (clinical TB), and gene ontology category depicted by colour. Red: Pathways significantly upregulated in category. Blue: Pathways significantly downregulated in category. The 3D Alginate-Collagen model most closely reflects clinical TB gene expression.*

The 2D model displays significant dysregulation of the REACTOME pathways, though more often in the opposite direction to the clinical TB samples (4 in the same and 7 in the opposing direction), whereas the microsphere models show concordance in the direction of pathways differentially expressed. Moreover, the 3D Alginate-Collagen model shares a majority of the 40 pathways with

clinical TB samples. The evidence of far greater overlap in gene ontology and DEGs between clinical TB samples and the 3D Alginate-Collagen model, in comparison to the 2D and 3D Alginate model, demonstrates the 3D Alginate-Collagen model is best reflective of human TB. Hence the 3D Alginate-Collagen model of TB was chosen to perform host-directed therapy.

4.3 Discussion

The microsphere study used the PBMCs of six healthy donors in three different primary cell culture methods with and without the stimulus of Mtb infection, leading to six conditions studied: uninfected 2D; infected 2D; uninfected 3D Alginate; infected 3D Alginate; uninfected 3D Alginate-Collagen; and infected 3D Alginate-Collagen. RNA was extracted on day 4 of Mtb infection and underwent sequencing by Illumina HiSeq technology. The same bioinformatic pipeline was used for analysis as for the lymph node study, including differential gene expression, correlation analysis with Markov Clustering, gene ontology and gene set enrichment analysis. Multiple methods were used to ensure the analysis was comprehensive and accurate. On comparison of the cell culture models with clinical TB lymph node samples, infected 3D Alginate-Collagen best reflects the pathways dysregulated in clinical TB samples (Figure 62).

An essential feature of all the cell culture models studied is the presence of primary human cells. Mtb is a pathogen exclusive to humans, with host-pathogen coevolution of at least 70,000 years (Gagneux, 2012). More importantly, a better understanding of the host immune response to Mtb is critical to finding targets for host-directed therapy. The mouse model is the most widely used animal model to study Mtb, but it is limited by the absence of true caseous necrosis or cavities during Mtb infection (Cardona, 2010), which may be explained by the lack of functional ortholog of MMP-1 (matrix metalloproteinase 1) (Balbin *et al.*, 2001), the collagenase found to be significantly upregulated in clinical TB samples. The most similar animal model to humans is the non-human primate model, which does form caseous necrosis during Mtb infection, though unlike in humans, 50% of animals progress after infection (Young, 2009). This microsphere study incorporated human PBMCs, and isolated RNA on day 4 of Mtb infection to try to emulate the granuloma in its early stages of evolution, a process thought to change in dimensionality over time (Pagan and Ramakrishnan, 2018). However, it could be argued that studying the cellular models on day 14 may have provided more insight into the immunopathology of TB granulomas, and so a future study of kinetics of gene expression changes over time would be worthwhile.

The differences in TB susceptibility are incompletely understood, with a quarter of the world's population infected with Mtb, but only 10% proceeding to develop active TB disease (Vynnycky and Fine, 2000; Shea *et al.*, 2014). Multiple risk factors have already been identified, such as low socioeconomic status, malnutrition and vitamin D deficiency (Martineau *et al.*, 2011), and immune

suppression (Langan *et al.*, 2020). Donor-to-donor variation is also evident in this microstudy, where the infected cell culture models display a degree of overlap with uninfected cell culture models on principal component analysis (Figure 38) and hierarchical clustering (Figure 40). Interestingly, Y linkage is evident in a cluster of genes seen in the top 50 most variable genes (Figure 41), as being male is another risk factor, with males almost twice as likely to develop TB as females worldwide (WHO, 2020).

The vast quantity of data obtained from RNA sequencing, and the numerous methods and ways to apply each method, make it ever more important to understand and employ diverse methods to gain the most biologically relevant results (Byron *et al.*, 2016). The analysis performed is dictated by the question asked. Here, the complexity of bioinformatic analysis for the six cell culture conditions is evident from the three main comparisons I made, and driven by three questions. Firstly, to understand the baseline activity and processes innate to each model, particularly the differences between 2D and 3D cell culture, the three cell culture models were compared with each other in the absence of Mtb infection. Secondly, the effect of Mtb infection on each model was analysed by separately comparing each model with and without Mtb infection. Thirdly, the effect of collagen in Mtb infection was analysed by comparison of the three cell culture models only when infected.

The behaviour of uninfected cell culture models was analysed and demonstrated the 2D model to have the highest baseline activity, evidenced by the most differentially expressed genes (DEGs) (Figure 42) and most processes and pathways using gene ontology (Table 16). Adhesion to plastic in the 2D model is a factor likely to contribute to higher levels of cell activation. Another factor is the likely regulatory effect of the 3D environment on cellular processes. Matrix rigidity has been shown to modulate cell biology, such as determination of cell lineage (Engler *et al.*, 2006), and cell motility (Ulrich, de Juan Pardo and Kumar, 2009) and gene expression in the cancer field (Huang *et al.*, 2019). Given that the uninfected models were next used as a comparator to their infected state, this first comparison is important. My findings further support the conclusion that adherence to tissue culture plastic may introduce changes in gene expression that then confound the analysis of subsequent changes induced by a pathogen. However, cell culture models remain crucial to understanding TB in human cells, though the challenges of these models, including the absence of other important immune cells and vasculature, are not easily overcome (Fonseca *et al.*, 2017).

The effect of Mtb on each cell culture model was studied by comparing each infected model with its uninfected counterpart. The highest number of DEGs are found in the 2D model (Figure 43), of which those that are upregulated are likely to relate to the cell cycle identified on gene ontology (Table 17) and gene set enrichment analysis (Figure 55). Given that the 2D model also has the highest level of cellular activity when uninfected, the same factors such as adhesion to plastic may account for the high level of gene expression, in keeping with previous cell studies (Puschmann *et al.*, 2013), with infection providing an additional stimulus to pathway upregulation. Conversely, gene ontology finds the most diverse biological processes and pathways in both 3D models and include upregulation of phagocytosis, apoptosis and IFNγ signalling (Table 17). Gene set enrichment analysis shows the most overexpressed pathways in the microsphere models relate to extracellular matrix organisation and are exhibited to a higher degree in the 3D Alginate-Collagen model (Figure 57), consistent with published works (Ravi *et al.*, 2015), while secretory granules are the only cellular component uniquely overexpressed in the 3D Alginate-Collagen model (Figure 58). Therefore, these findings are consistent with the emerging concept of the importance of 3D organisation and extracellular matrix regulation of human immune responses (Sorokin, 2010; Schwartz and Chen, 2013).

An alternative method of analysis, correlation analysis using Markov Clustering (Pearson's correlation coefficient 0.88), finds Cluster 14 to be the only cluster underexpressed in uninfected microsphere models but upregulated during Mtb infection, particularly in the 3D Alginate-Collagen model (Figure 46). Cluster 14 is comprised of 7 genes which are individually involved in a number of biological processes. The chemoattractant CCL22 *(CCL22* gene, C-C motif chemokine ligand 22) acts on monocytes, natural killer cells and activated T lymphocytes (Godiska *et al.*, 1997), while CCL24 (*CCL24* gene, C-C motif chemokine ligand 24) is a chemoattractant for resting T lymphocytes (Patel *et al.*, 1997). CD1B (*CD1B* gene, CD1b Molecule) mediates the presentation of self or microbial lipid and glycolipid antigens to T lymphocytes (Briken *et al.*, 2000). CD1B shares a genetic locus on chromosome 1q23.1 with nestin (*NES* gene, Nestin), an intermediate filament involved in cell division, particularly axon growth in nerve cells (Dahlstrand *et al.*, 1992). IGFBP6 (*IGFBP6* gene, Insulin like growth factor binding protein 6) has diverse effects by prolonging the half life of insulin like growth factors, a family of proteins implicated in diabetes mellitus and cancer (Bach, 2005). Inhibition of fibrinolysis by plasminogen activator inhibitor 1 (PAI-1), the principal inhibitor of tissue plasminogen activator and urokinase, is encoded by *SERPINE1* (Serpin Family E Member 1) (Fay *et al.*, 1997). Lastly, TIE1 (*TIE1* gene, Tyrosine Kinase With Immunoglobulin Like And EGF Like Domains 1) plays a critical role in angiogenesis by inhibiting angiopoietin 1 signalling (Seegar *et al.*, 2010).

Altogether, gene ontology analysis of Cluster 14 finds chemotaxis, angiogenesis and extracellular matrix organisation among the top biological processes (Table 14) or pathways (Table 15). Interestingly, VEGF (vascular endothelial growth factor) signalling is activated in the zebrafish model of TB (Uusi-Makela and Ramet, 2018) and in humans (Polena *et al.*, 2016), suggesting angiogenesis to be a potential host therapeutic target to treat TB. Another potential target for host-directed therapy is the extracellular matrix, whose crucial role in cell signalling and granuloma formation are discussed in 1.2 The role of the extracellular matrix.

The effect of collagen before and during Mtb infection was studied in the cell culture models. Prior to infection, addition of collagen to the microsphere model demonstrates fewer DEGs than the 3D Alginate model, behaving more similarly to 2D cell culture (Figure 42). Gene ontology of uninfected models also shows upregulation of a developmental pathway in the 3D Alginate-Collagen model relative to 2D cell culture (Table 16). However, what is most striking is the effect of collagen on gene expression in the 3D models during Mtb infection. Comparison of the three infected cell culture models demonstrates the highest number of DEGs (Figure 44) in the 3D Alginate-Collagen model, together with the highest number of upregulated processes and pathways on gene ontology analysis (Table 18). These pathways relate to angiogenesis, cell migration, phagocytosis and the immune response, all important in TB (O'Garra and Berry, 2013). Collagen is the primary lung matrix fibril in humans, and TB granulomas are dynamic environments which become fibrotic over time. The role of collagen is not merely structural but plays a significant part in altering cell behaviour (Sorokin, 2010). Cell adhesion molecules have been shown to affect signalling pathways in cancer progression (Christofori, 2003). Furthermore, mammalian collagen receptor genes are known to have multiple biological functions including fibroblast proliferation, matrix metalloproteinase expression, arterial wound repair and immune cell regulation (Leitinger and Hohenester, 2007). Therefore, modulation of the host response by collagen is not surprising, but this modulation may then impact the analysis of specific interventions.

Modelling the 3-dimensional environment is crucial to replicating *in vivo* granulomatous conditions (Elkington *et al.*, 2019). Granulomas comprise a core of multinucleate giant cell surrounded by $CD4^+$ T cells and fibroblasts (Pagan and Ramakrishnan, 2018). Advanced cell culture models allow incorporation of multiple cell types and temporal studies of conditions including the granuloma (Elkington *et al.*, 2019), a necessary feature considering granulomas evolve with the clinical course of TB. Studying host-pathogen interactions in a 3D model adds another level of complexity when using

live Mtb culture, as all work must be performed in containment level 3 facilities. Therefore all material is required to be autoclaved, formaldehyde treated, lysed or sterilely filtered for microorganism capture before leaving the laboratory, hence investigations are limited to the local facilities. The emerging importance of the extracellular matrix and how it influences cell-cell adhesion, cell signalling, cell migration and diffusion gradients demonstrates the advantage 3D cell culture models can offer (Schwartz and Chen, 2013).

A number of other 3-dimensional cell culture models exist. A multicellular lung tissue model developed in Sweden utilises a scaffold consisting of a fibroblast collagen layer with stratified epithelial cells (Parasa *et al.*, 2014). The lung tissue model displays active formation of extracellular matrix proteins and is able to secrete mucus, though limitations of this model include a lack of T cells due to the use of cell lines to form the matrix, and the limitation of MHC-incompatability. Another model is a static collagen and fibronectin extracellular matrix model, used to model Mtb dormancy (Kapoor *et al.*, 2013). This collagen model shows features of Mtb reactivation on immunosuppression. However, technical limitations include difficulty in releasing cells without using collagenase, and a relatively low throughput. Lastly, a human granuloma model composed of PBMCs has been developed (Guirado *et al.*, 2015). The main limitation of this PBMC-based model is the lack of extracellular matrix. Overall, each model has its strengths and weaknesses, and validation against the clinical gene expression patterns is essential to confirm the *in vivo* validity.

Comprehensive bioinformatic analysis shows the 3D Alginate-Collagen model is the cell culture model most reflective of clinical TB samples. Of the three humanised cell culture models, the 3D Alginate-Collagen model demonstrates the highest number of DEGs (Figure **61**), and most overlap of REACTOME pathways using gene ontology (Figure 62) and significantly enriched canonical pathways on gene ontology (Figure 32 and Figure 57). The 3D microenvironment and extracellular matrix are likely to play a major role in TB pathogenesis (Ong and Friedland, 2014; Marakalala *et al.*, 2016) and are well represented in the 3D Alginate-Collagen model. However, limitations of this model are also seen on gene ontology analysis of the 3D Alginate-Collagen model, which demonstrated no significant upregulation of extracellular matrix organisation pathways (Figure 62), a key feature of clinical TB samples used in this study. Hence, limitations of the 3D Alginate-Collagen model may include the lack of fibroblasts which need to be considered for a future approach (O'Kane *et al.*, 2010). Despite this limitation, the 3D Alginate-Collagen model clearly has the most similar

transcriptomic signature to that of human TB and was therefore used for my translational experiments.

Chapter 5: Integration of clinical and microsphere studies

5.1 Introduction

Having chosen the cell culture model with the transcriptome most closely resembling clinical TB samples, the next step was to find appropriate pathways which could be targeted and thus the effect on Mtb growth and the host response observed. Potential targets communal to the clinical TB samples and 3D Alginate-Collagen model needed to be identified using a systematic and non-hypothesis driven process to maximise the chance of finding a novel therapeutic target for TB therapy. My initial goal was to find the common pathways between the clinical TB samples and bioelectrospray model in order to perform a targeted intervention to skew the host-pathogen interaction in favour of bacterial control.

5.2 Results

5.2.1 Target selection using the 3D Alginate-Collagen cell culture model

Upregulated differentially expressed genes (DEGs) shared by infected 3D Alginate-Collagen cell culture and clinical TB samples were analysed, because inhibition of cellular proteins may be easier to target than activation of them. In order to further understand the differences and communalities between TB and sarcoidosis, DEGs identified during comparison of clinical TB and control samples (Figure 16) were divided according to whether they were unique to TB or communal to TB and sarcoidosis. An almost identical number of upregulated DEGs were identified in two groups: DEGs unique to TB and upregulated in the 3D Alginate-Collagen model; and DEGs communal to both TB and sarcoidosis and upregulated in the 3D Alginate-Collagen model (Figure 63).

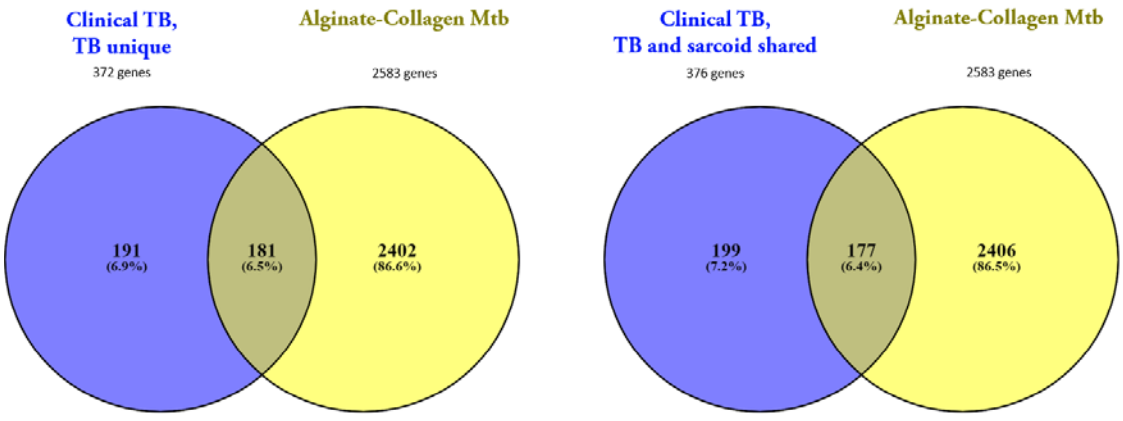


Figure 63: Venn diagram showing the number of upregulated DEGs in clinical TB relative to control samples, either specific to clinical TB or communal to clinical TB and sarcoidosis (absolute $log_2FC \geq 1.5$, adjusted P value < 0.05), overlapping upregulated DEGs in the infected 3D Alginate-Collagen model with its uninfected state (absolute $log_2FC \geq 0.18$, adjusted P value < 0.05). *Overlap of 181 genes specific to TB, and 177 genes communal to TB and sarcoidosis is seen in the 3D Alginate-Collagen model.*

The next step in target selection was to consider the type of protein that may be inhibited most easily in the microsphere culture system. Intracellular enzymes may be more abundant and easier to target than cell surface receptors, and are often targeted clinically, such as COX by aspirin and tyrosine kinases by imatinib. Therefore, I first narrowed down to these, and out of the 358 DEGs shared by the clinical TB and infected 3D Alginate-Collagen samples, 90 genes were found to be intracellular enzymes (Figure 64). Each of the 90 genes was explored to find whether their corresponding chemical inhibitors existed and were available for purchase. Consequently 36 available chemical inhibitors were identified and researched. Target specificity was crucial in the target selection process, and so drugs with multiple targets were excluded. Drug toxicity studies were reviewed and drugs ranked accordingly from drugs in human use, drugs in clinical trials, to drugs not toxic in mammalian models, drug not toxic in other animal models, and finally those only used in cell culture models. The final factors that were considered in target selection were the DEG fold change values in the clinical TB compared to control samples, DEG fold changes in infected 3D Alginate-Collagen compared to uninfected 3D Alginate-Collagen, and transcript abundance.

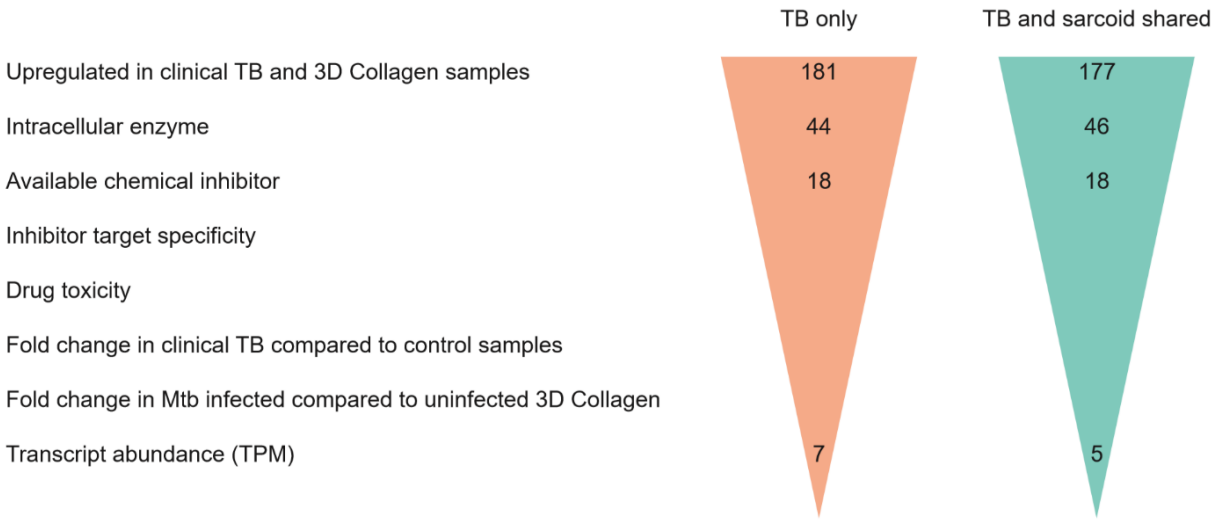


Figure 64: Selection process used to identify targets for host-directed therapy experiments. *Of the 358 genes initially identified, 12 potential candidates were selected for study in the 3D Alginate-Collagen cell culture model.*

As a result, the target selection process generated twelve intracellular target enzymes, seven of which are unique to TB and five of which are communal to TB and sarcoidosis. The ranking of these twelve targets is detailed in Table 19.

Gene	Name	Animal model safety	Clinical TB log_2FC	Infected 3D Alginate-Collagen log_2FC	Clinical Control mean TPM
CA2	carbonic anhydrase 2	In human use	2.55	3.36	150
TYMP	thymidine phosphorylase	Phase III trials	2.73	2.05	2713
SRC	SRC proto-oncogene, non-receptor tyrosine kinase	Phase III trials	2.29	1.88	181
HSD11B1	hydroxysteroid 11-β dehydrogenase 1	Phase II trials	7.74	2.40	8
NAMPT	nicotinamide phosphoribosyltransferase	Phase I trials	1.80	1.75	1497
LRRK2	leucine rich repeat kinase 2	Safe in monkeys	2.06	1.42	520
HK2	hexokinase 2	Safe in rats	3.94	1.35	114
SPHK1	sphingosine kinase 1	Safe in mice	2.50	2.39	393
RIPK2	receptor interacting serine/threonine kinase 2	Safe in mice	2.28	1.25	463
FADS2	fatty acid desaturase 2	Safe in mice	2.04	0.47	420
CTSL	cathepsin L	Safe in zebrafish	1.95	1.42	1437
FURIN	furin	NA	2.08	1.87	400

Table 19: Twelve selected potential host-directed targets ranked from 1 to 12. *Ranking was according to the order of criteria used in the systematic selection process.*

5.2.2 Inhibition of twelve enzymes to identify potential host-directed therapy targets in TB

The next step after target selection was to determine the dose of inhibitor to add to Mtb-infected 3D Alginate-Collagen culture. Published works using the inhibitors were reviewed, noting both drug concentration and model in which they were applied (Table 20). Since microspheres of the 3D model require drugs to diffuse a greater distance to reach cells, in comparison to 2D cell culture, higher concentrations of inhibitor may be required to see any effect.

Target	Inhibitor	Model	Concentration used previously	Initial concentration
CA2	Brinzolamide	In human use	26-78µM	4µM
CTSL	SID 26681509	Human PBMCs and endothelial cells	0.05 - 100µM	0.05µM
FADS2	SC 26196	Primary human cells, primary cell spheroids, and cancer cell lines	1 - 100µM	1µM
FURIN	SSM3 trifluoroacetate	Glioma cell line and rat macrophages	1 - 250µM	1µM
HK2	3-Bromopyruvic acid	Human RBC culture, human/rat glioma cell spheroids, and cancer cell lines	1µM – 3mM	1µM
HSD11B1	PF 915275	Primary human hepatocytes, human preadipocyte/keratinocyte culture, and sebocyte cell line	0.01µM – 10µM	0.05µM
LRRK2	GNE-7915	Cynomolgus monkeys serum levels, neuroblastoma cell line, primary rat/mouse brain neuron culture, and mouse brain slices	0.1µM - 10µM	1µM
NAMPT	CHS 828	Human lung cancer cell culture, human cancer cell lines	5 - 50nM	5nM
RIPK2	GSK 583	Human PBMCs, human intestinal biopsy, human macrophage/monocyte cell line, NOD2 cell line	0.01µM - 10µM	0.01µM
SPHK1	PF-543	Mouse model, human whole blood, primary human cancer cells, human cancer cell lines, and primary rat microglial cells	1 - 50µM	1µM
SRC	KX2-391	Primary human cancer cells, human liver cell line, human cancer cell lines	0.012 – 6564nM	0.05µM
TYMP	Tipiracil hydrochloride	Human cancer cell lines	0.25 - 10µM	0.25µM

Table 20: Chemical inhibitors used to study 12 potential host-directed therapy targets. *Determination of initial inhibitor concentrations to use in the 3D Alginate-Collagen model were based on concentrations stated in published works.*

Initial experiments added each inhibitor on day 1 (24 hours after Mtb infection of PBMCs) at 2 different concentrations with a 5 fold difference. To minimise the risk of inhibitors binding to unintended target molecules, the lowest concentration of inhibitor selected was based on the lowest effective concentration of inhibitor published. All drugs were reconstituted in DMSO, except for 3-BPA (hexokinase 2 inhibitor) and tipiracil (thymidine phosphorylase) which were dissolved in

water. The DMSO condition matched the highest DMSO concentration used to reconstitute the inhibitors in each experiment.

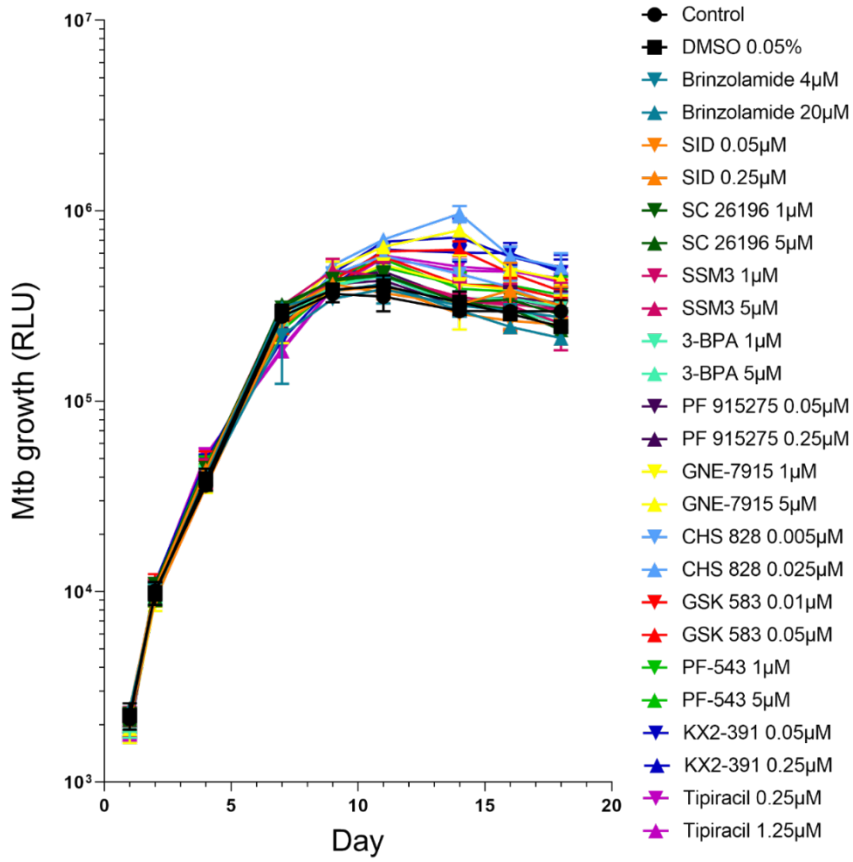


Figure 65: Initial host-directed therapy experiments in the 3D Alginate-Collagen model. *Poor Mtb growth is seen from day 9 onwards during which the twelve selected inhibitors had no effect. Error bars show standard deviation.*

Unfortunately, Mtb growth stopped after day 9, possibly due to the rapid growth seen in the first few days with subsequent energy source depletion (Figure 65). However, it was evident that the inhibitors had little effect on Mtb growth at these initial low concentrations. Thus the experiments were repeated using inhibitor concentrations 10 times higher than used in initial experiments, except for the CA2 (carbonic anhydrase 2) inhibitor, brinzolamide, which was only increased to 5 times the initial concentration in order to remain in the concentration range in current clinical use (Figure 66).

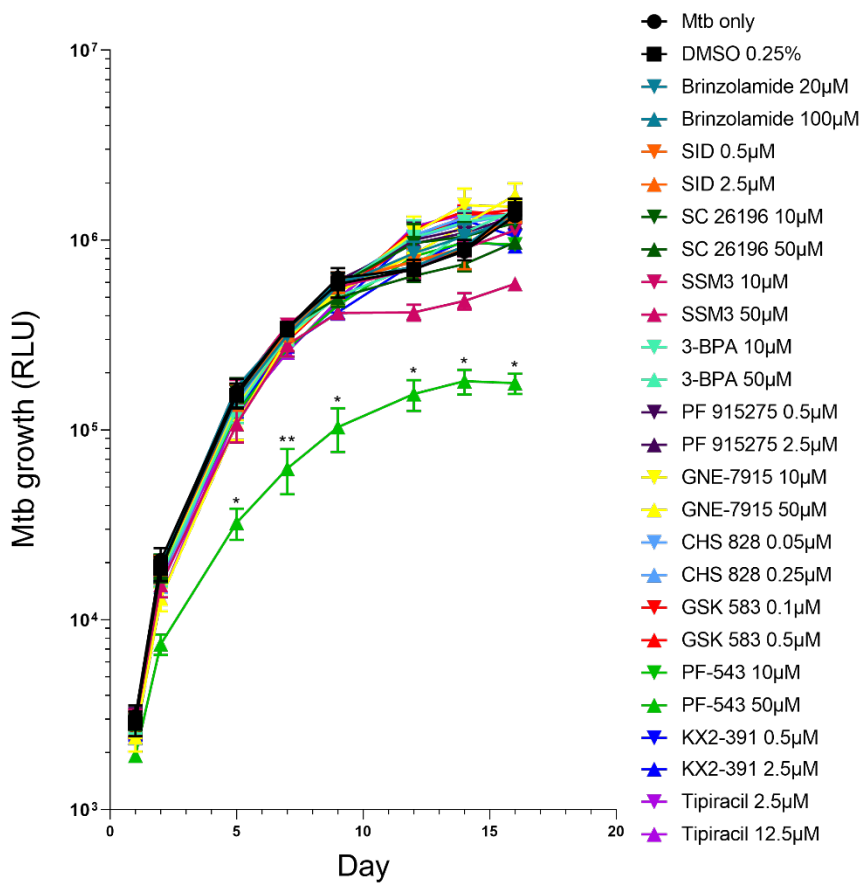


Figure 66: Effect of selected target inhibition on Mtb growth using higher inhibitor concentrations added on day 1. *SSM3 (pink) and PF-543 (bright green) control Mtb growth. Analysis by two way ANOVA, error bars show standard deviation. *p<0.05, **p<0.01*

Of the twelve enzyme candidates, two resulted in reduction in Mtb growth: the furin inhibitor SSM3 and sphingosine kinase 1 (SphK1) inhibitor PF-543 (Figure 66). Notably, inhibition of SphK1 significantly reduced Mtb growth. Therefore subsequent experiments used these higher inhibitor concentrations. The inhibitors were also added at both day 1 and day 7 to better reflect *in vivo* kinetics of drug application.

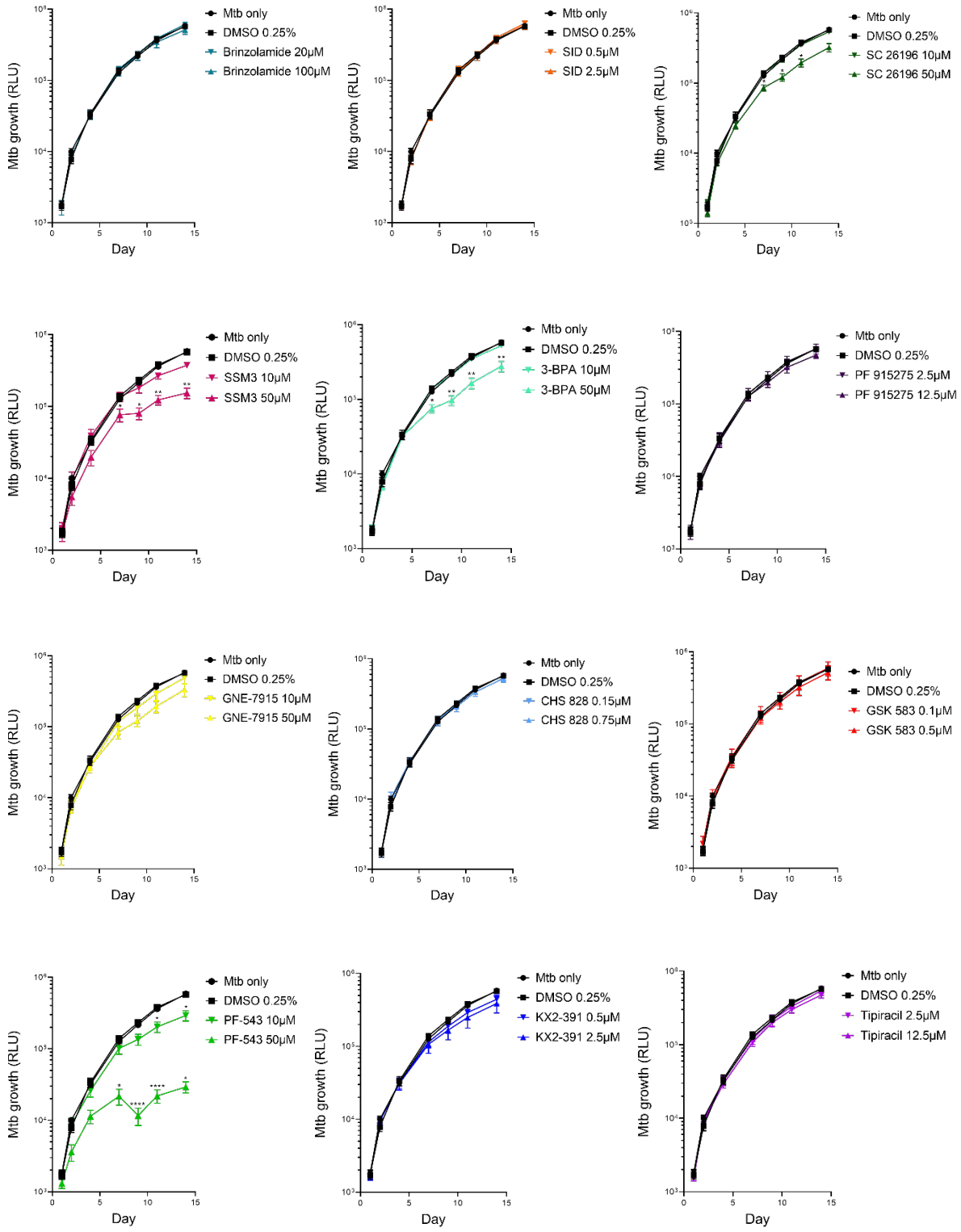


Figure 67: Effect of selected target inhibition on Mtb growth using higher inhibitor concentrations added on day 1 and 7. *Four inhibitors demonstrate significant control of Mtb: SC 26196 (dark green); SSM3 (pink); 3-BPA (turquoise); and PF-543 (bright green). PF-543 has the most significant effect on Mtb growth. Analysis by two way ANOVA, error bars show standard deviation. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.*

A significant reduction in Mtb growth resulted from 4 inhibitors: SC 26916 (fatty acid desaturase 2 inhibitor), SSM3 (furin inhibitor), 3-BPA (hexokinase 2 inhibitor), and PF-543 (sphingosine kinase 1

inhibitor) (Figure 67). Notably, control of Mtb growth occurred after adding SSM3 and PF-543 on day 7, and both inhibitors had the greatest effect on Mtb growth with a dose response observed. This effect was seen in 2 separate donors at this stage of experiments.

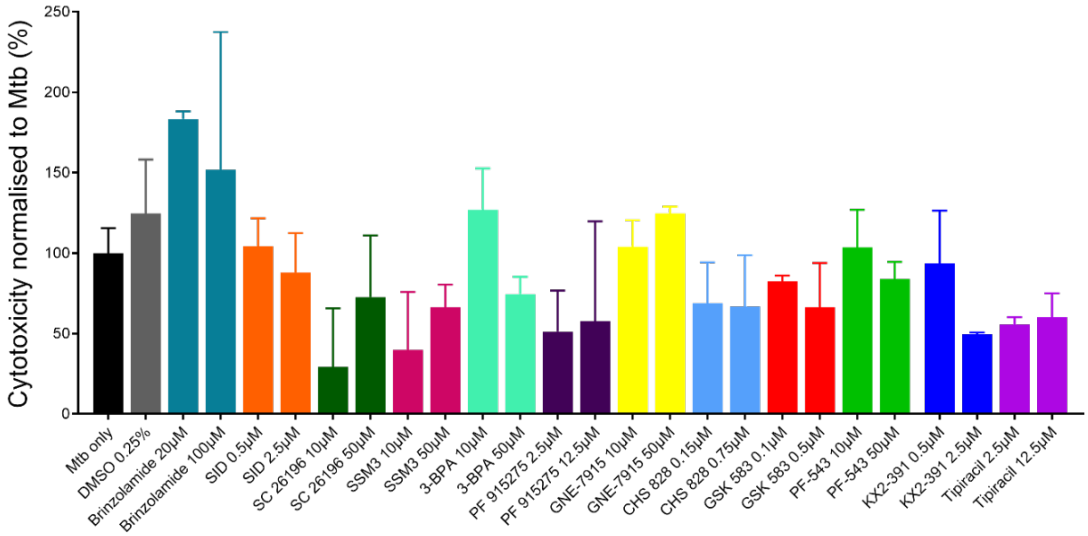


Figure 68: Effect of inhibitors on cellular toxicity on day 7 using CytoTox-Glo assay, normalised to Mtb alone. *Comparison of inhibitors with their relevant control (Mtb only or DMSO 0.25%) finds no significant cell toxicity. Mean values of triplicate are shown. Analysis by Student's t test, error bars show standard deviation.*

Concurrent cellular toxicity studies in each Mtb growth experiment were performed on day 7 to ensure excessive cell death was not contributing to reduction in Mtb growth (Figure 68). Two cytotoxicity assays were utilised, CytoTox-Glo and LDH (lactate dehydrogenase) assay. Brinzolamide (carbonic anhydrase 2 inhibitor) was the only inhibitor to have a greater cytotoxic effect than DMSO alone using CytoTox-Glo assay, with the same effect seen when performing the LDH (lactate dehydrogenase) assay on day 7 (Figure 69), though wide error bars are seen. Importantly, inhibition of the two most promising targets, SphK1 and furin, did not cause high levels of cell toxicity.

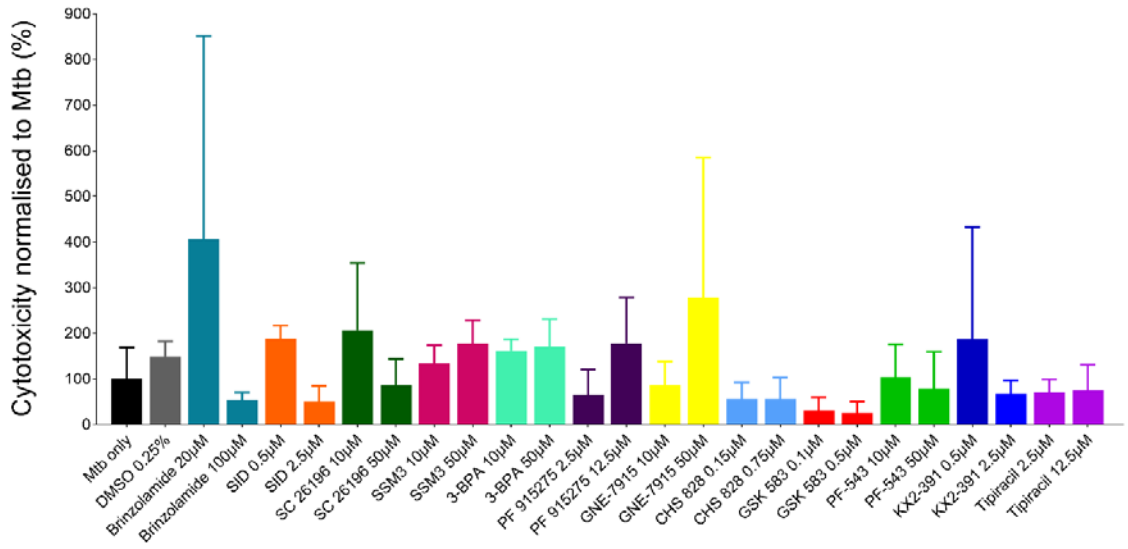


Figure 69: Effect of inhibitors on cellular toxicity on day 7 using LDH assay, normalised to Mtb alone. *Comparison of inhibitors with their relevant control (Mtb only or DMSO 0.25%) finds no significant cell toxicity. Mean values of triplicate are shown. Analysis by Student's t test, error bars show standard deviation.*

Another important consideration was whether the inhibitors were acting on Mtb proteins. BLAST (Basic Local Alignment Search Tool) with compositional matrix adjustment method was utilised to compare the nucleotide sequence of Mtb strain H37Rv with the nucleotide sequences of the two most promising candidates for host-directed therapy, SphK1 and furin (Table 21). *SPHK1* exhibits homology with Mtb protein diacylglycerol kinase, with higher probability sequence matches of up to 56%, while FURIN shows homology with Mtb protease of up to 47%.

	Mtb protein	**BLAST score**	**Expect value**	**Identities**
SPHK1 isoform 1	Diacylglycerol kinase catalytic region	47.4 bits	5E-04	34%
SPHK1 isoform 1	Diacylglycerol kinase	34.7 bits	5.7	56%
SPHK1 isoform 2	Diacylglycerol kinase catalytic region	46.6 bits	1E-03	49%
SPHK1 isoform 3	Diacylglycerol kinase catalytic region	47.4 bits	5E-04	49%
SPHK1 isoform 3	Diacylglycerol kinase	34.7 bits	6.6	56%
FURIN	Subtilisin DY	37.4 bits	2.6	30%
FURIN	Protease	37.0 bits	3.9	36%
FURIN	Protease, partial	34.3 bits	8.7	47%

Table 21: BLAST results on comparison of SPHK1 and FURIN with Mtb strain H37Rv nucleotide sequence. *SPHK1 and FURIN show a high degree of homology with Mtb.*

To confirm drug application was having an effect on the host cell, instead of directly affecting the pathogen itself, SphK1 and furin were added to Mtb grown in Middlebrook 7H9 broth alone (Figure 70). In the absence of PBMCs, inhibition of sphingosine kinase 1 by PF-543 had no effect on Mtb growth, whereas furin inhibition by SSM3 significantly reduced Mtb growth and therefore suggests SSM3 is acting directly on Mtb. This effect was seen in 3 separate donors.

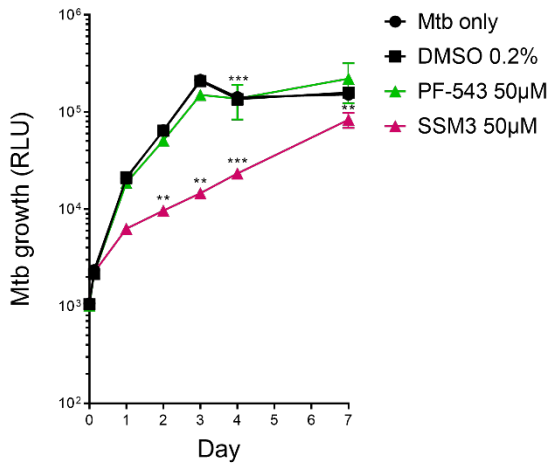


Figure 70: Effect of PF-543 (sphingosine kinase 1 inhibitor) and SSM3 (furin inhibitor) on Mtb growth in broth alone. *PBMCs are not present. PF-543 (green) has no effect on Mtb growth, in contrast to SSM3 (pink) which slows Mtb growth. Analysis by two way ANOVA, error bars show standard deviation. **p<0.01, ***p<0.001*

The dose-dependent effect of PF-543 on control of Mtb growth via host-directed mechanisms, and lack of excessive cell toxicity, observed in multiple donors led to sphingosine kinase 1 being taken forward and investigated as a potential host-directed therapeutic target.

5.2.3 Validation of sphingosine kinase 1 as a key regulator in TB

A closer look at sphingosine kinase 1 (SphK1) within its signalling pathway in clinical TB, relative to control samples, demonstrates both SphK1 (*SPHK1*) and S1P lyase (*SGPL1*) are significantly upregulated with $\log_2$ fold changes of 2.5 and 1.05 respectively (Figure 71). Sphingosine-1-phosphate (S1P) is known to act as an intracellular messenger on multiple receptors. As a result of SphK1 and S1P lyase upregulation, there may be depletion of ceramide and changes in S1P signalling effects due to the overall pathway diversion from ceramide to PEA in TB granulomas.

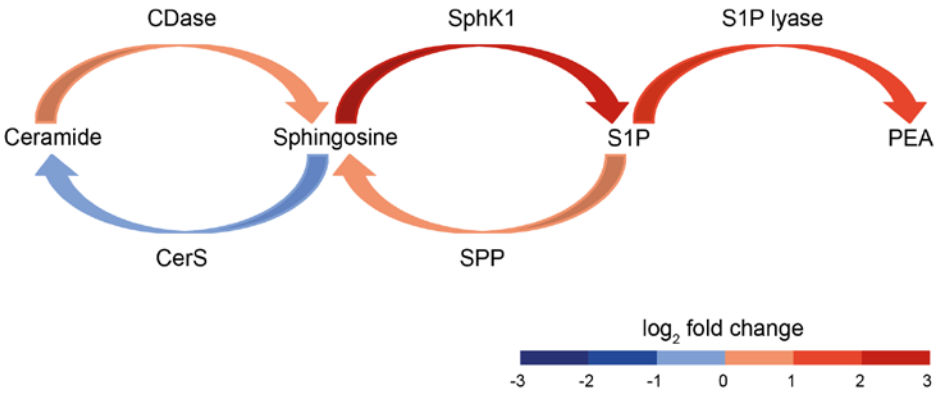


Figure 71: Graphical representation of changes in sphingolipid signalling pathway genes in TB granulomas relative to control lymph nodes (adjusted P value < 0.05 for SphK1 and S1P lyase, otherwise more than 0.05). *Upregulation of SphK1 and S1P lyase together with downregulation of CerS may lead to ceramide depletion. CDase, ceramidase; CerS, ceramide synthase; S1P, sphingosine-1-phosphate; SPP, sphingosine-1-phosphate phosphatase; PEA, phosphoethanolamine.*

To interrogate this pathway further, three compounds were chosen based on their specificity within the SphK1 signalling pathway. The first compound was another SphK1 inhibitor, SKI 178, with a different chemical structure to PF-543. The second compound was a downstream inhibitor acting on four of the five known sphingosine-1-phosphate (S1P) receptors, FTY720, otherwise known as fingolimod, a drug currently used to treat the relapsing and remitting form of multiple sclerosis. Thirdly, a specific SphK1 activator, K6PC-5, was selected. The mechanism of action of these three compounds, together with PF-543, is illustrated on the KEGG Sphingolipid Signalling pathway (Figure 72).

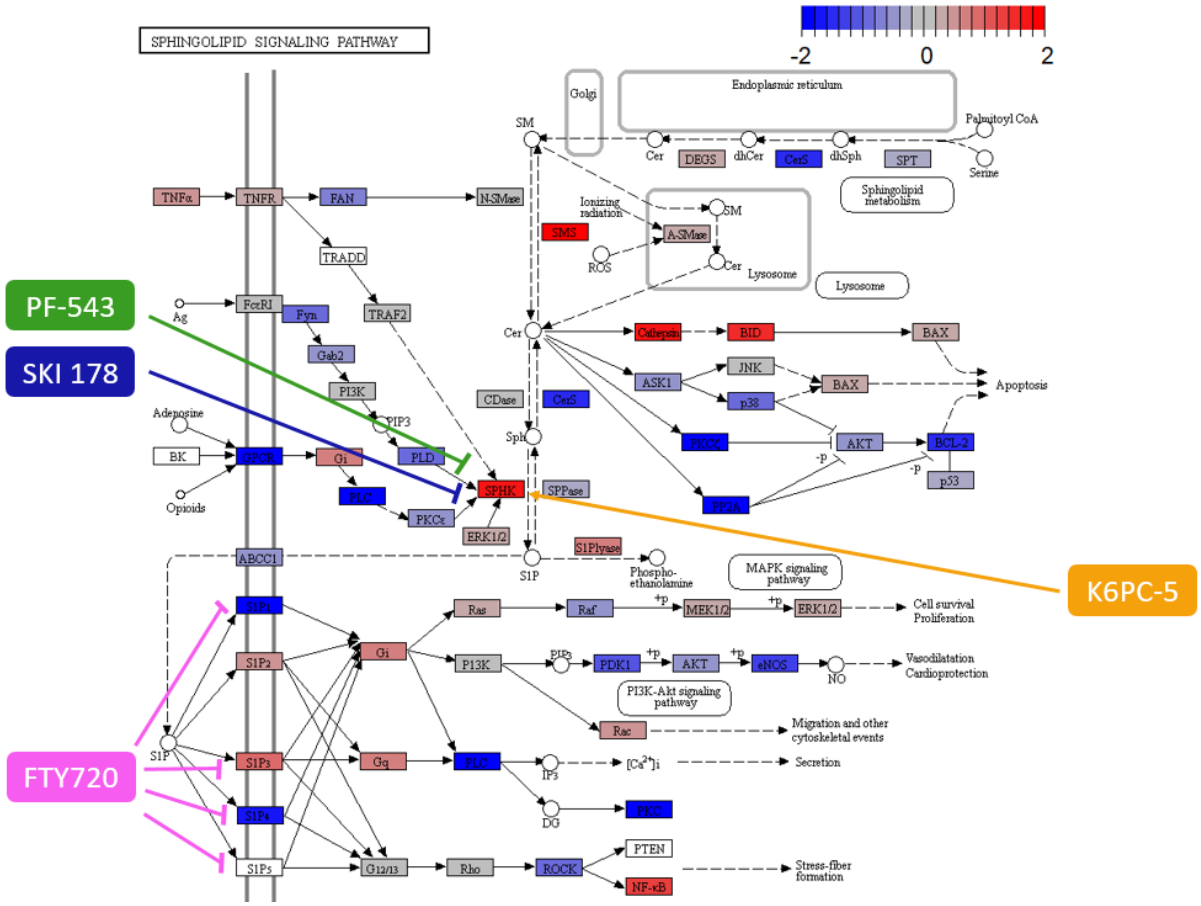


Figure 72: Expression of genes regulated in clinical TB in comparison to control specimens, overlaid on the KEGG Sphingolipid Signalling pathway, analysed by egsea (R environment). *The mechanism of action of four compounds is displayed: SphK1 inhibitor PF-543 (green); another SphK1 inhibitor SKI 178 (blue), S1P receptors 1, 3, 4 and 5 inhibitor FTY720 (pink); and SphK1 activator K6PC-5 (orange). Fold changes have been converted to a Z-score, with upregulated genes in red and downregulated genes in blue.*

It is evident that SphK1 is highly upregulated within the sphingolipid signalling pathway in TB samples, and may play a major role in TB pathogenesis. To investigate this in the 3D Alginate-Collagen model, PF-543 was added again at 50µM, the higher concentration used in earlier experiments which had maximal effects on controlling Mtb growth. All compounds were added at day 1 and 7. The other compounds were used at two concentrations with 10 fold difference, with

the lower concentration based on published studies of each compound (Table 22). In addition to the four compounds mentioned, sphingosine-1-phosphate (S1P) itself was applied, both alone and in combination to see if any effects from inhibitors could be reversed. Due to S1P having a half life of between 15 and 120 minutes, three concentrations were used and added on days 1, 2, 4, 7, 9, and 11, after Mtb growth was measured.

Target	Compound	Model	Concentration used previously	Concentration used initially
SPHK1	SKI 178 (inhibitor)	Cancer cell lines	0.5 - 5µM	1µM
S1P receptors 1/3/4/5	FTY720 (inhibitor)	In human use, mouse model, primary human cells, human T cell lines	1 - 5µM	1µM
SPHK1	K6PC-5 (activator)	Mouse model, primary human cells, human cancer and neuronal cell lines	0.1 - 100µM	10µM
-	S1P	Mouse model, primary human cells, human endothelial and kidney cell lines	0.001 - 10µM	0.1µM

Table 22: Chemical compounds used to study the sphingolipid signalling pathway. *Determination of initial compound concentrations to use in the 3D Alginate-Collagen model were based on concentrations stated in published works.*

As expected, SphK1 inhibition by PF-543 resulted in control of Mtb growth (Figure 73). However, inhibition of SphK1 by another compound, SKI 178, had no effect on Mtb growth, which may be explained by the fact that SKI 178 preparation was more difficult due to being light sensitive and requiring a water bath to dissolve the precipitation seen when reconstituted with DMSO. S1P receptor inhibition by FTY720, at the higher concentration of 10µM, resulted in a non-significant reduction in Mtb growth, particularly after the second addition of drug on day 7, and further controlled Mtb growth in combination with PF-543. SphK1 activation by K6PC-5 demonstrated dose dependent divergent effects on Mtb growth, whereby the higher concentration 100µM significantly reduced Mtb growth and the lower concentration 10µM resulted in significant increase in Mtb growth. Lastly, the addition of S1P had little effect on Mtb growth, both alone and in combination with PF-543.

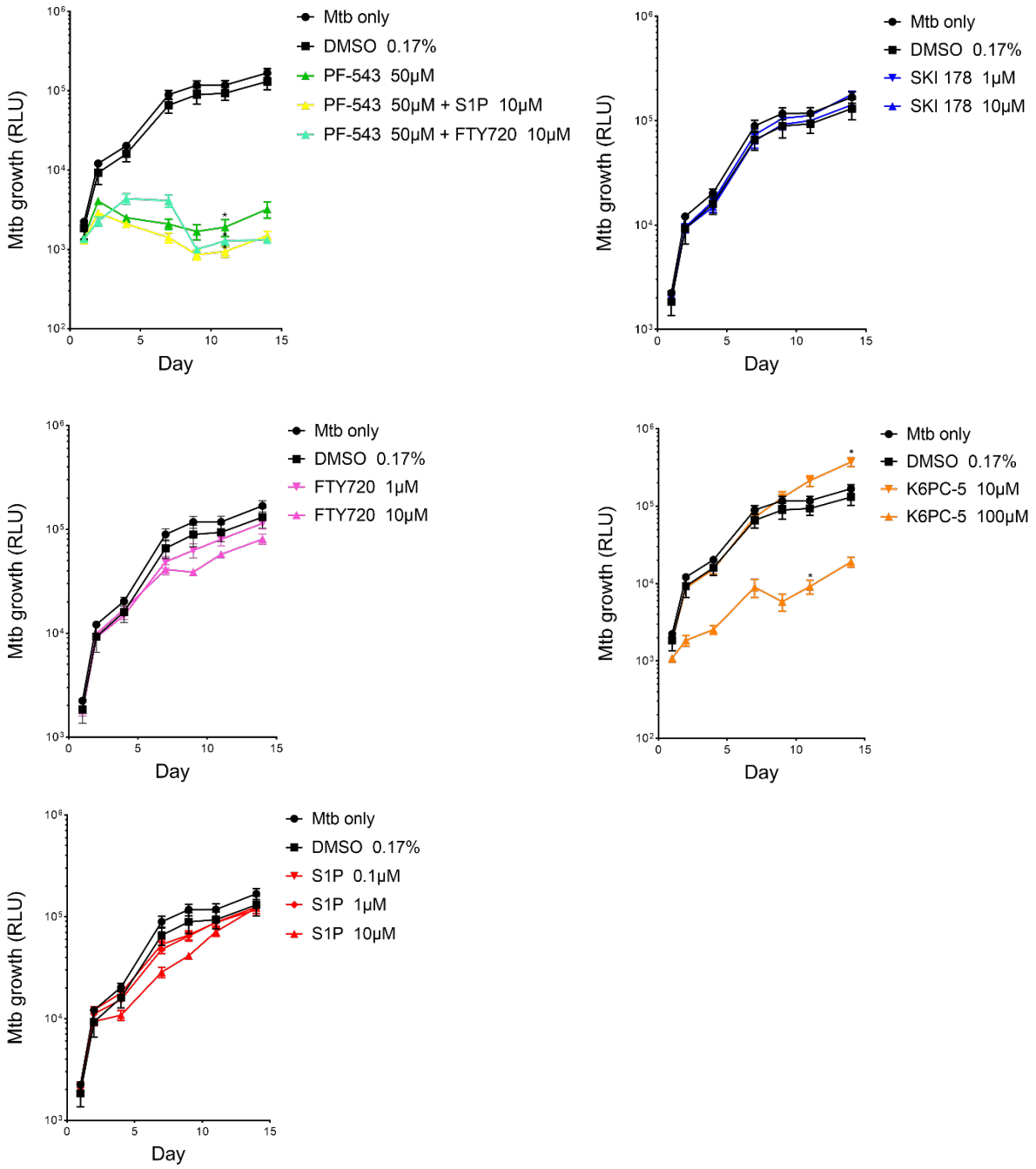


Figure 73: Effect of compounds acting on the sphingolipid signalling pathway on Mtb growth. *SphK1 appears to play a key role in Mtb control, where inhibition by PF-543 (bright green) and activation by K6PC-5 (orange) cause divergent effects. Analysis by two way ANOVA, error bars show standard deviation. *p<0.05*

To ensure cellular toxicity was not responsible for the effects on Mtb growth observed, cellular toxicity studies were concurrently performed on day 7 (Figure 74). No significant excess cell toxicity was observed in comparison to DMSO, though wide error bars are present.

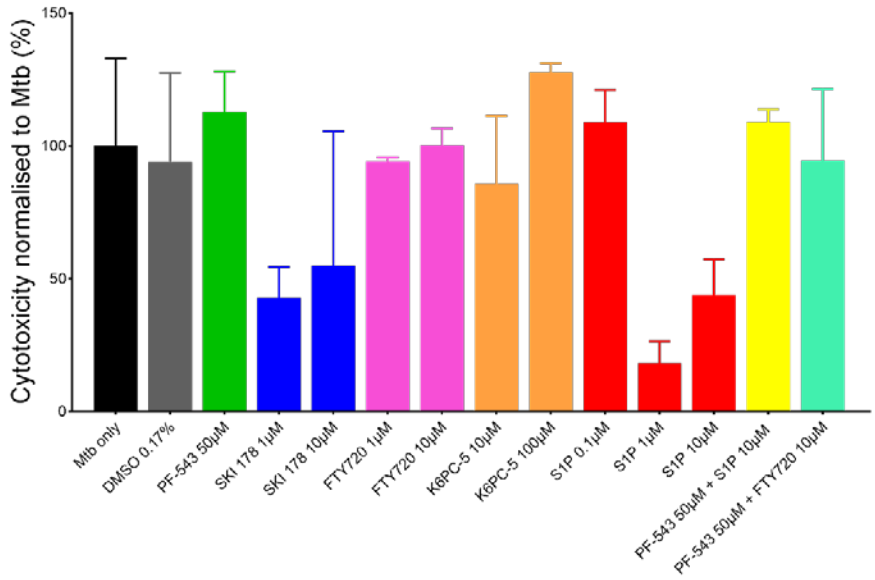


Figure 74: Effect of compounds acting on the sphingolipid signalling pathway on cellular toxicity on day 7 using CytoTox-Glo assay, normalised to Mtb alone. *Comparison of inhibitors with their relevant control (Mtb only or DMSO 0.17%) finds no significant cell toxicity. Mean values of triplicate are shown. Analysis by Student's t test, error bars show standard deviation.*

Published works have used K6PC-5 at a concentration of 10µM after having found cytotoxic effects of high concentration K6PC-5 100µM on human cells and mice (Youm *et al.*, 2008; Liu *et al.*, 2018). Together with the dose divergent effects observed at higher concentrations in this study (Figure 73), repeat studies in the 3D Alginate-Collagen model were performed using a 10 fold lower concentration of SphK1 activator K6PC-5. These experiments demonstrate a dose response, particularly after addition of compounds on day 7, in both direct inhibition and activation of SphK1 by PF-543 and K6PC-5, at the optimal concentrations, respectively (Figure 75). The same lack of effect on Mtb growth by SKI 178, FTY720 and S1P was also observed.

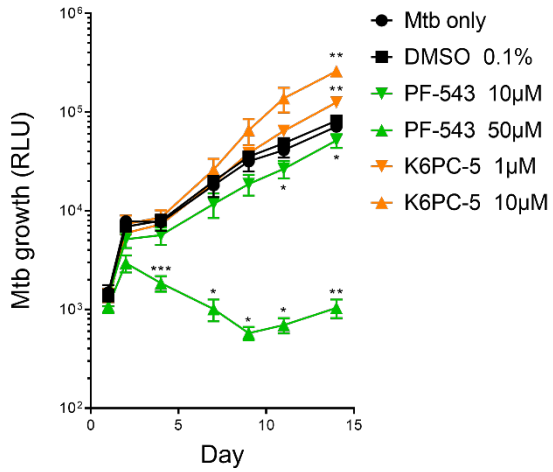


Figure 75: Inhibition of SphK1 by PF-543 and activation by K6PC-5 have divergent and dose dependent responses on Mtb growth. *Analysis by two way ANOVA, error bars show standard deviation. *p<0.05, **p<0.01, ***p<0.001*

So far, SphK1 certainly appeared to have a key role in regulation of Mtb growth. Therefore modulation of SphK1 was expected to associate with changes in secreted matrix metalloproteinases (MMPs) and cytokines implicated in TB pathogenesis. Luminex analysis was performed using a multiplex array using 96 well plate to measure levels of human MMPs and cytokines from supernatant taken on day 7 after Mtb infection. A total of four donors whose PBMCs were used in 3D Alginate-Collagen model experiments were used, three of which had PF-543 and two which had K6PC-5 added, with each condition in triplicate for each donor. Results were considered reliable where levels of MMP and cytokine were detectable in the Mtb control, and trends consistent across donors. The results from one donor are shown in Figure 76.

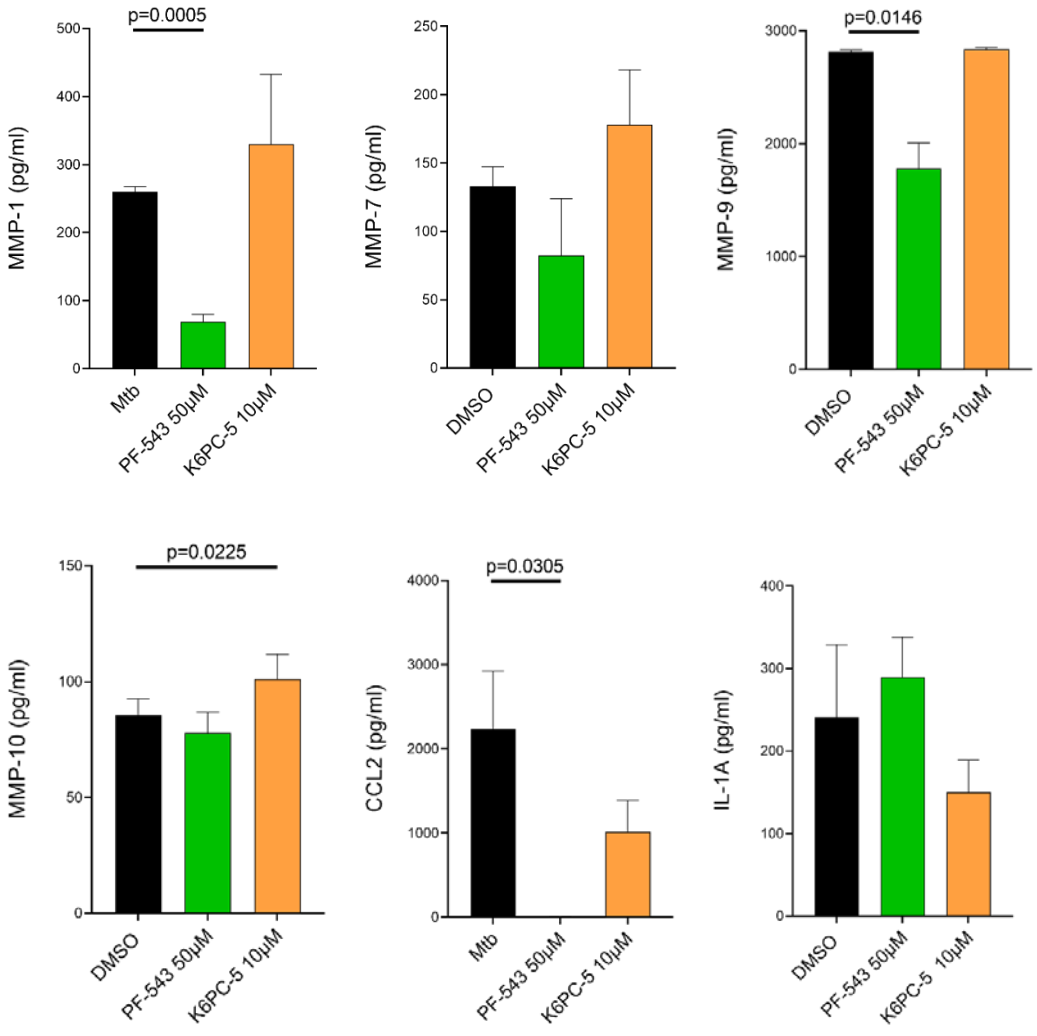


Figure 76: Matrix metalloproteinase (MMP) and cytokine levels taken from Mtb infected microsphere supernatant on day 7, measured by Luminex. *Inhibition and activation of SphK1 have opposing effects on MMP and cytokine secretion, including significantly lower concentrations of MMP-1 and CCL2. Analysis by Student's t test, error bars show standard deviation.*

Divergent effects of SphK1 inhibition and activation are seen in levels of both MMPs and cytokines, namely MMP-1, MMP-7, MMP-9, MMP-10, CCL2 (C-C motif chemokine ligand 2, previously known as

MCP-1, monocyte chemotactic protein 1), and IL-1A (Figure 76). Notably, MMP-1, MMP-9 and CCL2 genes are significantly upregulated in clinical TB samples relative to control, with MMP-1 and CCL2 differentiating TB and sarcoidosis in cytokine-mediated signalling pathway (Figure 24). Thus suppression of MMP-1 and CCL2 is likely to be an important correlation with inhibition of SphK1 by PF-543 and control of Mtb growth.

Finally, to confirm the presence of SphK1 in TB granulomas, human lung TB sections were stained with SphK1 antibody (Figure 77). SphK1 expression is seen in macrophages and multinucleate giant cells within and surrounding the granulomas. The control of Mtb growth with SphK1 inhibition, increase in Mtb growth with SphK1 activation, suppression of important secreted proteins such as MMP-1 and CCL2 with SphK1 inhibition, and definitive expression in human lung TB granulomas, highlights a major role for SphK1 in TB pathogenesis.

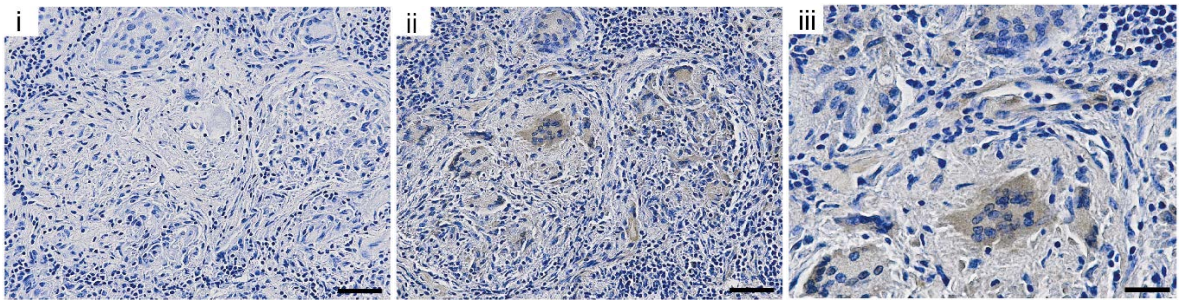


Figure 77: Immunohistochemistry of human lung TB granulomas stained with SphK1 antibody. *No reactivity is observed in the absence of antibody (**i**), while SphK1 expression is demonstrated in a subset of macrophages and multinucleate giant cells (brown stain, **ii** and **iii**). Scale bars **i** and **ii**: 50µm, **iii**: 25µm.*

5.2.4 Determining the mechanism by which sphingosine kinase 1 regulates M.tuberculosis

The next step was to determine the host mechanism by which SphK1 inhibition controls Mtb growth. It was interesting to observe the effect of PF-543 occurred within 2 hours of adding the compound to the microspheres on day 1. The first reading of Mtb luminescence was consistently lower than control microspheres in every donor as shown earlier (Figure 67, Figure 73). Hence SphK1 inhibition may affect cellular processes which occur early and rapidly in the host-pathogen interaction. I then utilised STRING (Search Tool for the Retrieval of Interacting Genes/Proteins), an online database for protein interaction network analysis, to find out more about the known biological roles of SphK1. The top 10 proteins which interact with SphK1 are illustrated in Figure 78.

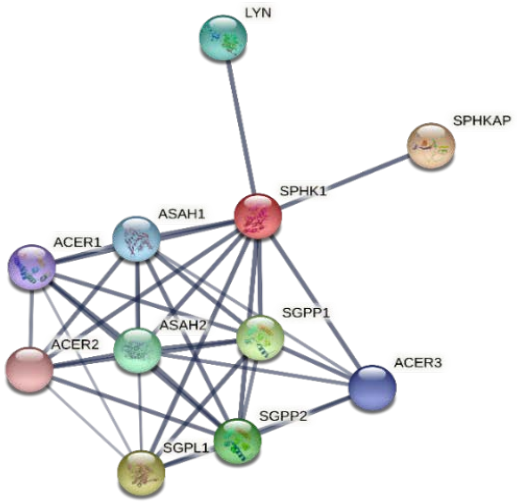


Figure 78: STRING network of top 10 protein interactions with SPHK1. *ASAH1 (light blue), LYN (dark green) and SGPL1 (yellow) are significantly upregulated in clinical TB samples. Edge confidence demonstrated with grey line thickness of 0.9 (thick) and 0.7 (thin).*

Three of the top ten interacting proteins are significantly upregulated in clinical TB samples: *ASAH1* (N-acylsphingosine amidohydrolase 1); *LYN* (Lck/Yes-related novel protein tyrosine kinase); and *SGPL1* (S1P lyase 1) (Table 23). *ASAH1* encodes a lysosomal enzyme involved with sphingosine metabolism, catalysing the degradation of ceramide into sphingosine and free fatty acid. *LYN* encodes a tyrosine kinase involved in the regulation of multiple cellular processes including apoptosis, migration and chemokine response. *SGPL1* encodes sphingosine-1-phosphate lyase 1, catalysing the breakdown of sphingosine-1-phosphate (S1P), the product of SphK1 action, into fatty aldehydes and PEA (phosphoethanolamine), as identified earlier in the sphingosine signalling pathway (Figure 71).

Gene symbol	Gene name	Function	Log_2 FC	Adj. P value
ASAH1	N-Acylsphingosine Amidohydrolase 1	Lysosomal enzyme which degrades ceramide into sphingosine and free fatty acid.	1.41	3.10E-04
LYN	Lck/Yes-Related Novel Protein Tyrosine Kinase	Regulation of proliferation, apoptosis, degranulation, migration, adhesion, response to chemokines and growth factors, and initiation of B cell response.	0.61	4.03E-02
SGPL1	Sphingosine-1-Phosphate Lyase 1	Cleaves phosphorylated sphingoid bases, e.g. S1P, into fatty aldehydes and phosphoethanolamine.	1.05	4.29E-04

Table 23: DEGs in the top 10 protein interactions with SPHK1, identified by STRING, found in clinical TB relative to control samples. *Genes encode proteins with functions associated with ceramide and S1P break down and the immune response.*

Other lysosomal enzymes, including α-galactosidase (GLA), galactosylceramidase (GALC), β1-galactosidase (GLB1) and glycosylceramidase beta (GBA), in the ceramide pathway, are also significantly upregulated in clinical TB samples (Figure 79). Ceramide production is upstream of S1P production (Figure 71), and it is difficult to know which components of pathways are directly

affected by Mtb and which are due to the host response. Clearly the balance of ceramide generation and conversion to sphingolipids is complex, and overexpression of lysosomal enzymes may be part of the host attempt to regulate a disrupted lysosome system. Interestingly, one of the genes of the TB unique Cluster 21, *LGALS17A* (Figure 19), encodes a galectin, a family of proteins exhibiting antimicrobial activity and which have an affinity for β-galactosides, the substrate for GLB1 (Nabi, Shankar and Dennis, 2015).

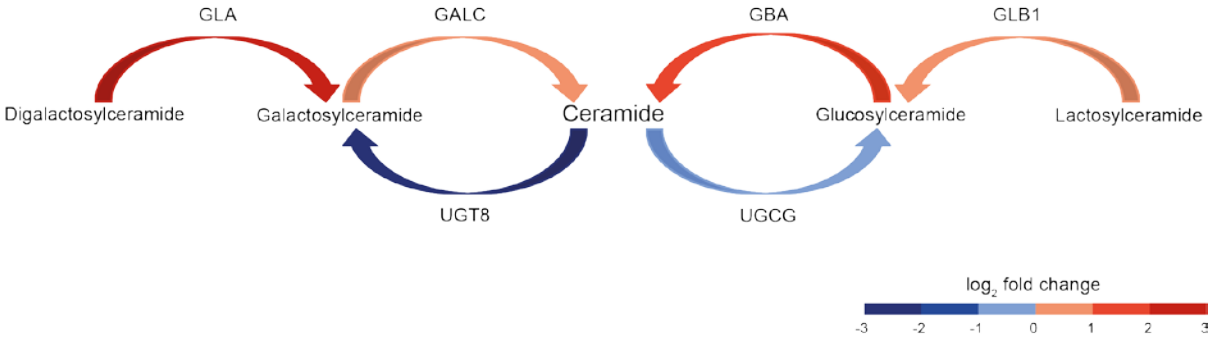


Figure 79: Graphical representation of changes in sphingolipid metabolism pathway genes in TB granulomas relative to control lymph nodes (adjusted P value < 0.05). *Upregulation of genes involved with ceramide production is seen. GALC, galactosylceramidase; GBA, glycosylceramidase beta; GLA, galactosidase alpha; GLB1, galactosidase beta 1; UGCG, UDP-glucose ceramide glucosyltransferase; UGT8, UDP glycosyltransferase 8)*

Altogether, these analyses provided evidence implicating the lysosome as the intracellular location through which SphK1 may regulate Mtb. Firstly, the lysosome is the only cellular component expressed to a lesser degree in TB relative to sarcoidosis on correlation analysis (Table 8). Secondly, the lytic vacuole is the only downregulated cellular component in TB relative to sarcoidosis on gene set enrichment analysis (Figure 35). Thirdly, the lysosome is the only KEGG pathway downregulated in TB in comparison to sarcoidosis on gene set enrichment analysis (Figure 36). Disruption of the phagolysosomal process may also explain why the effect of SphK1 inhibition on Mtb growth is rapid (Figure 66, Figure 75). To control Mtb, human macrophages need to phagocytose the bacilli and fuse with lysosomes to acidify and degrade Mtb, which Mtb inhibits (Russell, 2003). Hence, it was hypothesised that SphK1 inhibition may accelerate phagolysosomal fusion within Mtb infected cells. The final set of experiments here aimed to address this hypothesis, using monocytes stained with pH sensitive dye.

Monocytes were isolated from fresh PBMCs and stained with pH sensitive dye, pHrodo, for 30 minutes at 37°C, and transferred to a 96 well plate where half the wells were infected with Mtb H37Rv. Monocytes were infected with a multiplicity of infection (MOI) of 1, aiming for each monocyte to be infected. Fluorescence readings were taken by GloMax Discover every 5 minutes for 4 hours at 37°C, then incubated overnight for a final reading at 24 hours. Unfortunately, the first

experiment measured emitted fluorescence at a suboptimal wavelength 500-550 nm, so the only reading obtained was at 24 hours when wavelength was changed to 580-640 nm (Figure 80).

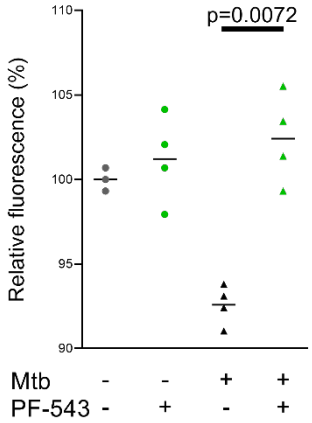


Figure 80: Relative fluorescence signal in human monocytes labelled with intracellular pH stain pHrodo taken 24 minutes after Mtb infection with or without concurrent PF-543 50μM addition *Increased fluorescence signal indicates lower pH. Mtb infection increases intracellular pH, and the effect is reversed with SphK1 inhibition by PF-543. Normalized data from one donor, analyzed by paired t test, horizontal bars indicate the mean.*

The fluorescence recorded at 24 hours found Mtb infected monocytes to be lower than uninfected monocytes, indicating a higher pH with Mtb infection. This effect was reversed when SphK1 inhibitor, PF-543, was present. Repeat experiments with two donors and optimised emission wavelength utilised fresh monocytes immediately infected with Mtb as previously. A second intracellular pH assay, LysoSensor, was used to confirm the findings from the pHrodo assays, where frozen monocytes were thawed, transferred to a 96 well plate for incubation at 37°C for 2 hours, then infected with Mtb H37Rv for a further 1 hour at 37°C. In both assays, Mtb infection was quickly followed by the addition of compounds occurred in quick succession, and fluorescence readings were taken every 5 minutes for 40 minutes (Figure 81).

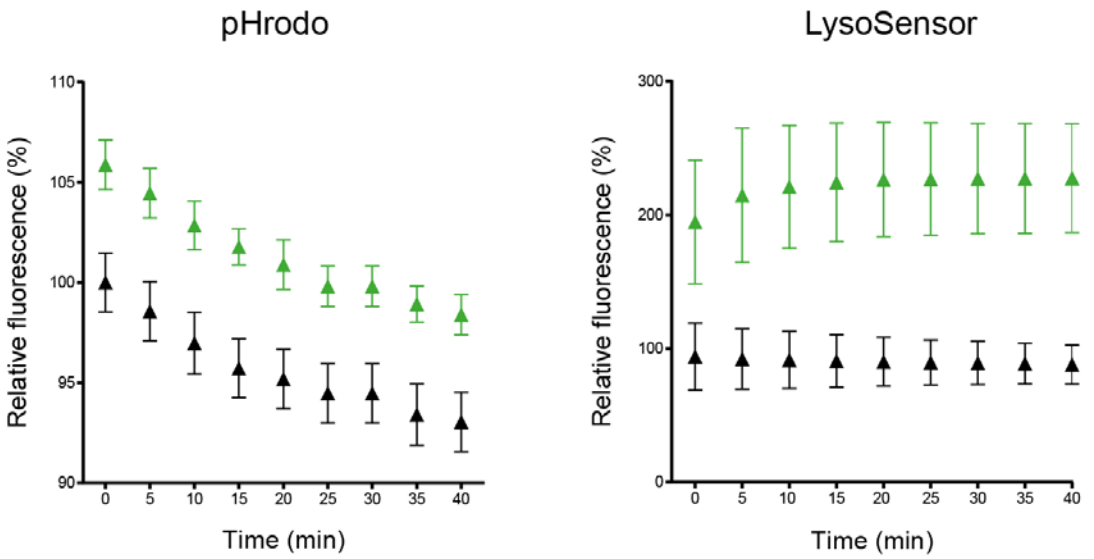


Figure 81: Relative fluorescence signal in human monocytes labelled with intracellular pH stains pHrodo or LysoSensor measured for 40 minutes after Mtb infection. *Higher fluorescence readings indicate more acidic (lower pH) conditions. Fluorescence measured at 5 minute intervals after Mtb infection for 40 minutes, treated with DMSO 0.1% (black triangles) or PF-543 50μM (green triangles). SphK1 inhibition by PF-543 lowers intracellular pH, with the effect lasting at least 40 minutes. Normalized data shown from one donor, triangles indicate mean, vertical bars indicate standard deviation.*

The addition of PF-543 quickly results in reversal of fluorescence signal from less acidic conditions (lower fluorescence signal) in Mtb infected monocytes to more acidic environment. The speed at which SphK1 inhibition alters the intracellular pH may relate to the persistently lower Mtb readings observed in the microsphere experiments taken straight after PF-543 addition on day 1. The fluorescence signal at the 5 minute time point is shown using readings from 2 donors (Figure 82), demonstrating significant increase with the addition of PF-543 in Mtb infected monocytes.

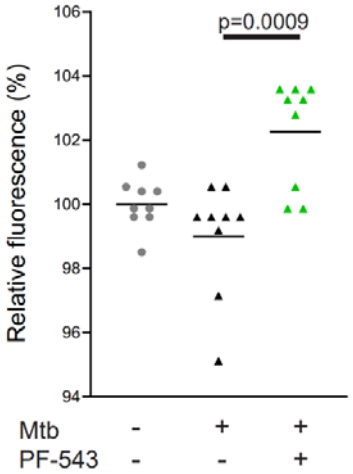


Figure 82: Relative fluorescence signal in human monocytes labelled with intracellular pH stain pHrodo taken 5 minutes after Mtb infection with or without concurrent PF-543 50μM addition. *Increased fluorescence signal indicates lower pH. SphK1 inhibition by PF-543 lowers intracellular pH. Normalized data from two separate donors, analyzed by paired t test, horizontal bars indicate the mean.*

These last experiments provide evidence for an important role of SphK1 in the disruption of intracellular pH regulation during Mtb infection, and will require further investigation to further unpick the mechanism.

5.3 Discussion

Integration of the clinical and 3D Alginate-Collagen RNAseq studies identified shared differentially expressed genes (DEGs) (Figure 63) which underwent a methodical selection process to generate twelve intracellular host target enzymes (Table 19). Selective inhibition of these target enzymes showed sphingosine kinase 1 (SphK1) inhibition by PF-543 to cause a striking reduction in Mtb growth which is dose-dependent and mediated by the host cell (Figure 66). Furthermore, selective activation of SphK1 has the opposite action, significantly increasing Mtb growth in a dose-dependent manner (Figure 75). Luminex analysis of supernatant taken on day 7 of Mtb infected microspheres shows divergent effects of inhibition and activation of SphK1 on MMP-1 (matrix metalloproteinase 1) and CCL2 (C-C motif chemokine ligand 2, previously known as *MCP1*, monocyte chemotactic protein 1) (Figure 76), both identified on gene ontology analysis of clinical TB samples relative to sarcoidosis samples (Figure 27). Immunohistochemistry of human lung TB samples confirm the presence of SphK1 within granuloma-associated multinucleate giant cells as well as a subset of surrounding macrophages (Figure 77). Intracellular pH staining of fresh monocytes suggests disruption of lysosomal function by Mtb which is reversed by inhibition of SphK1 (Figure 82), putting forward a potential mechanism by which SphK1 inhibition controls Mtb growth.

A systematic screening approach was crucial to identifying SphK1 as a potential target (Figure 64). Using all significant DEGs upregulated in the 3D Alginate-Collagen model without a fold change threshold maximised the number of potential targets included. I focussed on upregulated intracellular enzymes, taking the example of aspirin, brinzolamide, and imatinib, as they are suitable for resource-poor settings given they are orally available and mechanistically easier to target (Holtzman, 1992; Robertson, 2005). Prioritising selective drugs proven safe in early clinical trials or in mammals was another important step in selecting appropriate targets to study (Hawn and Vandal, 2013). However, the challenge of identifying putative targets among hundreds of significant DEGs, especially with high degrees of fold change, is evident (Byron *et al.*, 2016) and leaves many potential targets yet to be studied.

Studying the host response to another organism presents further challenges. Here, the *SPHK1* gene has a large overlap with Mtb's diacylglycerol kinase, bringing into question whether PF-543 is targeting the host or Mtb itself (Table 21), so this needed excluding. We concluded when PF-543 was shown to have no effect on Mtb grown in broth (Figure 70). Thus PF-543 requires the presence

of host cells to control Mtb growth. A dose-dependent effect on Mtb growth is also seen with another compound, SSM3, a furin inhibitor (Figure 67). The *FURIN* gene also has a great deal of overlap with a Mtb protease, including the subtilisins (Table 21). However, in sharp contrast to SphK1 inhibition, furin inhibition leads to reduced Mtb growth in the absence of host cells (Figure 70). SSM3 is likely to be acting directly on the Mtb serine proteases, which include the mycosin protease family (*mycP*), highly conserved subtilisin-like serine proteases found in Mtb to be essential for type VII secretion system function (Brown *et al.*, 2000; van Winden *et al.*, 2019). Additionally, where host cells are available, SSM3 may also be acting on the host furin enzyme, a ubiquitous proprotein convertase with a broad range of substrates including MMP-1 (Thomas, 2002). Of note, furin has recently been proposed as one of eight proteins in a diagnostic biosignature for TB (Yang *et al.*, 2020).

SphK1 is a critical enzyme in the sphingosine-ceramide cellular rheostat, which modulates diverse cellular processes (Spiegel and Milstien, 2003). SphK1 phosphorylates sphingosine to sphingosine-1-phosphate (S1P), which is then irreversibly degraded by S1P lyase, both enzymes of which are significantly upregulated in clinical TB samples (Figure 71). S1P is a signalling molecule with a complex mechanism of action dependent upon its intracellular or extracellular location, and regulates diverse processes including transcription, angiogenesis, atherosclerosis, apoptosis, and autophagy (Maceyka *et al.*, 2012). Together with the short half life of S1P, these factors may explain why addition of S1P had no effect on Mtb growth in the microsphere model. Interestingly, FTY720 otherwise known as fingolimod, is an inhibitor of S1P receptors already in clinical use for relapsing and remitting multiple sclerosis (Chun and Hartung, 2010). FTY720 demonstrated non-significant control of Mtb growth and is likely to be influenced by S1P localisation (Figure 73). Ultimately, by diverting the sphingolipid pathway away from S1P synthesis, SphK1 inhibition will concurrently increase cellular ceramide levels (Schnute *et al.*, 2012).

SphK1 inhibition suppresses Mtb growth surprisingly rapidly, suggesting the mechanism may be through increasing phagolysosomal fusion, which Mtb inhibits as part of its intracellular survival strategy (Vergne *et al.*, 2005; Russell, 2011). Consistent with this conclusion, SphK1 inhibition normalizes intracellular pH, and previous studies have shown that ceramide accumulation increases phagolysosomal fusion. Ceramide accumulates causes Mtb killing by regulating actin nucleation (Anes *et al.*, 2003), although a report from an alternative 2D model systems differs from this conclusion (Thompson *et al.*, 2005). Taking these findings with my observations suggests that

directing sphingosine metabolism to S1P, and away from ceramide, is part of Mtb's intracellular survival strategy, and reversing this process through SphK1 inhibition increases Mtb killing.

The role of MMP-1 and CCL2 is highlighted both by bioinformatic analysis (Figure 24) and microsphere studies, whereby SphK1 inhibition results in reduced secretion of these molecules (Figure 76). A key question is whether these secreted proteins are involved in the pathway which leads SphK1 inhibition to killing Mtb, and if indeed they are critical to the killing of Mtb death or simply a product of the pathway, or if they are induced by Mtb death itself. To address the former suggestion, animal models are likely to be required and is discussed below, whereas experiments using UV-irradiated Mtb could help answer if Mtb death is responsible for the secretion of these matrix metalloproteinases and chemokines.

Translating the findings from an *in vitro* model to clinical use presents a huge challenge for all potential therapeutic candidates (Ledford, 2011). The 3D Alginate-Collagen model is limited by lack of cell diversity and organ systems which exist in animal models. A reasonable next step would be to use a mouse model to study the mammalian response to Mtb. SphK1 knockout mice have been used to study S1P signalling (Allende *et al.*, 2004), and humanised models such as SGM3 (Stem cell factor, Granulocyte-macrophage colony-stimulating factor, and Interleukin 3 expressing) mice used to study cytokine release and neurotoxicity (Norelli *et al.*, 2018). However, these mouse models may still be limited by the absence of lung cavitation during Mtb infection and lack of functioning MMP-1 ortholog (Balbin *et al.*, 2001). Therefore an alternative animal model such as the rabbit model may be a better choice, which expresses MMPs and develops cavities (Kubler *et al.*, 2015). In due course, clinical trials will be necessary to ensure drug safety and benefit to the patient, the latter of which requires months of treatment and monitoring in the case of TB. It is likely that the therapeutic use of any host-directed treatment will be an adjunct to anti-microbial therapy in TB. Clearly the move to phase II clinical trials is required, in conjunction with overcoming the complexities of model limitations.

Chapter 6: Conclusion and future work

6.1 Summary

This work is likely to be the first published unbiased analysis of mediastinal lymph node TB, without the confounder of prior antibiotic or immunosuppressant treatment. Firstly, I performed RNA sequencing of granulomas taken from pre-treatment clinical TB and sarcoidosis lymph node samples, and control lymph node tissue. In parallel, I undertook a cellular study of three humanised culture models of TB, linking both studies by the same optimised bioinformatic pipeline. The clinical study demonstrates a significant overlap between TB and sarcoidosis samples. However, a cluster composed of seven genes, including *MMP1*, is unique to TB on correlation analysis and associates with overexpression of extracellular matrix turnover and the inflammatory response, similar processes shared by TB and sarcoidosis using differential gene expression analysis but expressed to a much higher degree in TB. Similarly, angiogenesis appears to be a defining feature in TB samples. In contrast, sarcoidosis is characterised by overexpression of the lysosome, a feature shared with TB though to much less extent.

The cellular study compared the 2D cell culture, 3D Alginate model and 3D Alginate-Collagen model, each incorporating donor PBMCs. The 3D Alginate-Collagen model best reflects clinical TB samples on transcriptomic analysis, hence it was later used in target selection experiments. Therefore I employed both a bioinformatic and experimental screening process to identify potential therapeutic targets. Structured exploration of new host-directed therapies found SphK1 to significantly control Mtb growth, whose expression is confirmed on immunohistochemistry of TB patient samples. Using the same approach, there are potentially many more host-directed targets to be uncovered from this work.

6.2 Future work

My PhD trained me in translational medical research, including analysis of clinical samples by unbiased techniques, bioinformatics analysis and advanced cell culture models. I am planning to stay in academic medicine, and hope to obtain an Academic Clinical Lecturer position or further funding to do this. There are a number of ways I would like to take this work forward, although each approach represents a significant programme of work and so I will need to focus my aims over time.

i) SphK1 translation pathway

Two approaches could be undertaken to further determine the role of SphK1 in TB. Moving to an animal model is likely to be necessary, and mouse experiments are currently running in collaboration with Adrie Steyn's laboratory to see if PF-543 has an effect on Mtb-infected mice. Other animal models to consider are the SphK1 knockout mouse, humanised NSG-SGM3 mouse, or rabbit. Alternatively, I could explore the cellular biology of SphK1 by modulating the pathway with the addition of ceramide, or modify cells using CRISPR-Cas9 technology to alter the genome to disrupt the sphingosine-ceramide rheostat at targeted points. Integration with imaging such as confocal microscopy could also study the real-time intracellular localisation of various components of the SphK1 pathway during Mtb infection.

ii) Disease comparisons using LCM-RNAseq approach

My approach of Laser Capture Microdissection followed by RNA sequencing (LCM-RNAseq) to understand disease mechanism can be applied to many conditions. One approach would be to compare Mtb granulomas in different loci, such as the lung versus lymph node, and different clinical manifestations, such as lung versus miliary, using TB meningitis samples as the source of the miliary granulomas. My work has gained insight into the mechanism of sarcoidosis, and further bioinformatic analysis directly comparing sarcoidosis and TB may find more features differentiating these diseases. Using a larger cohort of patients in different clinical stages of sarcoidosis would also allow possible characterisation of the transcriptomic signature at these stages of clinical disease. The next step would be to develop a model for sarcoidosis. Current sarcoidosis models range from unstimulated PBMC-based models to animal models. The Schlesinger laboratory have developed a popular sarcoidosis model utilising coated beads to stimulate PBMCs, ironically with Mtb, to cause granuloma formation (Crouser *et al.*, 2017). Once established, a model of sarcoidosis could be used to better understand the role of the lysosome in the disease. Similarly, the LCM-RNAseq approach should be employed to study basic granuloma biology, applying it to other granulomatous conditions such as Chronic Granulomatous Disease and other infectious granulomatous diseases including schistosomiasis and leishmaniasis.

iii) More carefully defined microenvironments

The temporal and geographical variations of granulomas likely represent different clinical states of TB. I studied the entire granuloma by laser capture, but in fact this probably encapsulates a number of micro-environments itself. My current data can be processed by cell sorting analysis using tools such as CIBERSORT to determine cellular composition. The progression of TB could be then be studied by comparative analysis of intact, caseating, and fibrosing granulomas, and

surrounding tissue, spatially guided by Micro-Computed Tomography or adjacent histological sections to direct microenvironment specific laser capture. Other parameters including quantitative PCR for mycobacterial load could be observed. Furthermore, cross-correlative analysis using single cell RNA sequencing of TB lung specimens to compare peripheral tissue and site of disease would confirm the cellular composition and associated protein secretion. Patient samples obtained by Alisdair Leslie's laboratory are likely to be amenable to this analysis. Therefore, analysis of clinical specimens from different microenvironments and compartments may further build the insights gained in my analysis to date.

iv) Model development

Despite the success of 3D Alginate-Collagen microsphere model, the absence of upregulated pathways relating to extracellular matrix organisation needs addressing (Figure 62). The addition of other cellular components including fibroblasts seems likely to be required to optimise the model, though will need to be derived from the same donor providing PBMCs to prevent MHC incompatibility. Lung derived cells such as alveolar epithelial and mesenchymal cells, or the use of human pluripotent stem cells to generate these cells, could supplement the addition of fibroblasts. Other donor factors including smoking and HIV status could be studied to see the impact on Mtb and associated host response.

Bibliography

Agrawal, R. *et al.* (2016) 'Tuberculosis or sarcoidosis: Opposite ends of the same disease spectrum?', *Tuberculosis (Edinb)*, 98, pp. 21-6.
Al Shammari, B. *et al.* (2015) 'The Extracellular Matrix Regulates Granuloma Necrosis in Tuberculosis', *J Infect Dis*, 212(3), pp. 463-73.
Alhamdoosh, M. *et al.* (2017) 'Combining multiple tools outperforms individual methods in gene set enrichment analyses', *Bioinformatics*, 33(3), pp. 414-424.
Allende, M.L. *et al.* (2004) 'Mice deficient in sphingosine kinase 1 are rendered lymphopenic by FTY720', *J Biol Chem*, 279(50), pp. 52487-92.
Anes, E. *et al.* (2003) 'Selected lipids activate phagosome actin assembly and maturation resulting in killing of pathogenic mycobacteria', *Nat Cell Biol*, 5(9), pp. 793-802.
Archer, N.S., Nassif, N.T. and O'Brien, B.A. (2015) 'Genetic variants of SLC11A1 are associated with both autoimmune and infectious diseases: systematic review and meta-analysis', *Genes Immun*, 16(4), pp. 275-83.
Bach, L.A. (2005) 'IGFBP-6 five years on; not so 'forgotten'?', *Growth Horm IGF Res*, 15(3), pp. 185-92.
Balbin, M. *et al.* (2001) 'Identification and enzymatic characterization of two diverging murine counterparts of human interstitial collagenase (MMP-1) expressed at sites of embryo implantation', *J Biol Chem*, 276(13), pp. 10253-62.
Barber, D.L. *et al.* (2019) 'Tuberculosis following PD-1 blockade for cancer immunotherapy', *Sci Transl Med*, 11(475).
Baughman, R.P. *et al.* (2001) 'Clinical characteristics of patients in a case control study of sarcoidosis', *American Journal of Respiratory and Critical Care Medicine*, 164(10), pp. 1885-1889.
Bekkering, S. *et al.* (2014) 'Oxidized low-density lipoprotein induces long-term proinflammatory cytokine production and foam cell formation via epigenetic reprogramming of monocytes', *Arterioscler Thromb Vasc Biol*, 34(8), pp. 1731-8.
Bellamy, R. *et al.* (1998) 'Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans', *N Engl J Med*, 338(10), pp. 640-4.
Bray, N.L. *et al.* (2016) 'Near-optimal probabilistic RNA-seq quantification', *Nat Biotechnol*, 34(5), pp. 525-7.
Briken, V. *et al.* (2000) 'Human CD1b and CD1c isoforms survey different intracellular compartments for the presentation of microbial lipid antigens', *J Exp Med*, 192(2), pp. 281-8.
Brites, D. and Gagneux, S. (2015) 'Co-evolution of Mycobacterium tuberculosis and Homo sapiens', *Immunol Rev*, 264(1), pp. 6-24.
Brown, G.D. *et al.* (2000) 'The mycosins of Mycobacterium tuberculosis H37Rv: a family of subtilisin-like serine proteases', *Gene*, 254(1-2), pp. 147-55.
Brownell, I. *et al.* (2011) 'Evidence for mycobacteria in sarcoidosis', *Am J Respir Cell Mol Biol*, 45(5), pp. 899-905.
Byron, S.A. *et al.* (2016) 'Translating RNA sequencing into clinical diagnostics: opportunities and challenges', *Nat Rev Genet*, 17(5), pp. 257-71.
Cali, J.J. *et al.* (1991) 'Mutations in the bile acid biosynthetic enzyme sterol 27-hydroxylase underlie cerebrotendinous xanthomatosis', *J Biol Chem*, 266(12), pp. 7779-83.
Cardona, P.J. (2010) 'Revisiting the natural history of tuberculosis. The inclusion of constant reinfection, host tolerance, and damage-response frameworks leads to a better understanding of latent infection and its evolution towards active disease', *Arch Immunol Ther Exp (Warsz)*, 58(1), pp. 7-14.
Casanova, N.G. *et al.* (2020) 'Differential transcriptomics in sarcoidosis lung and lymph node granulomas with comparisons to pathogen-specific granulomas', *Respir Res*, 21(1), p. 321.
Chen, E.S. and Moller, D.R. (2015) 'Etiologies of Sarcoidosis', *Clinical Reviews in Allergy & Immunology*, 49(1), pp. 6-18.

Chen, E.S. *et al.* (2010) 'Serum amyloid A regulates granulomatous inflammation in sarcoidosis through Toll-like receptor-2', *Am J Respir Crit Care Med*, 181(4), pp. 360-73.
Chen, J. *et al.* (2009) 'ToppGene Suite for gene list enrichment analysis and candidate gene prioritization', *Nucleic Acids Res*, 37(Web Server issue), pp. W305-11.
Christofori, G. (2003) 'Changing neighbours, changing behaviour: cell adhesion molecule-mediated signalling during tumour progression', *EMBO J*, 22(10), pp. 2318-23.
Chun, J. and Hartung, H.P. (2010) 'Mechanism of action of oral fingolimod (FTY720) in multiple sclerosis', *Clin Neuropharmacol*, 33(2), pp. 91-101.
Clayton, K. *et al.* (2017) 'Gene Expression Signatures in Tuberculosis Have Greater Overlap with Autoimmune Diseases Than with Infectious Diseases', *Am J Respir Crit Care Med*, 196(5), pp. 655-656.
Conesa, A. *et al.* (2016) 'A survey of best practices for RNA-seq data analysis', *Genome Biol*, 17, p. 13.
Costa-Silva, J., Domingues, D. and Lopes, F.M. (2017) 'RNA-Seq differential expression analysis: An extended review and a software tool', *Plos One*, 12(12).
Costabel, U. and Hunninghake, G.W. (1999) 'ATS/ERS/WASOG statement on sarcoidosis. Sarcoidosis Statement Committee. American Thoracic Society. European Respiratory Society. World Association for Sarcoidosis and Other Granulomatous Disorders', *Eur Respir J*, 14(4), pp. 735-7.
Crouser, E.D. *et al.* (2017) 'A Novel In Vitro Human Granuloma Model of Sarcoidosis and Latent Tuberculosis Infection', *Am J Respir Cell Mol Biol*, 57(4), pp. 487-498.
Dahlstrand, J. *et al.* (1992) 'Characterization of the human nestin gene reveals a close evolutionary relationship to neurofilaments', *J Cell Sci*, 103 (Pt 2), pp. 589-97.
Davis, J.M. and Ramakrishnan, L. (2009) 'The role of the granuloma in expansion and dissemination of early tuberculous infection', *Cell*, 136(1), pp. 37-49.
Divangahi, M., Khan, N. and Kaufmann, E. (2018) 'Beyond Killing Mycobacterium tuberculosis: Disease Tolerance', *Front Immunol*, 9, p. 2976.
Dobin, A. *et al.* (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1), pp. 15-21.
Drain, P.K. *et al.* (2018) 'Incipient and Subclinical Tuberculosis: a Clinical Review of Early Stages and Progression of Infection', *Clin Microbiol Rev*, 31(4).
Dubos, R.J. and Dubos, J. (1987) *The white plague : tuberculosis, man, and society*. New Brunswick: Rutgers University Press.
Durinck, S. *et al.* (2009) 'Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt', *Nat Protoc*, 4(8), pp. 1184-91.
Elkington, P. and Friedland, J.S. (2011a) 'MMP-1 drives immunopathology in human tuberculosis and transgenic mice ', *Journal of Clinical Investigation*, 121(5), pp. 1827-1833.
Elkington, P. and Friedland, J.S. (2011b) 'Tuberculosis Immunopathology: The Neglected Role of Extracellular Matrix Destruction ', *Science Translational Medicine*, 3(71).
Elkington, P. *et al.* (2019) 'In Vitro Granuloma Models of Tuberculosis: Potential and Challenges', *J Infect Dis*.
Elkington, P., Tebruegge, M. and Mansour, S. (2016) 'Tuberculosis: An Infection-Initiated Autoimmune Disease?', *Trends Immunol*.
Elkington, P.T. *et al.* (2018) 'Implications of Tuberculosis Reactivation after Immune Checkpoint Inhibition', *Am J Respir Crit Care Med*.
Engler, A.J. *et al.* (2006) 'Matrix elasticity directs stem cell lineage specification', *Cell*, 126(4), pp. 677-89.
Fay, W.P. *et al.* (1997) 'Human plasminogen activator inhibitor-1 (PAI-1) deficiency: characterization of a large kindred with a null mutation in the PAI-1 gene', *Blood*, 90(1), pp. 204-8.
Fennelly, K.P. *et al.* (2012) 'Variability of infectious aerosols produced during coughing by patients with pulmonary tuberculosis', *Am J Respir Crit Care Med*, 186(5), pp. 450-7.
Flynn, J.L. and Chan, J. (2001) 'Immunology of tuberculosis', *Annu Rev Immunol*, 19, pp. 93-129.
Fonseca, K.L. *et al.* (2017) 'Experimental study of tuberculosis: From animal models to complex cell systems and organoids', *PLoS Pathog*, 13(8), p. e1006421.

Frieden, T.R. *et al.* (2003) 'Tuberculosis', *Lancet*, 362(9387), pp. 887-99.
Gagneux, S. (2012) 'Host-pathogen coevolution in human tuberculosis', *Philos Trans R Soc Lond B Biol Sci*, 367(1590), pp. 850-9.
Godiska, R. *et al.* (1997) 'Human macrophage-derived chemokine (MDC), a novel chemoattractant for monocytes, monocyte-derived dendritic cells, and natural killer cells', *J Exp Med*, 185(9), pp. 1595-604.
Goodwin, S., McPherson, J.D. and McCombie, W.R. (2016) 'Coming of age: ten years of next-generation sequencing technologies', *Nat Rev Genet*, 17(6), pp. 333-51.
Gribbin, J. *et al.* (2006) 'Incidence and mortality of idiopathic pulmonary fibrosis and sarcoidosis in the UK', *Thorax*, 61(11), pp. 980-985.
Guirado, E. *et al.* (2015) 'Characterization of Host and Microbial Determinants in Individuals with Latent Tuberculosis Infection Using a Human Granuloma Model', *Mbio*, 6(1).
Gupta, U.D. and Katoch, V. (2005) 'Animal models of tuberculosis', *Tuberculosis*, 85(5-6), pp. 277-293.
Hansson, G.K. and Libby, P. (2006) 'The immune response in atherosclerosis: a double-edged sword', *Nat Rev Immunol*, 6(7), pp. 508-19.
Hawn, T.R. and Vandal, O. (2013) 'Host-Directed Therapeutics for Tuberculosis: Can We Harness the Host? ', *Microbiology and Molecular Biology Reviews*, 77(4), pp. 608–627.
Holtzman, E. (1992) 'Intracellular Targeting and Sorting: How are macromolecules delivered to specific locations?', *Bioscience*, 42(8), pp. 608-620.
Hong, C.J. *et al.* (2020) 'Fas-apoptotic inhibitory molecule 2 localizes to the lysosome and facilitates autophagosome-lysosome fusion through the LC3 interaction region motif-dependent interaction with LC3', *FASEB J*, 34(1), pp. 161-179.
Houben, R.M. and Dodd, P.J. (2016) 'The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling', *PLoS Med*, 13(10), p. e1002152.
Huang, W. *et al.* (2019) 'Regulatory networks in mechanotransduction reveal key genes in promoting cancer cell stemness and proliferation', *Oncogene*, 38(42), pp. 6818-6834.
Iannuzzi, M.C., Rybicki, B.A. and Teirstein, A.S. (2007) 'Medical progress: Sarcoidosis', *New England Journal of Medicine*, 357(21), pp. 2153-2165.
Jimenez-Corona, M.E. *et al.* (2006) 'Gender differentials of pulmonary tuberculosis transmission and reactivation in an endemic area', *Thorax*, 61(4), pp. 348-53.
Judson, M.A. (1999) 'An approach to the treatment of pulmonary sarcoidosis with corticosteroids - The six phases of treatment', *Chest*, 115(4), pp. 1158-1165.
Kakumanu, P. *et al.* (2008) 'Patients with pulmonary tuberculosis are frequently positive for anti-cyclic citrullinated peptide antibodies, but their sera also react with unmodified arginine-containing peptide', *Arthritis Rheum*, 58(6), pp. 1576-81.
Kapoor, N. *et al.* (2013) 'Human Granuloma In Vitro Model, for TB Dormancy and Resuscitation', *Plos One*, 8(1).
Keane, J. *et al.* (2001) 'Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent', *N Engl J Med*, 345(15), pp. 1098-104.
Kim, D. *et al.* (2019) 'Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype', *Nat Biotechnol*, 37(8), pp. 907-915.
Kim, M.J. *et al.* (2010) 'Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism', *EMBO Mol Med*, 2(7), pp. 258-74.
Koch, R. (1882) 'The etiology of tuberculosis', *Berliner Klinischen Wochenschrift*, 15, pp. 221-230.
Koth, L.L. *et al.* (2011) 'Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis', *Am J Respir Crit Care Med*, 184(10), pp. 1153-63.
Kubler, A. *et al.* (2015) 'Mycobacterium tuberculosis dysregulates MMP/TIMP balance to drive rapid cavitation and unrestrained bacterial proliferation', *J Pathol*, 235(3), pp. 431-44.
Langan, E.A. *et al.* (2020) 'Immune checkpoint inhibitors and tuberculosis: an old disease in a new context', *Lancet Oncol*, 21(1), pp. e55-e65.

Ledford, H. (2011) 'Translational research: 4 ways to fix the clinical trial', *Nature*, 477(7366), pp. 526-8.
Leitinger, B. and Hohenester, E. (2007) 'Mammalian collagen receptors', *Matrix Biol*, 26(3), pp. 146-55.
Li, D. *et al.* (2002) 'LOX-1 inhibition in myocardial ischemia-reperfusion injury: modulation of MMP-1 and inflammation', *Am J Physiol Heart Circ Physiol*, 283(5), pp. H1795-801.
Li, W. *et al.* (2015) 'Comprehensive evaluation of AmpliSeq transcriptome, a novel targeted whole transcriptome RNA sequencing methodology for global gene expression analysis', *BMC Genomics*, 16, p. 1069.
Lin, P.L. and Flynn, J.L. (2010) 'Understanding Latent Tuberculosis: A Moving Target', *Journal of Immunology*, 185(1), pp. 15-22.
Liu, H. *et al.* (2018) 'K6PC-5 Activates SphK1-Nrf2 Signaling to Protect Neuronal Cells from Oxygen Glucose Deprivation/Re-Oxygenation', *Cell Physiol Biochem*, 51(4), pp. 1908-1920.
Locke, L.W., Schlesinger, L.S. and Crouser, E.D. (2020) 'Current Sarcoidosis Models and the Importance of Focusing on the Granuloma', *Front Immunol*, 11, p. 1719.
Maceyka, M. *et al.* (2012) 'Sphingosine-1-phosphate signaling and its role in disease', *Trends Cell Biol*, 22(1), pp. 50-60.
Marakalala, M.J. *et al.* (2016) 'Inflammatory signaling in human tuberculosis granulomas is spatially organized', *Nat Med*, 22(5), pp. 531-8.
Martineau, A.R. *et al.* (2011) 'High-dose vitamin D(3) during intensive-phase antimicrobial treatment of pulmonary tuberculosis: a double-blind randomised controlled trial', *Lancet*, 377(9761), pp. 242-50.
Mattila, J.T. *et al.* (2013) 'Microenvironments in tuberculous granulomas are delineated by distinct populations of macrophage subsets and expression of nitric oxide synthase and arginase isoforms', *J Immunol*, 191(2), pp. 773-84.
McCarthy, D.J., Chen, Y. and Smyth, G.K. (2012) 'Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation', *Nucleic Acids Res*, 40(10), pp. 4288-97.
Mitchell, D.N. *et al.* (1977) 'Sarcoidosis - Histopathological Definition and Clinical Diagnosis', *Journal of Clinical Pathology*, 30(5), pp. 395-408.
Nabi, I.R., Shankar, J. and Dennis, J.W. (2015) 'The galectin lattice at a glance', *J Cell Sci*, 128(13), pp. 2213-9.
Nakagawa, T.Y. and Rudensky, A.Y. (1999) 'The role of lysosomal proteinases in MHC class II-mediated antigen processing and presentation', *Immunol Rev*, 172, pp. 121-9.
Nishino, M. *et al.* (2018) 'Sarcoid-Like Granulomatosis of the Lung Related to Immune-Checkpoint Inhibitors: Distinct Clinical and Imaging Features of a Unique Immune-Related Adverse Event', *Cancer Immunol Res*, 6(6), pp. 630-635.
Norelli, M. *et al.* (2018) 'Monocyte-derived IL-1 and IL-6 are differentially required for cytokine-release syndrome and neurotoxicity due to CAR T cells', *Nat Med*, 24(6), pp. 739-748.
O'Garra, A. and Berry, M.P.R. (2013) 'The Immune Response in Tuberculosis ', *Annual Review of Immunology*, 31, pp. 475-527.
O'Kane, C.M. *et al.* (2010) 'STAT3, p38 MAPK, and NF-kappaB drive unopposed monocyte-dependent fibroblast MMP-1 secretion in tuberculosis', *Am J Respir Cell Mol Biol*, 43(4), pp. 465-74.
Ogongo, P., Porterfield, J.Z. and Leslie, A. (2019) 'Lung Tissue Resident Memory T-Cells in the Immune Response to Mycobacterium tuberculosis', *Front Immunol*, 10, p. 992.
Ogongo, P. *et al.* (2021) 'Tissue resident-like CD4+ T cells secreting IL-17 control Mycobacteria tuberculosis in the human lung', *J Clin Invest*.
Olive, A.J. and Sassetti, C.M. (2018) 'Tolerating the Unwelcome Guest; How the Host Withstands Persistent Mycobacterium tuberculosis', *Front Immunol*, 9, p. 2094.
Ong, C.W.M. and Friedland, J.S. (2014) 'Tuberculosis, Pulmonary Cavitation, and Matrix Metalloproteinases', *American Journal of Respiratory and Critical Care Medicine*, 190(1), pp. 9-18.

Opie, E.L. and Aronson, J. (1927) 'Tubercle bacilli in latent tuberculosis lesions and in lung tissue without tuberculous lesions', *Archives of Pathology and Laboratory Medicine*, 4(1).
Pagan, A.J. and Ramakrishnan, L. (2015) 'Immunity and Immunopathology in the Tuberculous Granuloma', *Cold Spring Harbor Perspectives in Medicine*, 5(9).
Pagan, A.J. and Ramakrishnan, L. (2018) 'The Formation and Function of Granulomas', *Annu Rev Immunol*, 36, pp. 639-665.
Parasa, V.R. *et al.* (2014) 'Modeling Mycobacterium tuberculosis early granuloma formation in experimental human lung tissue', *Disease Models & Mechanisms*, 7(2), pp. 281-288.
Patel, V.P. *et al.* (1997) 'Molecular and functional characterization of two novel human C-C chemokines as inhibitors of two distinct classes of myeloid progenitors', *J Exp Med*, 185(7), pp. 1163-72.
Peters, W. and Ernst, J.D. (2003) 'Mechanisms of cell recruitment in the immune response to Mycobacterium tuberculosis', *Microbes Infect*, 5(2), pp. 151-8.
Pimentel, H. *et al.* (2017) 'Differential analysis of RNA-seq incorporating quantification uncertainty', *Nat Methods*, 14(7), pp. 687-690.
Polena, H. *et al.* (2016) 'Mycobacterium tuberculosis exploits the formation of new blood vessels for its dissemination', *Sci Rep*, 6, p. 33162.
Pop, M. and Salzberg, S.L. (2008) 'Bioinformatics challenges of new sequencing technology', *Trends in Genetics*, 24(3), pp. 142-149.
Puschmann, T.B. *et al.* (2013) 'Bioactive 3D cell culture system minimizes cellular stress and maintains the in vivo-like morphological complexity of astroglial cells', *Glia*, 61(3), pp. 432-40.
Rainer, J., Gatto, L. and Weichenberger, C.X. (2019) 'ensembldb: an R package to create and use Ensembl-based annotation resources', *Bioinformatics*, 35(17), pp. 3151-3153.
Raudvere, U. *et al.* (2019) 'g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)', *Nucleic Acids Res*, 47(W1), pp. W191-W198.
Ravi, M. *et al.* (2015) '3D cell culture systems: advantages and applications', *J Cell Physiol*, 230(1), pp. 16-26.
Repnik, U., Knezevic, M. and Jeras, M. (2003) 'Simple and cost-effective isolation of monocytes from buffy coats', *J Immunol Methods*, 278(1-2), pp. 283-92.
Ritchie, M.E. *et al.* (2015) 'limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic Acids Res*, 43(7), p. e47.
Robertson, J.G. (2005) 'Mechanistic basis of enzyme-targeted drugs', *Biochemistry*, 44(15), pp. 5561-71.
Rohde, K. *et al.* (2007) 'Mycobacterium tuberculosis and the environment within the phagosome', *Immunol Rev*, 219, pp. 37-54.
Rosen, Y. (2021) 'Pathology of Granulomatous Pulmonary Diseases', *Arch Pathol Lab Med*.
Rual, J.F. *et al.* (2005) 'Towards a proteome-scale map of the human protein-protein interaction network', *Nature*, 437(7062), pp. 1173-8.
Russell, D.G. (2003) 'Phagosomes, fatty acids and tuberculosis', *Nat Cell Biol*, 5(9), pp. 776-8.
Russell, D.G. (2011) 'Mycobacterium tuberculosis and the intimate discourse of a chronic infection', *Immunol Rev*, 240(1), pp. 252-68.
Russell, D.G. *et al.* (2009) 'Foamy macrophages and the progression of the human tuberculosis granuloma', *Nat Immunol*, 10(9), pp. 943-8.
Sahraeian, S.M.E. *et al.* (2017) 'Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis', *Nat Commun*, 8(1), p. 59.
Salgame, P. (2011) 'MMPs in tuberculosis: granuloma creators and tissue destroyers', *J Clin Invest*, 121(5), pp. 1686-8.
Sathaliyawala, T. *et al.* (2013) 'Distribution and compartmentalization of human circulating and tissue-resident memory T cell subsets', *Immunity*, 38(1), pp. 187-97.
Sathyamoorthy, T. *et al.* (2015) 'Membrane Type 1 Matrix Metalloproteinase Regulates Monocyte Migration and Collagen Destruction in Tuberculosis', *J Immunol*, 195(3), pp. 882-91.

Schnute, M.E. *et al.* (2012) 'Modulation of cellular S1P levels with a novel, potent and specific inhibitor of sphingosine kinase-1', *Biochem J*, 444(1), pp. 79-88.
Schwartz, M.A. and Chen, C.S. (2013) 'Cell biology. Deconstructing dimensionality', *Science*, 339(6118), pp. 402-4.
Scriba, T.J. *et al.* (2017) 'Sequential inflammatory processes define human progression from M. tuberculosis infection to tuberculosis disease', *PLoS Pathog*, 13(11), p. e1006687.
Seegar, T.C. *et al.* (2010) 'Tie1-Tie2 interactions mediate functional differences between angiopoietin ligands', *Mol Cell*, 37(5), pp. 643-55.
Sharma, S.K. *et al.* (2005) 'Miliary tuberculosis: new insights into an old disease', *Lancet Infect Dis*, 5(7), pp. 415-30.
Shea, K.M. *et al.* (2014) 'Estimated rate of reactivation of latent tuberculosis infection in the United States, overall and by population subgroup', *Am J Epidemiol*, 179(2), pp. 216-25.
Shen, C.Y. *et al.* (2013) 'Autoantibody prevalence in active tuberculosis: reactive or pathognomonic?', *BMJ Open*, 3(7).
Shigehara, K. *et al.* (2001) 'IL-12 and IL-18 are increased and stimulate IFN-gamma production in sarcoid lungs', *J Immunol*, 166(1), pp. 642-9.
Skoura, E., Zumla, A. and Bomanji, J. (2015) 'Imaging in tuberculosis', *Int J Infect Dis*, 32, pp. 87-93.
Soneson, C., Love, M.I. and Robinson, M.D. (2015) 'Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences', *F1000Res*, 4, p. 1521.
Song, Z. *et al.* (2005) 'Mycobacterial catalase-peroxidase is a tissue antigen and target of the adaptive immune response in systemic sarcoidosis', *J Exp Med*, 201(5), pp. 755-67.
Sorokin, L. (2010) 'The impact of the extracellular matrix on inflammation', *Nat Rev Immunol*, 10(10), pp. 712-23.
Spagnolo, P. *et al.* (2018) 'Pulmonary sarcoidosis', *Lancet Respir Med*, 6(5), pp. 389-402.
Spiegel, S. and Milstien, S. (2003) 'Sphingosine-1-phosphate: an enigmatic signalling lipid', *Nat Rev Mol Cell Biol*, 4(5), pp. 397-407.
Sturgill-Koszycki, S. *et al.* (1994) 'Lack of acidification in Mycobacterium phagosomes produced by exclusion of the vesicular proton-ATPase', *Science*, 263(5147), pp. 678-81.
Subramanian, A. *et al.* (2005) 'Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles', *Proc Natl Acad Sci U S A*, 102(43), pp. 15545-50.
Supek, F. *et al.* (2011) 'REVIGO summarizes and visualizes long lists of gene ontology terms', *PLoS One*, 6(7), p. e21800.
Tezera, L.B. *et al.* (2017) 'Dissection of the host-pathogen interaction in human tuberculosis using a bioengineered 3-dimensional model', *Elife*, 6.
Thomas, G. (2002) 'Furin at the cutting edge: from protein traffic to embryogenesis and disease', *Nat Rev Mol Cell Biol*, 3(10), pp. 753-66.
Thompson, C.R. *et al.* (2005) 'Sphingosine kinase 1 (SK1) is recruited to nascent phagosomes in human macrophages: inhibition of SK1 translocation by Mycobacterium tuberculosis', *J Immunol*, 174(6), pp. 3551-61.
Ulrich, T.A., de Juan Pardo, E.M. and Kumar, S. (2009) 'The mechanical rigidity of the extracellular matrix regulates the structure, motility, and proliferation of glioma cells', *Cancer Res*, 69(10), pp. 4167-74.
Uplekar, M. *et al.* (2015) 'WHO's new end TB strategy', *Lancet*, 385(9979), pp. 1799-1801.
Uusi-Makela, M. and Ramet, M. (2018) 'Hijacking Host Angiogenesis to Drive Mycobacterial Growth', *Cell Host Microbe*, 24(4), pp. 465-466.
Valeyre, D. *et al.* (2014) 'Sarcoidosis', *Lancet*, 383(9923), pp. 1155-1167.
van Winden, V.J.C. *et al.* (2019) 'Protease domain and transmembrane domain of the type VII secretion mycosin protease determine system-specific functioning in mycobacteria', *J Biol Chem*, 294(13), pp. 4806-4814.
Vergne, I. *et al.* (2005) 'Mechanism of phagolysosome biogenesis block by viable Mycobacterium tuberculosis', *Proc Natl Acad Sci U S A*, 102(11), pp. 4033-8.

Vilaplana, C. and Cardona, P.-J. (2014) 'The lack of a big picture in tuberculosis: the clinical point of view, the problems of experimental modeling and immunomodulation. The factors we should consider when designing novel treatment strategies', *Frontiers in Microbiology*, 5(55), pp. 1-8.
Vukmirovic, M. *et al.* (2017) 'Identification and validation of differentially expressed transcripts by RNA-sequencing of formalin-fixed, paraffin-embedded (FFPE) lung tissue from patients with Idiopathic Pulmonary Fibrosis', *BMC Pulm Med*, 17(1), p. 15.
Vynnycky, E. and Fine, P.E. (2000) 'Lifetime risks, incubation period, and serial interval of tuberculosis', *Am J Epidemiol*, 152(3), pp. 247-63.
Walker, N.F. and Elkington, P.T. (2012) 'Doxycycline and HIV Infection Suppress Tuberculosis-induced Matrix Metalloproteinases ', *American Journal of Respiratory and Critical Care medicine*, 185(9), pp. 989-997.
Wenzel, C. and Steigemann, P. (2014) '3D high-content screening for the identification of compounds that target cells in dormant tumor spheroid regions', *Experimental Cell Research*, 323(1), pp. 131-143.
WHO (2020) 'World Health Organization. Global Tuberculosis Report 2020'.
Workman, V.L. and Jayasinghe, S.N. (2014) 'Controlled Generation of Microspheres Incorporating Extracellular Matrix Fibrils for Three-DimensionalCell Culture', *Advanced Functional Materials*, 24(8), pp. 2648-2657.
Yang, Q. *et al.* (2020) 'Identification of eight-protein biosignature for diagnosis of tuberculosis', *Thorax*, 75(7), pp. 576-583.
Youm, J.K. *et al.* (2008) 'K6PC-5, a sphingosine kinase activator, induces anti-aging effects in intrinsically aged skin through intracellular Ca2+ signaling', *J Dermatol Sci*, 51(2), pp. 89-102.
Young, D. (2009) 'Animal models of tuberculosis', *Eur J Immunol*, 39(8), pp. 2011-4.
Yu, G. and He, Q.Y. (2016) 'ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization', *Mol Biosyst*, 12(2), pp. 477-9.
Ziegenhagen, M.W. *et al.* (2002) 'Exaggerated TNFalpha release of alveolar macrophages in corticosteroid resistant sarcoidosis', *Sarcoidosis Vasc Diffuse Lung Dis*, 19(3), pp. 185-90.
Zissel, G. and Muller-Quernheim, J. (2016) 'Specific antigen(s) in sarcoidosis: a link to autoimmunity?', *European Respiratory Journal*, 47(3), pp. 707-709.