Thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2011

Measuring exposure to non-tuberculous mycobacteria:

defining and testing novel antigens

Anna Checkley

Lincoln College

Acknowledgements

There are so many people who have helped me with this project in different ways, and I am grateful to them all. My supervisor Helen McShane provided constant support- working with you has been great fun, and I have learnt lots. David Wyllie taught me a huge amount about bioinformatics, and I could not possibly have done this project without his help.

I am grateful to the many other people in the Jenner Institute who have also helped in different ways. Iman Satti helped with cultured ELISpot techniques, provided useful discussions, and proof read sections of my thesis whilst on holiday in Sudan. Helen Fletcher and Ansar Pathan provided very helpful advice and discussion, and Tara Berthoud shared both her laboratory experience and constant cheerfulness. Tess Lambe, Alex Spencer, Hazel Poyntz and Simone de Cassan kindly helped with mouse immunisations and culling; additional thanks go to Simone for her meticulous proof reading. Thanks also to Steph Harris for help with shipping samples to South Africa, to Drew Worth for help using the flow cytometer and Julie Furze for the smooth running of the lab. Finally I would like to thank the whole TB group at the Jenner Institute for helpful discussions and great pub lunches and dinner parties.

Everyone at the South African TB Vaccine Initiative (SATVI), Cape Town made my time in South Africa hugely enjoyable and rewarding. Thanks to my co-supervisor, Willem Hanekom, for making me so welcome, and for useful discussions. Thanks to Tom Scriba, for his enthusiasm, expertise and helpfulness, to Ben Kagina for friendship and constant amusement, and finally to all the others who helped in different ways; Cheryl Day, Brian Abel, Nazma Mansoor, Noella Moshi, Andreia Soares, Alana Keyser, William Kwong Chuong, Erica Smit, Lesedi Lerumo, Yolande Brown, Saroj Pillay and Keshia de Long.

Tanya Golubchik from the Department of Statistics, Oxford University, kindly performed a genome assembly on recalcitrant sequencing data from the *M. fortuitum* sequencing.

All blood and cord blood donors also deserve my thanks and gratitude, as do Oxford Immunotec for the kind donation of PBMC samples. Thanks to Henry Bettinson from the Chest Clinic, Oxford, for his help in patient recruitment and to both Ian Bowler, Dept Microbiology, Oxford and Malcolm Yates, TB Reference Laboratory, for the provision of microbiological samples. Carmella Elan-Gaston's efficiency and helpfulness made Lincoln College a great place to be based.

And finally the largest thank you goes to my parents, Marilyn and Stuart Checkley, and especially my husband, James Barr, for the constant love, support and encouragement they have provided throughout.

This work was funded by the Medical Research Council, UK (grant number G0701309), the Lancet International Fellowship, the Sir Rataji Dalal Fellowship and the Sir Herbert Hunt Travelling Scholarship.

Abstract

The tuberculosis (TB) vaccine bacillus bacillus Calmette-Guerin (BCG) has low efficacy against pulmonary TB in the tropics, where most TB disease occurs, and where an effective TB vaccine is most needed. It is possible that new TB vaccines in development may be similarly affected, in particular those based on the current BCG vaccine. There are several hypotheses to explain this phenomenon, one of which is that exposure to related non-tuberculous mycobacteria (NTM) reduces the efficacy of the BCG vaccine. Some animal data suggest that exposure to NTM results in reduced efficacy of BCG, although results from animal experiments are variable. IFN-γ responses to purified protein derivative (PPD) in humans also suggest that individuals living in regions where BCG has a low efficacy make robust immune responses to PPD suggestive of NTM exposure.

The definition of antigens which are present in common NTM but absent from *Mycobacterium tubercuosis (M. tuberculosis)* and BCG would allow more detailed testing of this hypothesis. In this thesis, the definition and testing of such antigens are described. First, a bioinformatics pathway was used to define proteins which are present in common NTM and absent from the *M. tuberculosis* family (*M. tuberculosis,* BCG and other related mycobacteria). Following this, overlapping peptides from proteins selected during this process were tested in a mouse model of NTM infection, in humans from the UK and US who had NTM isolated from sputum samples, in healthy South Africans with known positive and negative responses to QuantiFERON®-TB Gold In-Tube test (Cellestis) and in cord blood samples. Low level but significant gamma interferon (IFN-γ) responses were detected by *ex vivo* ELISpot in the healthy South Africans, and to the same peptide pools in individuals from the UK and US using a proliferation assay. No responses were detected in the cord blood

cells, or in the mouse model. The implications of these findings, and the potential for further development of a test of NTM exposure, are discussed.

Chapters

Abbreviations

AAM	Alternately activated macrophage
BCG	Bacillus Calmette-Guerin
BLAST	Basic Local Alignment Search Tool
BLASTN	Nucleotide BLAST
BLASTP	Protein BLAST
bp	Base pairs
CAM	Classically activated macrophage
CFP-10	Culture filtrate protein-10
CFU	Colony forming unit
Contig	Contiguous DNA sequence (assembled from WGS sequence information)
CXR	Chest X-ray
EBI	European Bioinformatics Institute
ELISA	Enzyme-linked immunosorbent assay
ELISpot	Enzyme-linked immunosorbent spot
EMBL	European Molecular Biology Laboratory
ESAT-6	Early secretory antigenic target-6
FCS	Foetal calf serum
H	Hour
hsp-65	Heat-shock protein 65

IFNγ	Gamma interferon
IGRA	Interferon gamma release assay
IL-2	Interleukin 2
IL-7	Interleukin 7
i.v.	Intravenous
M.	Mycobacterium
MDR-TB	Multi drug-resistant tuberculosis
MEM	Modified Eagle's Medium
Mins	Minutes
NCBI	National Centre for Biotechnology Information
NK	Natural killer
NTM	Non-tuberculous mycobacteria
OD	Optical density
ORF	Open reading frame
PBMC	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PPD	Purified protein derivative
PPD-A	Purified protein derivative from *M. avium*
PPD-T	Purified protein derivative from *M. tuberculosis*
PHA	Phytohaemagglutinin
PMA	Phorbol 12-myristate 13-acetate
PBS	Phosphate-buffered saline

PPD	Purified protein derivative
R0	RPMI medium, 2mM L-glutamine and 1% penicillin/streptomycin
R10	R0 plus 10% heat-inactivated FCS
RBC	Red blood cells
RCT	Randomised controlled trial
RPM	Revolutions per minute
RPMI/AB/glutamine	RPMI + 12.5% human AB serum + 2mM L L-glutamine
rpoB	RNA polymerase B
RT	Room temperature
SATVI	South African TB Vaccine Initiative
Sec	Seconds
SFC	Spot forming cells
SSI	Statens Serum Institut
SNP	Single nucleotide polymorphism
TB	Tuberculosis
TBLASTN	Translated nucleotide BLAST
TDR-TB	Totally drug-resistant tuberculosis
TNF-α	Tumour necrosis factor alpha
Treg	Regulatory T cells
TST	Tuberculin skin test
US	United States
WGS	Whole genome shotgun

XDR-TB	Extremely drug-resistant tuberculosis

Table of Contents

Table of figures

Table of tables

1 Introduction

1.1 Tuberculosis

1.1.1 Epidemiology

Tuberculosis (TB) remains a major threat to global public health, with an estimated 9.4 million new cases and 1.7 million deaths occurring worldwide in 2009 [1]. Substantial case finding and treatment efforts [2] over the past decade have been associated with a small decrease in TB prevalence, but a global population increase means that more cases were still reported in 2009 than in previous years [1].

TB incidence rates are particularly high in countries with high human immunodeficiency virus (HIV) infection prevalence; Africa alone accounts for 4 out of 5 HIV-associated tuberculosis cases worldwide. Increasing rates of drug resistance add to the difficulty of treatment, particularly in resource-poor settings, where second line drug combinations are often unaffordable [3]. In 2008, approximately 390,000-510,000 cases of tuberculosis were caused by multi-drug resistant strains of *Mycobacterium tuberculosis* (*M. tuberculosis*) (MDR-TB), those resistant to both rifampicin and isoniazid, about 3.6% of the global burden of tuberculosis [3]. Among MDR-TB cases there was an estimated case fatality of 26% (among HIV negative individuals). Extremely drug resistant *M. tuberculosis* (XDR-TB) strains, resistant to isoniazid, rifampicin, fluoroquinolones and at least one injectable agent, now make up 5.4% of MDR-TB cases. Such strains are associated with even higher mortality, and the first cases of totally drug resistant tuberculosis (TDR-TB), resistant to all first and second line antibiotics, are now being reported [4].

1.1.2 Clinical disease

On exposure to a *M. tuberculosis* aerosol from an infectious patient, an HIV negative individual has approximately a 10% lifetime risk of developing active clinical tuberculosis [5], with roughly half that risk in the first few years after infection. Most infected individuals develop latent tuberculosis, in which tests of exposure such as the tuberculin skin test (TST) and interferon gamma release assays (IGRAs) are positive, but the individual is free of symptoms (see Section 1.1.4 for discussion on TST and IGRAs).

On exposure to *M. tuberculosis,* some individuals are infected, progressing rapidly to active disease (particularly in the case of children), some become latently infected, and some clear the infection immediately, with no immunological evidence of *M. tuberculosis* infection [6]. Latent tuberculosis probably represents a spectrum of disease, ranging from individuals who have completely cleared the disease to those harbouring actively replicating mycobacteria [7]. This is reflected in very different pathological appearances in latent infection, from sterile fibrosis through caseous lesions to cavitating lesions containing huge numbers of replicating mycobacteria [7]. These very different disease states probably result from differences in immunological response to exposure to *M. tuberculosis,* as well as to other factors, including bacterial strain (see Section 1.2.2.3).

Active TB disease in adults usually occurs following reactivation of latent bacilli, and in immunocompetent adults it most often takes form of pulmonary tuberculosis, with presenting symptoms including fever, cough, sputum production with haemoptysis, and weight loss. Treatment of drug-sensitive pulmonary disease with 6 months of combination antibiotic therapy results in cure in 85% of individuals worldwide [8].

Tuberculosis disease is more severe in HIV co-infected individuals, who have increased susceptibility to infection [9], higher rates of re-activation from latency and more rapid progression of active disease [10]. This group of individuals are also often diagnosed late in the course of infection, as atypical presentations of infection are much more common, with more cases of smear-negative disease (no acid-fast bacilli seen on sputum microscopy) and either a normal Chest X-ray (CXR) or CXR changes which are not typical of TB disease in immunocompetent individuals.

Children are another group who are particularly susceptible to *M. tuberculosis* infection, and in whom diagnosis is also difficult [11]. They are more likely to develop active rather than latent disease following infection. Disseminated TB, consisting of miliary TB and TB meningitis, with a high mortality rate, is much more common, and pulmonary TB is more difficult to diagnose as sputum and gastric washing are more often smear negative, sputum is much more difficult to collect (hence the use of gastric washings), and the CXR is often not diagnostic.

1.1.3 Tuberculin skin test and interferon gamma release assays

The TST consists of an intradermal injection of purified protein derivative (PPD, a mixture of multiple proteins produced by *M. tuberculosis* in culture), which is read 48-72 hours later by measuring the extent of induration. A positive response, usually taken as greater than 10mm, is indicative of exposure to *M. tuberculosis* exposure; smaller responses may also be associated with exposure to other mycobacteria, such as BCG or non-tuberculous mycobacteria (NTM). The sensitivity of TST in detecting *M. tuberculosis* infection has been estimated as slightly greater than 70%, with a specificity of 66% [12]. A positive TST (with induration greater than 5mm) was shown to be associated with a risk of developing TB

disease of 1.6 per 100 person years among HIV positive individuals in Switzerland, a low TB incidence country [13].

More recently, two tests have been developed which measure T cell gamma interferon (IFN-γ) production in response to two antigens which are present in *M. tuberculosis,* and in very few other species of mycobacteria. TSPOT-TB (Oxford Immunotec) uses an enzyme-linked immunosorbent spot (ELISpot) assay to enumerate IFN-γ-producing, antigen-specific T cells, and QuantiFERON®-TB Gold In-Tube test (Cellestis) uses an enzyme-linked immunosorbent assay (ELISA) to quantify T cell IFN-γ production. Both have been used extensively in screening TB contacts for immunological evidence of TB exposure, and both have been shown to have a very high sensitivity and specificity in this setting, in particular not being confounded by BCG vaccination, as occurs with the TST [14]. They are also used in screening individuals for evidence of latent TB infection prior to commencing a wide range of immunosuppressive therapies.

Interpretation of results from IGRA is more complex in high TB prevalence countries, where latent infection is so widespread (and where evidence of latency is not usually an indication for treatment). A study comparing the sensitivity and specificity of IGRA in diagnosing latent TB infection in Cape Town, South Africa, found 76% sensitivity and 42% specificity for QuantiFERON®-TB Gold In-Tube compared to 84% sensitivity and 47% specificity for TSPOT.TB [15]. Interestingly, a recent South African study showed that QuantiFERON®-TB Gold In-Tube test (Cellestis) was not superior to TST in diagnosing latent TB infection among adolescents [16].

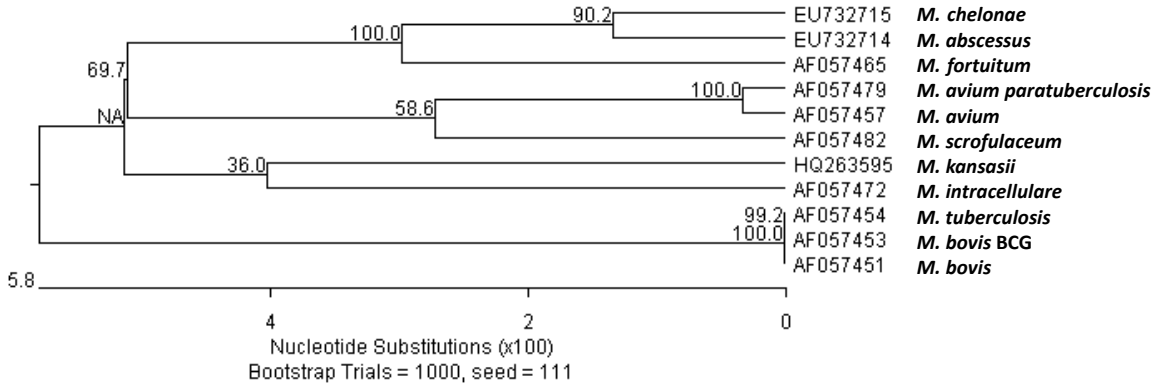
1.1.4 Mycobacteria

Mycobacteria are acid-fast, aerobic, Gram-positive rods which are resistant to decolourisation with acid after staining. There are many different species of mycobacteria, and they can be divided into the *Mycobacterium tuberculosis* (*M. tuberculosis)* complex, and non-tuberculous mycobacteria (NTM). Members of the *M. tuberculosis* complex cause TB disease in humans, and include *M. tuberculosis, M. africanum, M. bovis* and also BCG, derived from *M. bovis* (see Section 1.2). NTM are a highly variable group of organisms, and can be broadly divided into fast and slow growing mycobacteria. Fast growing organisms are visible on an agar plate after one to two weeks of incubation, whereas slow growing organisms are not visible until four to six weeks. Analysis of genetic polymorphisms, for example of the gene RNA polymerase B (rpoB), confirms that this microbiological division also reflects genetic relatedness, as shown in Figure 1.1 [17].

1.1.5 Tuberculosis immunology

On entry into the lungs, a *M. tuberculosis* bacillus is typically phagocytosed by an airway antigen presenting cell, usually an alveolar macrophage or dendritic cell. Infected dendritic cells then migrate to the draining lymph node, where T cell priming takes place 12-24 days after infection [18]. Within the phagosome of the infected macrophage, *M. tuberculosis* arrests the usual process of phagolysosomal fusion [19], avoiding proteolytic degradation. Mycobacterial replication then occurs, unless the macrophage is activated by an effector T cell. Infected alveolar macrophages are central to the formation of granulomas; multi-cellular structures which allow the containment, but also the persistence, of *M. tuberculosis*.

Figure 1-1: Phylogenetic tree of mycobacteria



Phylogenetic tree of mycobacterial species of interest. The tree was constructed using Megalign (DNAstar LazerGene) following Clustal W alignment, based on differences in the sequence of the RPOB gene (RPOB gene sequences downloaded from the NCBI Nucleotide database).

1.1.5.1 The granuloma

A granuloma is formed following the sequential recruitment of different cell types, resulting in a spherical structure with infected central macrophages at its core, lipid-filled 'foamy' macrophages surrounding them, and lymphoyctes (T, B and natural killer (NK) cells) and neutrophils on the outside [18]. Granuloma structure is highly variable, containing varying proportions of pro-inflammatory 'classically activated macrophages' (CAM) and anti-inflammatory 'alternately activated macrophages' (AAM), as well as varying proportions of mycobacteria, neutrophils and regulatory T cells (Tregs) [18]. Important cytokines include the following: interleukin-12 (IL-12) and tumour necrosis factor alpha (TNF-α) (Th1 cytokines produced by CAM), IL-10, transforming growth factor beta (TGF-β) and IL-6 (Th2 cytokines produced by AAM) gamma interferon (IFN-γ), IL-2 and TNFα (produced by Th1 T cells), IL-17 (produced by Th17 T cells) and IL-10 and TGF-β (produced by Treg cells). The proportions of the various cell types and cytokines influence the composition of the granuloma and its success in bacterial containment.

1.1.5.2 Cytokines

There are multiple different cytokines involved in the immune response against *M. tuberculosis*, and their relative importance can be deduced using both animal model experiments, and in humans from observing the effects of the loss of specific immune cell subsets or cytokines resulting from infection, gene deletion or pharmacological interventions. A Th1 cellular immune response, consisting of CD4 positive T cells, activated macrophages and cytokines including IFN-γ and IL-12 is critical for anti-mycobacterial immunity.

1.1.5.2.1 IFN-γ

IFN-γ is critically important in the anti-mycobacterial immune response [6]. It is produced by CD4 and CD8 positive T cells, and it induces macrophage activation and increased expression of MHC class I and II proteins on antigen presenting cells. IFN-γ knock-out mice are highly susceptible to *M. tuberculosis* infection, and have low levels of macrophage activation [20]. Individuals with mutations in the IFN-γ receptor or in the downstream STAT 1 signalling pathway are at dramatically increased risk of infection both by *M. tuberculosis* and by BCG and NTM [21].

1.1.5.2.2 Interleukin-12

The cytokine interleukin-12 (IL-12) is also important, as demonstrated by an increased susceptibility of children with interleukin 12 receptor β1 (IL-12Rβ1) deficiency to severe *M. tuberculosis* infection [22, 23]. In these patients, levels of IFN-γ production are reduced. IL-12 is produced by activated macrophages and dendritic cells on phagocytosis of *M. tuberculosis*, and it promotes the development of a Th1 polarised immune response, activating natural killer (NK) cells, and inducing CD4 T cell differentiation to Th1-type cells

[24]. IL-12 knockout mice have increased susceptibility to *M. tuberculosis* infection, and administration of endogenous IL-12 to mice can reduce the *M. tuberculosis* load [25, 26].

1.1.5.2.3 TNF-α

TNF-α is also essential to protection from *M. tuberculosis* disease, as demonstrated by a greatly increased susceptibility to *M. tuberculosi*s reactivation in individuals taking TNF antagonists for the treatment of chronic inflammatory diseases, such as rheumatoid arthritis, inflammatory bowel disease and psoriasis. This susceptibility is so great that preventative therapy with anti-tuberculous medication is now recommended in individuals with evidence of mycobacterial exposure prior to starting such therapies [27]. TNF-α is secreted by macrophages, dendritic cells and T cells, and is important in macrophage activation and granuloma formation, although it has also been implicated in the excessive destructive inflammation often seen in *M. tuberculosis* infection [24].

1.1.5.2.4 Other cytokines

Other cytokines which may be important in antituberculous immunity include IL-17, produced by Th17 cells, and the regulatory cytokines IL-10 and TGF-β, produced by regulatory T cells (Tregs) (expanded in Section 1.1.5.3).

1.1.5.3 Cell subsets

1.1.5.3.1 CD4 positive T cells

CD4 positive T cells are central to protection against infection and disease, as evidenced by the dramatically increased susceptibility to tuberculosis seen in humans in states of CD4 T cell depletion such as HIV infection [28, 29], and in genetic disorders affecting the IFN-γ signalling pathway [21]. The increased susceptibility of HIV-infected individuals to *M.*

tuberculosis infection and also to disease reactivation illustrates the importance of CD4 positive T cells in protection against both of these events [10]. It is the ability of these cells to produce IFN-γ which is critical to their protective function (see Section 1.1.5.2), although they have also been shown to exhibit cytotoxic activity [30, 31].

1.1.5.3.2 CD8 positive T cells

Although CD4 positive T cells are essential for anti-tuberculous immunity, they are probably not sufficient, and other T cell subsets are also important. CD8 positive T cells probably have two main functions in anti-tuberculous immunity; the production of Th1 cytokines such as IFN-γ, and the mediation of cytotoxicity, via the production of molecules which are directly toxic to mycobacteria, such as perforin and granulysin, and via Fas-Fas ligand interactions [6]. It is likely that CD8 positive T cells polarise to perform one of these two functions [32, 33]. CD8 positive T cells are important in antituberculous immunity; mice lacking functional CD8 positive T cells have increased susceptibility to *M. tuberculosis* infection [34]. They may also have a role in mediating immune surveillance in latent TB infection [35].

1.1.5.3.3 Th17 cells

Other T cell subsets have also been shown to mediate protective immunity in TB. Th17 cells are characterised by the production of interleukin 17 (IL-17). They are mainly known for their role in mediating autoimmune pathology, but there is evidence from animal models that they have a role in the immune response to *M. tuberculosis* infection. In mice, Th17 cells have been shown to mediate the recruitment of Th1 cells to the lung on *M. tuberculosis* challenge, and to confer increased protection from disease [36]. They have been shown to have a role in granuloma formation in the lung of BCG-infected mice [37], but have also been associated with increased immunopathology in mice [38]. In cattle

vaccinated using a prime-boost regimen, IL-17 production correlated with protection from *M. bovis* challenge [39]. The role of IL-17 cells in TB in humans is not clear, however, although high levels of Th17 cells were demonstrated in healthy mycobacteria- exposed adults in South Africa [40], and low levels have been shown in healthy volunteers vaccinated with the vaccine MVA85A [41].

1.1.5.3.4 Regulatory T cells

Regulatory T cells (Tregs) produce IL-10 and TGF-β, and inhibit cytokine production and proliferation by other cell types, for example CD4 positive T cells [42]. They have been shown to be important in regulating excessive inflammatory damage in *M. tuberculosis*-infected macaques [43], although excessive Treg activity may impair a protective immune response [44].

1.1.5.3.5 Polyfunctional T cells

There is evidence in some infections, for example *Leishmania major* infection in mice, that the presence of vaccine-induced polyfunctional T cells (those simultaneously expressing multiple cytokines, such as IFN-γ, IL-2 and TNF-α) are associated with protection from disease [45]. There is no evidence of protection from TB disease in humans [46] although in HIV infection polyfunctional T cells have been shown to be associated with long-term non-progression of disease [47, 48]. Polyfunctional T cells have been shown to be induced following vaccination with the new TB vaccines MVA85A [49, 50] and AERAS-402 [51], and to be present in higher numbers in active than in latent TB infection [52].

1.1.5.4 The innate immune response

It is likely that some form of innate immune response, possibly mediated by cells already present in the lung alveoli (for example, dendritic cells, alveolar macrophages or alveolar

endothelial cells), allows some individuals to clear *M. tuberculosis* without ever developing immunological evidence of infection, although there is no clear evidence for this. This is likely to be based on a rapidly activated innate immune response.

1.1.5.4.1 Macrophages

Macrophages are critical in the host defence against *M. tuberculosis* infection; they are the main cellular target of infection of *M. tuberculosis,* they are central to the structure of the granuloma (Section 1.1.5.1) and they produce the important Th1 and pro-inflammatory cytokines IL-12 and TNF-α (Section 1.1.5.2). *M. tuberculosis* bacilli can multiply within a resting macrophage, but are killed via the production of nitric oxide and related reactive nitrogen intermediates, and by phagolysosomal fusion, when the macrophage is activated [18, 24].

1.1.5.4.2 Gamma-delta T cells

Gamma-delta (γδ) T cells are T cells with both innate and adaptive properties, as they respond to antigens in the absence of MHC presentation, but undergo limited receptor gene rearrangement. They are thought to be important in the early immune response against *M. tuberculosis* infection, with a role in granuloma formation [53], and they have been shown to produce IFN-γ [54] and IL-17 [55].

1.1.5.4.3 Other cell types

Neutrophils form a part of the granuloma, and there is emerging evidence that excessive neutrophils may contribute towards active TB disease in humans [21, 56]. There is also emerging evidence that B cells may play a previously unrecognised role in the immune response to *M. tuberculosis* [57].

1.1.5.5 Correlates of protection

An important challenge facing those involved in developing new TB vaccines is in determining which combination of immunological markers represents a correlate of vaccine-induced protection from TB disease. The availability of a biomarker which could inform us which individuals were protected by a TB vaccine and which remained at risk of TB disease, would dramatically assist efforts to develop a new TB vaccine, as it could avoid the need for lengthy and expensive efficacy trials.

Currently, the most frequently used read outs of vaccine immunogenicity measure those aspects of the anti-tuberculous immune response which we know are essential for protection, in particular CD4 cell IFN-γ production. However, we know that IFN-γ is not the only cytokine involved in a protective immune response to *M. tuberculosis* infection. Other than T cell IFN-γ production, a number of other cytokines and cell subsets have been identified as potential correlates of vaccine-induced protection in animal models. Until we have a successful TB vaccine, however, they cannot be validated in humans.

The detection of Th17 cells in cattle was found to correlate with protection against *M. bovis* challenge following a prime-boost vaccination regimen [39] (Section 1.1.5.3). In same study, so did the detection of T cells which were defined as memory T cells (detected by ELISpot following 14 days of culture, but not further defined on the basis of expression of cell surface marker expression). It has also been suggested that polyfunctional T cells might be a marker of protective immunity, and there is evidence in mice that they may correlate with protection from *M. tuberculosis* [45], but no evidence in humans (Section 1.1.5.3). Kagina *et al* showed induction of greater levels of polyfunctional T cells in infants in whom BCG vaccination was delayed until 10 weeks of age, compared to those vaccinated at birth,

together with higher levels of antigen-specific CD4 positive T cells [58]. However, a large, prospective South African study at the same centre looking for immunological correlates of risk of TB disease following infant BCG vaccination failed to find a correlation between levels of polyfunctional T cells and protection from TB disease [59]. Interestingly, this study failed to describe any correlates of vaccine-induced protection after immunological testing [59].

Until we have a vaccine which has been proven to be efficacious in randomised controlled trials in multiple populations, we will not be able to define which combination of these or other factors represents a correlate of vaccine-induced protection against TB disease (or infection, and these may be different). *M. tuberculosis* is a complex pathogen, requiring a complex co-ordination of different aspects of the immune response, and it is very unlikely that any single immunological read-out will eventually provide a correlate of vaccine-induced protection. It is much more likely that multiple different read-outs will be combined to provide an estimate of risk of disease and vaccine-induced protection, and it may well be that these combinations are specific to stage of disease (latent or active), to specific vaccines and to different populations. Interestingly, the prospective South African study described above also studied the whole blood transcriptional response to BCG vaccination, and preliminary results suggest that a whole blood transcriptional signature can reliably predict the development of TB in BCG-vaccinated infants, but that within this population there were two very different groups, in whom two distinct signatures predicted TB infection (personal communication, H. Fletcher, Oxford University, W Hanekom, SATVI, University of Cape Town).

In the absence of correlates of protection, *in vitro* models have been developed which compare rates of killing of *M. tuberculosis*-[60] and BCG-infected [61] cells in whole blood

from vaccinated and unvaccinated volunteers, but there are technical challenges associated with using these complex assays reproducibly in multiple laboratories.

1.2 BCG

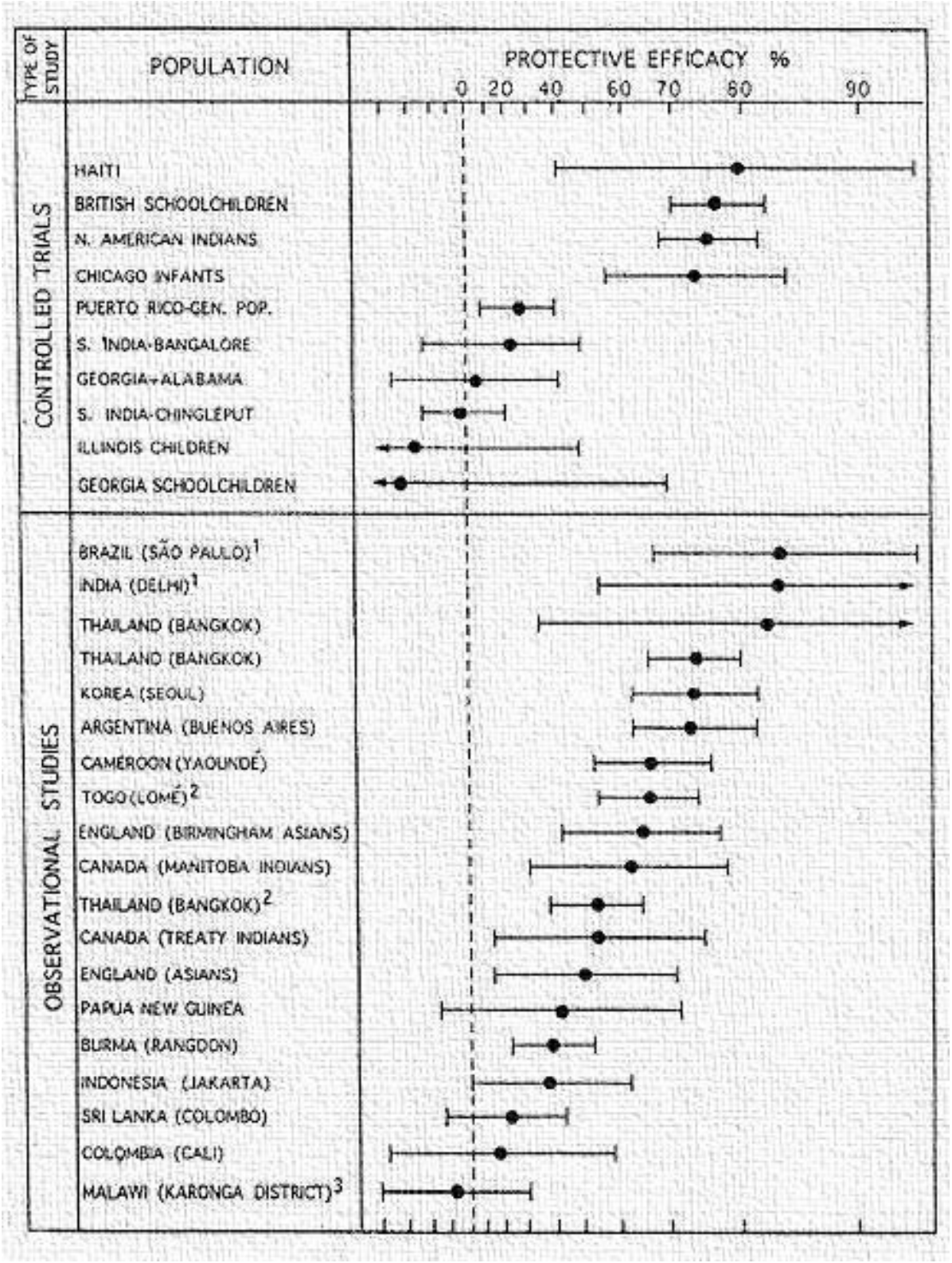
BCG is currently the only licensed tuberculosis vaccine. It was derived from an isolate of *M. bovis* at the Institut Pasteur, passaged multiple times from 1909-1921, and first given to a human in 1921. All current sub-strains of BCG are derived from this original strain [62]. It was introduced into the WHO Expanded Programme on Immunisation (EPI) in 1974, and is currently one of the most widely used vaccines, with coverage at birth in endemic countries reaching more than 88% in 2009 [62].

1.2.1 Variable efficacy of BCG

The efficacy of BCG is highly variable, with protection against pulmonary disease ranging from 80% in the UK [63] to 0% in Malawi [64]; 41% of this variability has been attributed to the latitude at which the study was conducted [65] (Figure 1.2). There are significant differences between its efficacy in preventing tuberculosis in adults and children. It has also been shown to have an efficacy of at least 50% against leprosy in a controlled trial carried out in northern Malawi [64].

European trials between the 1920s and 1940s established BCG's efficacy in Europe [66], and it was adopted widely after the Second World War, both in Europe and worldwide. A major randomised controlled trial (RCT) in the UK described an efficacy of 80% [67], and another in north American Indians showed an efficacy of 82%, with an efficacy of 55% when the same population was followed up 60 years later [68].

Figure 1-2: BCG efficacy varies with latitude



The protective efficacy of BCG, summarised as part of a meta-analysis of controlled trials and observational studies. From: Colditz, G.A., et al., *Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature.* 1994 [65].

Studies performed in the southern United States (US) in Georgia, Alabama and Puerto Rica, however, failed to demonstrate efficacy [69], so routine BCG vaccination was never adopted there. In response to the different results from the various BCG trials (and a suspicion that the differences may be partly explained by exposure to non-tuberculous mycobacteria (NTM)), the British MRC undertook a large trial in Chingleput, in southern India. The efficacy of BCG in this trial was shown to be 0% [70].

Despite its lack of efficacy in preventing adult pulmonary TB in tropical areas (the most common form of TB in this age group) BCG has an efficacy of approximately 73% in preventing tuberculosis meningitis, and of 79% in preventing disseminated tuberculosis in children under the age of 5 years (who account for 80% cases of disseminated tuberculosis) [71]. These forms of disease are very different to adult pulmonary tuberculosis, as they result from primary infection rather than reactivation, and occur as a result of haematogenous dissemination. It is not known whether the efficacy of BCG in this age group is due to its a greater effect on disease caused by haematogenous dissemination, on a waning of its effect with age [72], or to another mechanism. Interestingly, a Saudi Arabian study described 82% protection in individuals aged 5-14 years old, 67% in those aged 15-24 years and 20% in those aged 25-34 years [73].

1.2.2 BCG vaccination recommendations

The lack of effect of BCG on pulmonary TB means it has very little effect on disease transmission in tropical areas, but its effects on childhood disseminated TB have led the WHO to recommend vaccination as soon as possible after birth in high TB endemnicity countries. In view of the high risk of the development of disseminated BCG infection following vaccination of HIV-infected infants [74, 75], current WHO recommendations (in

the absence of availability of expensive tests such as polymerase chain reaction (PCR)) are to withhold vaccination in infants known to be HIV positive or who show signs of infection [76]. If PCR to detect neonatal HIV infection is available, HIV should be actively excluded in infants born to HIV positive mothers prior to administering BCG vaccine, but this is not a widely available test in regions of high HIV prevalence.

Many industrialised countries do not recommend routine BCG vaccination (e.g. the United States, which has never recommended it and the UK and Norway, which have recently changed their policies). In these countries, groups at higher risk of TB infection are vaccinated at birth [62]. Some industrialised countries have continued to recommend vaccinating in early adolescence, since this is a time of low tuberculosis incidence, just before the increase in incidence (and, more importantly, transmission) which is seen in adulthood. A small number of countries recommend BCG revaccination (or even several repeated vaccinations), although there is little evidence to support this recommendation [77-79]. Worldwide vaccination practices are summarised in the BCG World Atlas [80].

1.2.3 Reasons for the variable efficacy of BCG

There are many theories as to why BCG works less well in the tropics than in temperate regions. These include host factors such as genetics, malnutrition or helminth co-infection, vaccine factors such as BCG strain and the reliability of the cold chain, and microbial factors such as the virulence of local strains of *M. tuberculosis* and exposure to non-tuberculous mycobacteria (NTM) [81]. Differences in trial methodologies may also contribute [82].

1.2.3.1 BCG strain

There are several different strains of BCG currently in use, which have been produced as a result of propagation in different culture conditions (including time in culture). Freeze-drying conditions add further variation within and between strains, with large variations in the number of culturable particles retrievable per dose [83]. BCG vaccine quality control has been under the control of national Regulatory Authorities since the Danish Statens Serum Institut (SSI) relinquished the role in 1997, and as a result the level of quality control applied is highly variable between different countries. In order to prevent further deviation from the original strains, the WHO has kept lyophilised seed lots of the different vaccine strains since 1956. A number of vaccine strains are available, but the French Pasteur strain 1173 P2, the Danish strain 1331, the Glaxo-Evans strain 1077 and the Tokyo strain 172 account for about 90% of BCG vaccinations worldwide [62].

The ongoing passage of BCG has led to the progressive loss of various virulence determinants, starting with the RD1 region, which is absent from all strains of BCG. It is possible that as BCG strains undergo further passage they become less virulent; the RD2 region was deleted in about 1927, so earlier strains of BCG (such as Tokyo 172 and Pasteur 1173 P2) contain RD2, whereas newer strains such as Danish 1331 and Glaxo-Evans 1077, do not. Earlier strains have been shown to be more immunogenic in the mouse model compared to later strains [84, 85]. However, studies comparing the efficacy of different sub-strains of the vaccine in humans have shown variable results, with a trial in Hong Kong concluding that Pasteur vaccine provided greater protection over Glaxo [83], and case control studies in Indonesia and Columbia showing a decline in vaccine efficacy on switching from Tokyo 172 to Pasteur 1173 P2 (Indonesia) and from Glaxo to Danish 1331 (Columbia)

[69]. Notably, the freeze-dried Glaxo vaccine was used in two major trials; in the UK where an efficacy of 80% was demonstrated [67] and in Malawi where the efficacy was 0% [64, 86]. Also, the BMRC trial compared the very different vole bacillus vaccine based on *M. microti* with the Copenhagen sub-strain of BCG, and found similar levels of efficacy [63, 64].

When South Africa changed from using subcutaneously administered Tokyo 172 to intradermal Danish 1331, a comparison was undertaken of the number of cases of tuberculosis in infants under the age of two years in the two years immediately before and the two years immediately after the change. There was a suggestion that intradermal Danish 1331 may prevent disseminated tuberculosis more effectively (relative risk (RR) 0.54, 95% confidence interval (CI) 0.4-0.72), although there was no difference in the absolute rate of tuberculosis in the two cohorts [87]. This was an observational study, and may have been confounded by the increased surveillance in the latter part of the study leading to increased tuberculosis diagnosis (and potentially at an earlier stage), and by the coincidental increase in the prevalence of tuberculosis in the country over this period. A large randomised controlled trial in the same region showed no difference in the efficacy of Tokyo 172 vaccine whether it was administered via the intradermal or the subcutaneous route [88]. A comparison between T cell IFN-γ production and vaccine site scar size following vaccination with Danish 1331 and Glaxo-Evans 1077 in the UK revealed no differences between the two vaccines [89].

In summary, although individual examples of difference in efficacy following the use of different BCG strains exist, these are not consistent, and none of them are convincing enough to explain the variations in the efficacy of the vaccine that have been observed.

1.2.3.2 Human genetics

It has been proposed that genetic differences between populations receiving BCG may contribute to differences in vaccine efficacy. For example, it is known that polymorphisms in HLA-DR, HLA-DQ, vitamin D receptor, IFN-γ , IL-12 and MCP1 and NRAMP genes influence mycobacterial immunity [22, 23, 90-92]. There is, however, little evidence to support a large influence of population genetic differences; Asians born in the UK were protected following BCG vaccination in MRC BCG trials, unlike those living in South India. It has been difficult to test this hypothesis further, due to the lack of sufficiently large systematic studies on specific immigrant groups.

1.2.3.3 Different strains of M. tuberculosis

Modern *M. tuberculosis* is unable to undergo lateral gene transfer, which means that it mainly evolves by genetic deletion and duplication events [93]. This means that it has evolved in a strongly clonal manner, with the emergence of separate lineages all derived from a common ancestor. Differences in the virulence of the six main modern lineages have been noted in studies of animal models, with the West Beijing (W Beijing) strain causing a higher bacillary load, increased dissemination [94] and early death [95] in mice.

In humans, there are suggestions that certain lineages of *M. tuberculosis* are associated with a more severe disease phenotype. W Beijing strains have recently rapidly spread throughout the Western Cape of South Africa, a region that was already highly endemic for tuberculosis, suggesting that this lineage possessed a survival advantage over the previously endemic Euro-American lineages [96]. Also in the Western Cape of South Africa, Hesseling *et al* showed that the Beijing and S strains were associated with an increased prevalence of extra-pulmonary tuberculosis in children compared to Latin American Mediterranean (LAM)

strain, suggesting that it possessed a greater ability to disseminate haematogenously [97]. Kong and others also described an association between W Beijing strain and extra-pulmonary tuberculosis in Arkansas [98]. Others have described an association between infection with the W Beijing lineage and treatment failure [99] or relapse [100] and between this lineage and more rapid presentation with tuberculosis meningitis [101]. Thwaites *et al* also described an association between infection with the Euro-American genotype and consolidation on chest X-ray [101].

It has been suggested that BCG may confer reduced protection against the W Beijing strain, accounting for its rapid spread throughout the world. BCG protects mice poorly against challenge with W Beijing strain [102] and also rabbits, in the tuberculosis meningitis model [82, 103]. No association between failure of the BCG vaccine and the prevalence of W Beijing has been demonstrated in humans, however (although this hypothesis does underline the importance of using clinical rather than laboratory strains in the pre-clinical testing of new TB vaccines).

In conclusion, different strains of *M. tuberculosis* probably are associated with differing levels of severity of TB disease, and it will be important to investigate the protective effect of novel TB vaccines in the context of different clinical strains.

1.2.3.4 Vitamin D levels

Antimycobacterial immunity is reduced in vitamin D deficiency, and it has been shown in individuals with the tt genotype of the TaqI vitamin D receptor polymorphism that vitamin D supplementation can significantly reduce the time to sputum culture conversion [104]. There are two vitamin D receptor polymorphisms; the t allele is associated with an increase in phagocytosis of *M. tuberculosis in vitro*, and more rapid sputum culture conversion *in*

vivo, whereas the f allele is associated with a reduction in phagocytosis and slower sputum culture conversion in pulmonary tuberculosis [104]. Lalor *et al* showed a significantly higher prevalence of vitamin D deficiency in UK infants who had not been vaccinated with BCG than their vaccinated counterparts [105]; the mechanism behind this and the effect on subsequent tuberculosis susceptibility is not known. It is not known exactly how vitamin D contributes to anti-tuberculous immunity, but it has been shown to be linked to the production of anti-microbial peptides such as cathelicidin [106], and vitamin D supplementation has been shown to increase the containment of mycobacteria in a functional whole blood assay [107].

1.3 Non-tuberculous mycobacteria

One theory as to why BCG works less well in the tropics than in temperate regions is exposure to non-tuberculous mycobacteria (NTM) [81]. More than 55 species of NTM are known, of which half are capable of causing disease in humans [62] (see Section 1.1.4 and Figure 1.1 for phylogeny). The *M. avium-intracellulare* complex is the most ubiquitous, and the prevalence of NTM is higher in hot than in cold climates [108].

There have been many attempts to model NTM exposure in different animal models, using different NTM species, strains, routes and doses, different BCG strains, routes and doses and different *M. tuberculosis* or *M. bovis* challenge strains, routes and doses. Strains and species of animals also differ, not to mention experimental design, so perhaps it is not surprising that results also differ widely, to the extent of sometimes being completely contradictory. Animal models of NTM exposure should ideally mimic human exposure as closely as possible, but this is difficult to achieve since very little is known about human

NTM exposure. It is assumed that humans are exposed to multiple species of NTM, and in particular to members of the *M. avium* complex, and that oral and to a lesser extent respiratory are the main routes of exposure, and that levels of exposure are greater in tropical countries. Many factors are not known, however, including dose and frequency of exposure, and whether exposure is primarily to live or dead organisms. Given the above assumptions, many animal models are therefore based on using *M. avium* as the sensitising agent, via the oral route.

1.3.1 Animal models of NTM exposure

The mouse is the most commonly used animal model of NTM exposure; it is readily available and inexpensive, and it is a well-established model for tuberculosis infection, with multiple immunological reagents available with which to characterise the immune response [109]. There are, however, some important differences between tuberculosis in the human and in the mouse. Mice do not have a latent stage of infection; chronic infection ultimately results in death, and histopathology is different, with non-caseating granulomas [110].

1.3.1.1 Exposure to NTM may reduce BCG protection against tuberculosis challenge

Brandt *et al* found that mice sensitised to NTM (a mixture of *M. avium, M. scrofulaceum and M. vaccae*) had lower levels of BCG recoverable from liver and spleen following vaccination with BCG, and that this translated into reduced protection against aerosol *M. tuberculosis* challenge compared with mice receiving BCG vaccine only [111]. Interestingly prior NTM exposure did not reduce the efficacy of a non-replicating protein/adjuvant vaccine in this model. Comparing soil isolates of *M. avium, M. fortuitum* and *M. chelonae* as sensitising

agents, only *M. avium* sensitisation resulted in a reduction in the subsequent growth of BCG in vaccinated mice.

Flaherty used a mouse model with the dual advantages of administering BCG prior to NTM exposure (more likely to reflect the natural sequence of events worldwide, since BCG is generally given at birth), and using multiple low doses of a clinical isolate of orally administered *M. avium*, again more likely to reflect NTM exposure in humans [112]. *M. avium* exposure resulted in increased populations of cells expressing the death receptor CD95, and in greater numbers of activated T cells in the draining lymph nodes. More inflammation and necrosis were also seen in the lungs. The protective efficacy of BCG waned over time with on-going exposure to *M. avium*.

1.3.1.2 Exposure to NTM may confer partial protection against tuberculosis challenge

However, not all experimental systems led to the same conclusions. Orme and colleagues conducted a series of experiments modelling NTM exposure in mice, and came to the conclusion that *M. avium*-exposed mice had a similar outcome to those that received BCG alone [113]. Similarly, de Lisle and others used a guinea pig model (in which the animals were less susceptible to *M. avium* infection compared to most of the mouse models discussed) to compare oral and subcutaneous *M. avium* sensitisation followed by BCG vaccination [114]. They also found that *M. avium* sensitisation conferred equivalent amounts of protection from *M. bovis* challenge compared to BCG vaccination.

Edwards *et al* used high and low virulence strains of *M. tuberculosis* isolated from South India during the Madras tuberculosis trial [115], which demonstrated zero efficacy of BCG, and found that exposure to *M. avium* prior to challenge with the low (but not the high)

virulence strain of *M. tuberculosis* conferred equal amounts of protection to guinea pigs compared to vaccination with BCG [116]. *M. avium* exposure did not interfere with BCG-associated protection against either strain of *M. tuberculosis.*

1.3.1.3 Results vary depending on dose, strain and species of sensitising organism

It is interesting that the experiments described above led to such different results. It has been shown in some experimental systems that using a different species, strain or even dose of sensitising NTM can lead to very different results. Hernandez *et al* exposed mice to different doses of killed *M. vaccae* before infection with *M. tuberculosis,* and showed that a lower dose (10^7 CFU) resulted in a protective Th1 immune response (defined as a high IL-2 to IL-4 ratio), and a higher dose of 10^9 CFU resulted in a Th2 response (defined as a low IL-2 to IL-4 ratio), with shorter survival times and more severe pneumonia [117].

Demangel noted that sensitisation with *M. avium* resulted in slower growth of BCG following vaccination, and that sensitisation with the mycobacteria *M. vaccae* and *M. scrofulaceum* did not have this effect [118]. Interestingly, they went on to vaccinate similarly sensitised mice with BCG expressing RD1 antigens, which are not present in NTM or BCG, and found that in *M. avium*- sensitised mice, BCG-RD1-vaccination confers greater protection against *M. tuberculosis* challenge than it does in BCG-vaccinated mice. The authors postulated that the BCG-RD1 prime with antigens not present in NTM may have evaded the interference that NTM sensitisation seemed to generate.

Young *et al* demonstrated differences in immune response in mice to two different strains of *M. avium*, one of which had been observed to be significantly more pathogenic in cattle than the other [119]. The more pathogenic strain resulted in greater up-regulation of

dendritic cell surface markers, and greater levels of cytokine secretion. This strain also persisted for longer in the mouse, and in the guinea pig reduced the efficacy of BCG in protecting against *M. bovis* challenge [114]. Brandt also described a difference between levels of inhibition of BCG growth between different species of NTM, and in particular that 'slow growing' NTM (more closely related to the *M. tuberculosis* complex, see Section 1.1.4, Figure 1.1) persisted and replicated more, whereas the more distantly related fast growers had no effect on BCG growth rates [111].

1.3.1.4 The use of antibiotics to eliminate NTM infection

Most investigators did not treat animals with antibiotics after exposure to NTM, and in most cases *M. avium* infection was noted to be persistent. This raises the possibility that the model may actually have been one of dual NTM and TB infection, and that some of the results described may have been due to the non-specific effects of inflammatory changes within the lungs. For example, Orme *et al* were able to isolate over 10^6 *M. avium* organisms from the lungs of NTM-exposed mice up to 150 days after exposure; the amount of time they monitored for [120], and Flaherty *et al* were able to detect *M. avium* up until day 130 after exposure [112]. Because of these worries, Brandt and Demangel used antibiotics to clear NTM infection prior to BCG vaccination [111, 118].

The difficulty in interpreting the results from these different approaches is in deciding which of them most closely resembles human sensitisation with NTM. It seems unlikely that it resembles the on-going, replicating infection that is seen in many of the mouse models described, making efforts to eliminate the infection prior to BCG vaccination logical. Hatherill *et al* studied NTM species isolated from sputum and gastric washing samples from South African children investigated for possible *M. tuberculosis* infection, and found a high

proportion of NTM isolates in children who were ultimately not diagnosed with *M. tuberculosis* infection, and whose symptoms improved without treatment [121]. It is not clear whether the isolation of NTM from these clinical samples reflected infection or transient exposure, but the resolution of symptoms without treatment suggests that the picture is less likely to be one of progressive infection.

1.3.1.5 Differentiation of mycobacterial species in assessing outcomes

Importantly, not all investigators distinguished between NTM, BCG and *M. tuberculosis* by using antibiotics to selectively inhibit the growth of NTM, BCG or both on agar plates when enumerating bacterial loads [113, 116, 120]. This could have led to critical over-estimations of the *M. tuberculosis* burden isolated from lung specimens, and to an underestimation of the efficacy of BCG/ NTM exposure in preventing tuberculosis infection.

1.3.2 Human NTM exposure

BCG has an efficacy of 0% against pulmonary TB in Malawi [64] and it has been suggested that some of this failure may be due to exposure to NTM. Studies in Karonga district in the north of the country have demonstrated isolation of high levels of NTM from the soil [108]. Black *et al* have shown that adolescents living in Malawi had marked pre-existing T cell responses to purified protein derivative (PPD) prior to BCG vaccination. After vaccination with BCG the increase in response to PPD was minimal, in contrast to in adolescents in the UK, in whom baseline responses were very low and a marked increase in PPD-specific responses post BCG-vaccination was detected [122]. These pre-vaccination immune responses to PPD were attributed to NTM exposure; the most likely explanation, given that individuals were BCG naïve, and those with a tuberculin skin test result suggestive of tuberculosis exposure had been excluded from the study. Similarly, Burl *et al* demonstrated

IFN-γ responses to PPD in infants in the Gambia who had not been vaccinated with BCG; the assumption was that these were due to NTM exposure (the possibility of *M. tuberculosis* infection was investigated, with no diagnoses made) [123].

Tuberculin skin test (TST) positivity increases with age [16], and as discussed, the efficacy of BCG decreases with age [124]. It may be that this increase in skin test responsiveness is explained by NTM exposure, and that this exposure contributes to the observed decrease in BCG efficacy. A detailed meta-analysis assessed the effect of NTM exposure on TST results, including results from 12 studies and more than 10^6 individuals. Based on comparisons between TST reactions using PPD from *M. tuberculosis* and PPD from NTM species, the authors estimate that levels of TST 'false positivity' attributable to NTM exposure range from 0.1% in Montreal to 2.3% in India [125].

Given the possible importance of NTM exposure in mycobacterial immunity and BCG efficacy, it would be very useful to be able to reliably quantify NTM exposure in an individual. It is, however, hard to characterise NTM exposure with certainty, as there are currently no defined antigens which are specific to NTM and not also present in (and therefore confounded by exposure to) *M. tuberculosis* and BCG. Although previous studies have investigated NTM exposure using PPD derived from different NTM species [126], a high proportion of proteins are shared between PPDs of different species of NTM and *M tuberculosis*, so it is hard to definitively attribute an immune response to any particular species.

1.4 Novel tuberculosis vaccines and NTM

If exposure to NTM has an effect on BCG replication and hence immunogenicity and protective efficacy, this effect might also be seen with novel tuberculosis vaccines based on BCG. Viral vectored and protein/ adjuvant subunit vaccines do not replicate, so may not be susceptible to interference by this mechanism [111]. It will be important in the development of new tuberculosis vaccines to ensure that their efficacy is not reduced by exposure to NTM, and in order to do this, specific measures of NTM exposure are needed. NTM-specific antigens could also be of diagnostic use to the veterinary and medical fields. NTM infections (for example, *M. paratuberculosis*) cause considerable morbidity and economic losses in animal husbandry [127] and can cause human disease in certain situations.

1.5 Aim of the thesis

The aim of this thesis was to define antigens which are specific to NTM (i.e. not present in the *M. tuberculosis* complex), and which could be used to study NTM exposure, without being confounded by prior BCG vaccination or by exposure to *M. tuberculosis.* Such antigens could then be used to characterise exposure to NTM in the context of trials of the new TB vaccine MVA85A, and in particular to compare immune responses to NTM between individuals in the UK, where the immune response to the vaccine is higher, and those in South Africa, where the response is lower. In addition, since the range of responses is highly variable between individuals at each trial site, immune responses to NTM could be compared between high and low responders to the vaccine at both sites. Response to NTM-specific peptides prior to vaccination could be compared with peak response to MVA85A vaccine (seen 7 days post-vaccination). The overall aim was to establish whether there is an

association between a greater prior response to NTM and a smaller peak response to the vaccine.

2 Methods

2.1 Mycobacteria

2.1.1 Non-tuberculous mycobacteria

NTM which are reported to be frequently isolated in studies of environmental and clinical isolation from the UK and South Africa were selected following literature review [121, 128-134]. These were: *Mycobacterium avium, M. avium paratuberculosis, M. intracellulare, M. kansasii, M. gordonae, M. fortuitum* and *M. absessus.* Searches were performed in the NCBI (http://www.ncbi.nlm.nih.gov/sites/genome), Swissprot (http://www.uniprot.org/help/uniprotkb) and Sanger (http://www.sanger.ac.uk/) websites for genome and protein sequence information for the selected species.

A clinical isolate of *M. fortuitum* was obtained from the Health Protection Agency (HPA) National Mycobacterium Reference Laboratory, UK, courtesy of Dr Malcolm Yates, for the purpose of genomic sequencing. Clinical *M. avium* isolates for use in the mouse model were obtained from the John Radcliffe Hospital, Oxford, microbiology department, courtesy of Dr Ian Bowler, and an additional isolate (*M. avium* 2151 SmT) was obtained courtesy of Prof Ian Orme (Colorado State University, US).

Clonal cultures were established as follows: stock solutions were plated onto Middlebrook 7H10 plates (Sigma Aldrich) and incubated for 21 days at 30°C. Single colonies were picked and inoculated into sterile smash-proof glass universal containers (Scientific Laboratory Supplies, Nottingham, UK) containing 10 ml of Middlebrook 7H9 media and 0.05% Tween 20 (Sigma Aldrich), and incubated in a shaking incubator at 30°C. The optical density (OD) at

600nm was measured at twice weekly intervals using a spectrophotometer (ThermoScientific Genesys 10 uv), and at mid-log phase (optical density reading of 1, at 600nm) the bacteria were harvested by centrifugation. They were washed once in phosphate-buffered saline (PBS), and then either processed for DNA extraction or frozen at -80°C for later use.

They were stained using TB Quick Stain Kit (BD) which is based on the Kinyoun stain, using carbolfuchsin and methylene blue, to confirm the presence of mycobacteria. The manufacturer's instructions were followed: slides were fixed by passing briefly through a flame, then flooded with reagent A for 5 minutes. They were rinsed until the water ran clear, then reagent B was added for 3 minutes. They were briefly rinsed again, air dried, and examined using a light microscope at X100 magnification under oil immersion.

Prior to use, frozen *M. avium* stocks were thawed at room temperature (RT), and diluted to 10^6/ ml in sterile PBS for intravenous inoculation into C57BL6 mice. 10^5 CFU was administered per mouse in 200 µl sterile PBS (Section 2.6.2).

2.1.2 BCG

BCG was obtained from the Statens Serum Institute (SSI), Denmark and was reconstituted using Sauton media (SSI) as per the manufacturer's instructions, and 10^5 CFU was administered intravenously to C57BL6 mice, in a volume of 200 µl per mouse, within 2 hours of reconstitution (Section 2.6.2).

2.1.3 Quantitation of mycobacteria after intravenous administration to C57BL6 mice

After each intravenous inoculation, remaining BCG or *M. avium* was quantified following culture on Middlebrooks 7H11 plates (Sigma Aldrich), to confirm the dose that the mice received. Serial 10-fold dilutions of BCG or *M. avium* were made in PBS, and 100µL of each was distributed evenly over the plate using a microbiological spreader (Sigma Aldrich). Plates were incubated at 37°C for 21 days. Colony forming units (CFU) were counted for each dilution at which distinct colonies could be visualised, and live BCG/ *M. avium* concentration was calculated as CFU/ ml.

2.2 DNA extraction

Three different methods for the extraction of DNA from mycobacteria were compared, numbered one to three, in view of the known difficulty in extracting DNA from mycobacteria.

2.2.1 Method one

DNA was extracted from pellets of mycobacteria using a modification of the DNeasy Blood and Tissue kit (Qiagen). Mycobacterial pellets were re-suspended in 100 µl PBS in eppendorfs containing approximately 100 sterile 1 mm beads and beaten in a mini bead-beater (Glen Mills Inc), setting: high, for 2 minutes. 200 µL ATL buffer and 20µl proteinase K (Qiagen) were added and the tube was vortexed and then incubated in a shaking heat block at 56°C for 4 hours. The sample was heated to 95°C for 15 min, then cooled on ice for 5 min. Chicken egg-white lysozyme was added at a concentration of 0.5 µg/ml, and incubated for

an hour at 37°C. It was vortexed for 15 seconds, then 200 µL AL buffer was added, and from this point on the Qiagen protocol was followed. Eluted DNA was quantified using a spectrophotometer (ND100, NanoDrop, readings taken at 260/280 nm), and frozen at -20°C until use. Method from A. Minassian, DPhil thesis, Oxford University.

2.2.2 Method two

Mycobacterial pellets were re-suspended in 100 µl PBS in eppendorfs containing approximately 100 sterile 1 mm beads and incubated at 56°c for 72 hours, with daily vigorous vortexing for two minutes. Samples were briefly centrifuged at 13,000 RPM and the supernatant aspirated. 200µL buffer AL was added to the supernatant, and the QIAamp DNA Micro protocol (Qiagen, for purification of genomic DNA from small volumes of blood) followed from this point on. The only protocol modification from this point on was that the final incubation with AE buffer was increased from one to five minutes. Method from C. Sander, DPhil thesis, Oxford University.

2.2.3 Method three

Mycobacterial pellets were re-suspended in 200µL sterile water and an equal volume of chloroform added. The sample was vortexed then incubated at 80°C for 30-50 minutes. It was then transferred to -20°C for 10 minutes or until required. Following thawing, the sample was vortexed, centrifuged at 13,000 RPM for 3 minutes and the upper, aqueous phase was aspirated for use. Method from M Yates, London TB Reference Laboratory (personal communication).

2.3 Hain test

Mycobacterial species were confirmed using the Hain kit GenoType Mycobacterium CM (Hain Lifescience, Germany) according to the manufacturer's instructions. In brief, DNA was amplified in a mixture containing DNA, primers, nucleotides and polymerase incubation buffer. A VWR DuoCycler PCR machine was used for amplification, with the following cycles:

5 minutes	95°C	1 cycle
30 seconds	95°C	
2 minutes	58°C	10 cycles
25 seconds	95°C	
40 seconds	53°C	
40 seconds	70°C	20 cycles
8 minutes	70°C	1 cycle

Hybridisation was then performed using filter paper strips onto which specific primers (sequences not available) had been immobilised. Amplified DNA sample was incubated with denaturation buffer for 5 minutes at RT, then added to the strip together with hybridisation buffer, and incubated in a shaking water bath at 45°C for 30 minutes. The solution was aspirated then the strip was washed with stringent wash solution, followed by rinse solution, and was then incubated for 30 minutes at RT on a shaking platform in conjugate solution. Finally each strip was washed with rinse solution then distilled water, then 1ml diluted substrate was added and the strip was protected from light without shaking. When bands were visible on the strip the reaction was stopped with distilled water, and the strips

allowed to air dry. Results were read using the interpretation chart (see Figure 3.5, Chapter 3, Bioinformatics).

2.4 Next generation sequencing

DNA was submitted to the Wellcome Trust Centre for Human Genetics (Oxford University, UK), for high throughput sequencing using an Illumina GAIIx with single end reads of 72bp and a coverage of 140MB per lane. DNA quality was assessed following fragmentation using a Covaris S220 ultrasonic nebulisation system, and run on an Agilent 2100 Bioanalyzer to verify the size distribution of the DNA fragments. The resulting whole genome shotgun (WGS) fragments were assembled into contigs using the tool Velvet, by Dr Tanya Golubchik (Experimental Statistics Laboratory, Oxford University). Details of the parameters used in the genome assembly are outlined in Section 3.2.2.4 (Chapter 3, Bioinformatics). Nucleotide BLAST was used to compare the assembled contigs of *M. fortuitum* with the sequenced genome of *M. abscessus,* another fast-growing mycobacterial species. Default parameters were used, and hits between the two species were plotted graphically using Stata (version 9, Statacorp LP, College Station, Tx, see Section 3.2.2.4, Chapter 3 for detail).

2.5 Bioinformatics

All bioinformatics work was done with the help of Dr David Wyllie, postdoctoral fellow, Jenner Institute, Oxford University. A computer running Linux (OpenSuse 11) with 2 cores, and 4 gB of RAM was used for computation. Biopython version 2.6.2 (http://www.biopython.org/wiki/BioSQL), BLAST version 2.2.21 (http://BLAST.ncbi.nlm.nih.gov/BLAST.cgi?CMD=Web&PAGE_TYPE=BLASTDocs&DOC_TYPE=Download) and MySql version 5.1 (http://dev.mysql.com/downloads/) were downloaded.

Python scripting was used to automate all steps in the pipeline. All available completed bacterial genome sequences were downloaded from the NCBI and Sanger websites into a BioSql database (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/, http://biosql.org/wiki/Main_Page). All bacterial reference proteins were also downloaded (ftp://ftp.ncbi.nih.gov/refseq/release/microbial/microbial??.protein.gpff.gz), and mycobacterial proteins were extracted from these based on having the genus 'Mycobacterium'. This resulted in a database containing complete and incomplete genome sequences for all mycobacterial species of interest with the exception of *M. fortuitum,* and in addition curated reference protein sequences for all annotated genomes. BLAST results were stored in a MySql database, and were recovered using SQL and analysed in Stata (version 9, Statacorp LP, College Station, Tx).

The Prodigal tool (v1.1) was then used to predict open reading frames on the unannotated genomes [135] (http://compbio.ornl.gov/prodigal/), in order to derive predicted protein sequences for these species. Predicted protein sequences were modified to start with a methionine irrespective of the start codon, which may be GTG as well as TTG or ATG in high GC content genomes [136]. These protein sequences were also loaded in the BioSql database.

Mycobacterial species within the database were classified into 3 groups: group 1= NTM species of interest, group 2= TB family (*M. tuberculosis, M. africanum,* BCG, *M. bovis*), group 3= other mycobacterial species (Table 3.1, Chapter 3).

NCBI Protein BLAST was used to compare all protein sequences against themselves, using NCBI default parameters, including an expect value of 10 and word size of 3 [137]. Information from all BLAST searches was recorded in a MySQL database produced for the

purpose using the BioPython BLAST parser and python scripts. Proteins were selected if they 'hit' at least 4 NTM proteins and no proteins in the TB family, in other words if the sequence was present in at least 4 common NTM and absent from the TB family. TBLASTN was then run on the selected proteins (using default parameters), and the results stored in the database. NCBI BLASTClust (http://www.ncbi.nlm.nih.gov/IEB/ToolBox/C_DOC/lxr/source/doc/BLAST/BLASTclust.html#L40) was used to arrange the selected proteins into clusters. Various clustering parameters were tried (data not shown), and >50% identity and >50% length match was selected based on successful separation of known protein families into clusters. The locateP dataset [138], consisting of predicted cellular location of all bacterial reference proteins, was downloaded into the database, to allow linkage of this information to the proteins of interest. Clusters were exported to Excel, together with information on protein number and sequence, predicated cellular location (available for reference sequences only) and whether a protein had hit a member of the TB family in the TBLASTN. Clusters were selected for peptide generation if they:

1. Contained no proteins which had hit nucleotide sequences in the TB family

2. Either a. contained a majority of proteins with a prediction to be secreted (ratio prediction secreted: prediction intracellular $\geq$ 4:1, which is an arbitrary criterion)

Or b. contained proteins for which there was experimental evidence of immunogenicity.

Clusters were aligned using the alignment tool ClustalX (http://www.clustal.org/clustal2/), and .fasta and .aln files were saved. Clusters were visualised, and ends were manually trimmed, such that sequences within a cluster were of a similar length. This was done because it was found that the Prodigal tool sometimes selected initiator sites which were

substantially upstream of the canonical start site of other family members. We trimmed such segments off, and did not investigate them further.

20-mer peptides sequences overlapping by 10 amino acids were generated computationally. A further Protein BLAST was then carried out with these sequences against all downloaded bacterial reference protein sequences. BLAST parameters were modified for short sequence length, with a significant hit considered as an exact match over the length of the peptide sequence. Peptides which hit members of the TB family or common species of non-mycobacterial bacteria were excluded.

A final check was performed, whereby a random selection of proteins from different clusters was manually entered into Protein BLAST on the NCBI server (http://BLAST.ncbi.nlm.nih.gov/BLAST.cgi).

2.6 Animal work

2.6.1 Animal welfare

All work undertaken was done under the UK Home Office PPL 30/2412. All procedures were performed by licensed individuals in accordance with the UK Animals (Scientific Procedures) Act 1986. Immunisations and the retrieval of mouse spleens were kindly carried out by Dr Alex Spencer, Dr Theresa Lambe, Miss Hazel Poyntz and Miss Simone de Cassan, and spleens were then processed by myself as outlined in Section 2.8.2.

2.6.2 *M. avium* mouse infection model

6 week old C57BL6 mice (Harlan, UK) were infected intravenously with 10^5 CFU *M. avium* (clinical isolate, John Radcliffe Hospital, Oxford) or 10^5 CFU BCG (SSI, 'negative control'

mice). 6 weeks later they were culled. Serial 10-fold dilutions of BCG or *M. avium* were made in PBS, and 100µl of each was distributed evenly over a Middlebrook 7H11 plate using a microbiological spreader (Sigma Aldrich). Plates were incubated at 37°C for 21 days. Colony forming units (CFU) were counted for each dilution at which distinct colonies could be visualised, and live BCG concentration was calculated as CFU/ ml.

Splenocyte IFN-γ responses to phytohaemagglutinin (PHA)/ phorbol 12-myristate 13-acetate (PMA) (10µg/ml/50µg/ml (PHA/PMA)), PPD (10µg/ml), pool of antigen 85A peptides (10µg/ml) and pools of investigational peptides (4µg/ml) were analysed by overnight *ex vivo* ELISpot assay. Peptides were tested in 43 pools of sizes varying from 25 to 49 peptides. *Ex vivo* ELISpot was performed as previously described [139]. In brief, nitrocellulose bottomed 96-well Multiscreen HA filtration plates (Millipore, UK Ltd) were coated with anti-mouse-IFN-γ mAb (Mabtech, UK) overnight at 4°C. Spleens were crushed and filtered through a 70 µm cell strainer (BD, France), and erythrocytes were lysed using ACK lysis buffer (0.15M NH_4Cl, 1mM KCO_3, 0.1mM Na_2EDTA). Splenocytes were washed, re-suspended in Modified Eagle's Medium alpha-modification (MEM) (Sigma) (supplemented with 10% heat-inactivated FCS, 2mM L-glutamine and 1% penicillin/streptomycin (Gibco)), and counted using a CASY counter (Schärfe Systems, Germany).

2.7 Volunteer recruitment

2.7.1 Ethical approval

Ethical approval for the recruitment of NTM-exposed individuals from the Chest Clinic, Churchill Hospital, Oxford and for permission to collect cord blood samples from the National Blood Service, Oxford, was obtained from the Oxfordshire REC B (reference

09/H0605/75). Ethical approval was already in place in Cape Town, for the collection of blood samples from healthy volunteers (REC reference 126/2006). Cryopreserved PBMC from samples collected in Portland VA Medical Center, Portland, United States (US) were kindly donated by Oxford Immunotec (REC reference IRB00004835). Ethical approval was also already in place for the collection of blood samples from healthy volunteers from the Jenner Vaccine Institute for assay optimisation (Berkshire REC, reference 06/Q1602/146).

2.7.2 'NTM-exposed' individuals

Details of individuals who had had NTM isolated from a sputum sample on at least 2 occasions were obtained from the microbiology laboratory at the John Radcliffe Hospital, Oxford. These individuals were contacted and invited to participate in the study. Clinical information was collected, including name, age, country of birth, history of BCG vaccination, travel to TB endemic countries, known contact with TB patients, medical history, history of TB infection, treatment for NTM infection and history of immunosuppressive conditions or immunosuppressive medication. See Appendix One (Section 8.1) for copies of the consent form, patient information sheet and trial protocol. Inclusion and exclusion criteria for the study were as follows:

2.7.2.1 Inclusion criteria

- Isolation of NTM from sputum on at least two occasions over the previous two years
- Competent to consent as judged by the investigator
- Over the age of 18 years

2.7.2.2 Exclusion criteria

- Currently taking systemic immunosuppressive medication other than steroids

- Taken systemic immunosuppressive medication other than steroids in the last 3 months
- Other known cause of immunosuppression
- Previous TB
- Contact with a known case of TB
- Lived or worked in a TB-endemic country

50 ml blood from volunteers recruited into the study was collected into lithium heparin tubes, and PBMC were cryopreserved within an hour of collection.

2.7.3 Cord blood samples

Cord blood samples were obtained from the National Blood Service, Oxford, following their standard collection procedures. Cord blood was collected by a trained midwife immediately following delivery of a healthy baby, following consent from the mother obtained before or during the early stages of labour, according to National Blood Service protocol. Blood was drawn into a standard blood donor collection bag, containing CPD (citrate phosphate dextrose) anticoagulant. Blood samples were processed within 1-4 hours of collection. See Appendix One (Section 8.1) for copies of the consent form and patient information sheets from the National Blood Service, Oxford.

2.7.4 South African 'TB unexposed' healthy individuals

Blood was collected from healthy individuals living in the Western Cape, South Africa, who were known to have a negative result to the Quantiferon Gold in Tube test (Cellestis), and from a second cohort of individuals known to have a positive response. 50 ml blood was

drawn into lithium heparin tubes, and was processed within 1 hour of taking the blood. PBMC were isolated and used immediately; surplus PBMC were cryopreserved.

2.8 Immunology

2.8.1 Peptides, PPD, PHA and PMA

Peptides were synthesised by ProImmune, UK, at ≥50% purity grade, with mass spectrometry results supplied (most were of a higher purity). 'NTM-specific' peptides were dissolved in DMSO at 1mg/ml and stored at -20°C. Peptides were pooled in R10 (RPMI medium, 2mM L-glutamine, 1% penicillin/streptomycin, 10% heat-inactivated FCS), Modified Eagle's Medium alpha-modification (MEM) (Sigma) (supplemented with 10% heat-inactivated FCS, 2mM L-glutamine and 1% penicillin/streptomycin (Gibco)) or AIM-V medium (Invitrogen) at a concentration of 8 µg/ml per peptide. DMSO concentration in all pools was < 0.35%.

15-mer peptides spanning the length of the *M. tuberculosis* ESAT-6 and CFP-10 proteins and overlapping by 10 amino acids (Peptide Protein Research) were dissolved in DMSO as above, and used in one pool. 15-mer peptides spanning the length of the *M. tuberculosis* antigen 85A protein (Peptide Protein Research Ltd) and overlapping by 10 amino acids were dissolved in DMSO as above, and used in one pool. PPD-T (SSI) and PPD-A (Veterinary Laboratories Agency (VLA)) were stored at 1mg/ml at -20°C. PHA/PMA (both Sigma-Aldrich) was stored at -20°C and used at a concentration of 10µg/ml/ 50µg/ml (PHA/PMA) (*ex vivo* ELISpot). PHA was used at a final concentration of 10µg/ ml (proliferation assay).

2.8.2 Mouse ex-vivo IFN-γ ELISpot assay

After splenocyte preparation (Section 2.6.2), 3 x 10^5 cells were plated in 50µl into duplicate wells. 50µl peptide, PPD, PHA/PMA or MEM alone was added to the wells, and the plates were incubated for 18-20 h in a humidified 37˚C 5% CO_2 incubator. Plates were then washed and incubated with biotinylated anti-mouse-IFN-γ mAb (Mabtech, UK), followed by an incubation with a streptavidin alkaline phosphatase polymer (Mabtech, UK). Spots were developed by addition of colour development buffer, and counted using ELISpot software (AID, Germany).

Results are expressed as spot forming cells (SFC) per million PBMC. Background responses in media only wells were subtracted from those measured in peptide-stimulated wells. A plate was considered to have passed if there were at least 200 SFC in at least one positive control well. Data were analysed using Prism v5 (GraphPad software). Wilcoxon matched pairs test was used to determine if responses were significantly different from responses in the unstimulated well ($P<0.05$).

2.8.2.1 Pre-stimulation and 48 hour elispots

Modifications to the above protocol were made for the pre-stimulation and 48 hour ELISpots. For the pre-stimulation ELISpot, 3 x 10^5 splenocytes were plated into a 96 well U bottomed tissue culture plate, and antigens added at the same concentration used in the *ex vivo* ELISpot protocol. They were incubated at 37°C for 24 hours, then gently re-suspended and plated directly onto an ELISpot plate (Mabtech, UK). The standard ELISpot protocol was then followed.

The 48 hour ELISpot was identical to the 18-20 hour *ex vivo* ELISpot, but with the modification that ELISpot plates were incubated at 37 °C for 48 hours rather than for 18-20 hours.

2.8.3 Human Immunology

2.8.3.1 PBMC isolation

50ml leucesep falcon tubes were filled with 15ml leucesep (Bio-Greiner) and pulse-spun. 25ml fresh blood was decanted into each tube and centrifuged at 1000 RPM with the brake off for 13 min. The mononuclear cell layer was washed twice in R0 (RPMI medium (Sigma Aldrich), 2mM L-glutamine and 1% penicillin/streptomycin (both Gibco)), and re-suspended in 2 ml R10 (R0 plus 10% heat-inactivated FCS). Cells were counted using either a CASY counter or a haemocytometer.

2.8.3.2 Cryopreservation

Cells were centrifuged at 1800rpm for 5 min and the cell pellets re-suspended at 20 x 10^6 cells/ ml in FCS and incubated at 4°C for 30 min. An equal volume of FCS/ 20% DMSO was added, and 1ml aliquots were rapidly transferred to a chilled 'Mr. Frosty' container (Nalgene Labware) and placed in the -80°C freezer. Cells were transferred to liquid nitrogen storage within 24 hours.

2.8.3.3 Cell thawing

PBMC were transferred from liquid nitrogen onto dry ice and partially thawed in a water bath at 37°C, one vial at a time, then transferred rapidly into 10ml of R10 at 37°C. They were washed once with R0, re-suspended in 2 ml R10, and incubated overnight with 10U/ ml benzonase nuclease (Novagen) at 37°C, 5% CO_2 in a humidified incubator.

2.8.3.4 Human ex vivo IFN-γ ELISpot assay

PBMC IFN-γ responses to PHA/PMA (10µg/ml/ 50µg/ml (PHA/PMA)), PPD (10µg/ml), pool of ESAT-6/ CFP-10 peptides (10µg/ml) and pools of investigational peptides (4µg/ml) were analysed by overnight *ex vivo* ELISpot assay. Investigational peptides were tested in R10 media, in 20 pools, each pool containing between 58 to 85 peptides. Briefly, nitrocellulose bottomed 96-well Multiscreen HA filtration plates (Millipore, UK Ltd) were coated with anti-human-IFN-γ-mAb (Mabtech, UK) overnight at 4°C. 3 x 10^5 PBMC were plated in 50µl into duplicate wells. 50µl peptide, PPD, PHA/PMA or R10 alone was added to the wells, and the plates were incubated for 18-20 hours in a humidified 37˚C 5% CO_2 incubator. Plates were then washed and incubated with biotinylated anti-human-IFNγ mAb (Mabtech, UK), followed by an incubation with a streptavidin alkaline phosphatase polymer (Mabtech, UK). Spots were developed by addition of colour development buffer, and counted using ELISpot software (AID, Germany). Results are expressed as spot forming cells (SFC) per million PBMC. Background responses in media only wells were subtracted from those measured in peptide-stimulated wells. A plate was considered to have passed if there were at least 200 SFC in at least one positive control well. Data were analysed using Prism v5 (GraphPad software). Wilcoxon matched pairs test was used to determine if responses were significantly different from responses in the unstimulated well ($P<0.05$). Cut off for a positive response was calculated as 3 median absolute deviations (MADs) above the median.

2.8.3.5 Human ex vivo IFN-γ ELISpot assay modifications

Various modifications of the *ex vivo* ELISpot assay were evaluated, using PBMC from blood from volunteers from the Jenner Institute, who had known positive T cell IFN-γ responses to PPD.

2.8.3.5.1 Modification one: pre-stimulation

3 x 10^5 PBMC were plated in R10 into a 96 well U bottomed tissue culture plate with antigen at the same concentration as in the *ex vivo* ELISpot, and incubated for 24 hours in a humidified 37˚C 5% CO_2 incubator. Cells were re-suspended and plated in 50µl into duplicate wells on an ELISpot plate, and incubated for 18-20 hours in a humidified 37˚C 5% CO_2 incubator. From here the standard ELISpot protocol was followed (Section 2.8.3.4).

2.8.3.5.2 Modification two: pre-stimulation plus PBMC washing

PBMC were plated and incubated for 24 hours in a 96 well U bottomed tissue culture plate as in modification one. The plates were centrifuged at 1500 RPM for 7 minutes at RT, the supernatant flicked off, and the PBMC re-suspended in 100µL fresh R10. They were then plated into ELISpot plates, incubated for 18-20 hours in a humidified 37˚C 5% CO_2 incubator, and the standard ELISpot protocol followed.

2.8.3.5.3 Modification three: pre-stimulation plus PBMC washing plus peptide re-stimulation

This protocol was the same as modification two, except that PBMC were plated onto the ELISpot plate in a volume of 50µl and then peptides or antigens were added, also in a volume of 50µl.

2.8.3.5.4 Modification four: pre-stimulation plus PBMC washing, FACS tubes

This protocol was identical to modification two except the pre-stimulation step was performed in 10 ml polypropylene tubes (BD biosciences) rather than a 96 well plate. This was done as there were concerns that cell adhesion to 96 well plates may result in the transfer of reduced numbers of PBMC to the ELISpot plate.

2.8.3.6 Cultured ELISpot assay

PBMC were thawed (Section 2.8.3.3), and rested for either 4 hours or overnight. PBMC were plated in duplicate in 24 well plates in AIM-V media (Invitrogen) at $1x10^6$ cells/ml ($1x10^6$ cells total), with PPD (1μg/ml), ESAT-6/CFP-10 (1μg/ml), selected NTM-specific peptide pools (1μg/ml), and incubated in AIM-V media with interleukin 7 (IL-7, 10ng/ml, Gibco) for 12 days in a humidified 37˚C 5% CO_2 incubator. On day 3 interleukin 2 (IL-2, 20U/ml, Gibco) and IL-7 (10ng/ml) were added, and on days 7 and 10 half of the media was replaced with fresh media, and IL-2 (20U/ml) added. On day 12 the cells were washed 3 times with AIM-V, and counted using a Coulter Counter (Coulter AcT diff, Beckman Coulter). They were plated on a coated and blocked ELISpot plate for detection of IFN-γ (Section 2.8.3.4, human *ex vivo* ELISpot) at 3 x 10^4 or 1 x 10^4 cells per well, and incubated with the same antigens as previously, at the following concentrations: PPD 1 μg/ml, ESAT-6/CFP-10 peptides 1 μg/ml, NTM-specific peptides 4μg/ml. ELISpot plates were incubated for 18-20 hours and developed as described above (Section 2.8.3.4, human *ex vivo* ELISpot).

2.8.3.7 Human proliferation assay: Ki67

All cryopreserved PBMC samples from UK-based NTM-exposed individuals and cord blood samples were analysed, where remaining cell numbers permitted [140]. Cells were

incubated with media alone ('RPMI/AB/glutamine', RPMI with 12.5% human AB serum (Sigma-Aldrich) and 2mM L-glutamine (Gibco)) or PPD (2 μg/ml), NTM-specific peptide pool number 5, 6, 11 or 17 or ESAT-6/ CFP-10 pool (all 1μg/ml) for 6 days at 37°C with 5% CO_2. On day 3, PHA (1μg/ml) was added to one of the 'media only' wells. On day 6, PBMC were stained with LIVE/DEAD Fixable Violet Dead Cell Stain (Invitrogen) and fixed with BD FACS Lysing Solution (BD Biosciences). Following permeabilisation with Perm/Wash (BD Biosciences) cells were incubated with the following monoclonal antibodies: anti-CD3-QDot 605, anti-CD4-APC, anti-CD8-PerCP-Cy5.5 and anti-Ki67-PE. All antibodies were from BD Biosciences except for CD3-QDot 605, which was from Invitrogen, and were used in pre-determined optimal concentrations. Following washes, samples were acquired on a BD LSRII flow cytometer (BD Biosciences, San Jose, CA). Time gates (excluding fluctuations in fluorescence), antibody aggregate exclusion gates and forward scatter/ side scatter gates (selecting singlets) were followed by gating on live CD3 positive cells, then CD4 or CD8 positive cells, then Ki67 positive cells. Data were analysed using FlowJo Software version 8.8.3 (Treestar Inc.) and GraphPad Prism version 5. Proportion of Ki67 positive cells was used as the readout of proliferation [140]. A sample was considered to have passed if percentage proliferating cells from the un-stimulated control was less than 2%, if proliferation to either PPD or to PHA was greater than 10% and if proliferative responses to ESAT-6/CFP-10 peptides was less than 2%. In addition to frequency of proliferating cells, the stimulation index (SI) was calculated as response in test well/ response in un-stimulated well; a SI greater than 2% was considered positive.

2.8.3.8 Human proliferation assay: Oregon Green staining

Oregon Green staining was compared with Ki67 staining as a method of detecting proliferating cells. The proliferation assay was set up as above, following Oregon Green labelling of cells. On day one, prior to incubation of PBMC with antigens, PBMC were incubated with Oregon Green (Invitrogen) at 5µg/ml for 7 minutes (RT), with vigorous vortexing for 10 seconds half way through the incubation. An equal volume of sterile PBS was added and the cells were mixed gently and incubated for a further 4 minutes (RT). They were washed once with PBS, then re-suspended in RPMI/AB/glutamine at 1 x 10^6/ml after counting on a Coulter Counter (Coulter AcT diff, Beckman Coulter). The Ki67 protocol above was followed from this point on. The gating strategy was also as above, except the final gating was on Ki67 hi/ low cells versus Oregon Green high/ low cells.

2.8.3.9 Multiplex bead analysis

PBMC were thawed (Section 2.8.3.3) and rested overnight in AIM-V media and 10U/ ml benzonase at 37°C, 5% CO_2 in a humidified incubator. They were counted using a CASY counter and plated in a 96 well U-bottom plate at 1x10^6/ml (50,000 per well). They were incubated with PHA, PPD, ESAT-6/CFP-10 and NTM-specific peptide pools 5, 6, 11 and 17 (all reagents at 1 µg/ml) in a total volume of 200 µl for 3 days, at 37°C in a humidified incubator (5% CO_2).

On day 3 the plates were centrifuged at 1500 RPM for 7 minutes (RT) and 200µl supernatant was removed into a new 96-well plate. 25µl was used for 11-plex Th1/Th2 cytometric bead array (eBioscience), according to the manufacturer's instructions. The remaining supernatants were stored and thawed at a later stage for use in a TGF-β cytometric bead array (eBioscience), where again the manufacturer's instructions were followed.

For the 11-plex multiples assay, the protocol was as follows. In brief, standard curves were prepared by serial dilution of standard cytokine solutions, for each of the analytes to be measured. A bead mixture containing different beads, each type coated with different antibodies to a different analyte to be measured, was made after careful vortexing of each of the bead solutions. 25µl of either investigational sample (cell culture supernatant) or standard curve sample was added to a well of a customised 96-well plate, and the following reagents were added. 25µl bead mixture (vortexed) and 50µl biotin-conjugate mixture. The plate was incubated at RT protected from light on a shaking platform at 500 RPM for 2 hours. The plate was washed twice using a vacuum filter manifold (DA7C Charles Austin pump, Whatman filter block) using assay buffer, then 100µl assay buffer and 50µl streptavidin-PE solution were added to each well. The plate was incubated in a similar manner for one hour, and the plate again washed twice with assay buffer. Finally, 200µl assay buffer was added and samples were mixed well and transferred into flow cytometer acquisition tubes (BD biosciences); a further 300µl assay buffer was added prior to sample acquisition.

For the TGF-β assay, PBMC culture supernatants were first diluted 1 in 10 in assay buffer by adding 20µl sample to 180µl assay buffer. 20µl 1M hydrochloric acid was added, the sample was incubated for one hour at RT and 20µl 1M sodium hydroxide was added. From this point on the 11-plex protocol outlined above was followed.

Samples were acquired on a CyAn ADP flow cytometer (Beckman Coulter), using the gating strategy recommended. Samples were excluded where response to PHA or PPD was less than 10pg/ml.

2.9 Statistical analysis

All data were non-parametrically distributed, so non-parametric tests of significance were used throughout. Statistical analysis was performed, and graphs were drawn, using Prism 5 (version 5, GraphPad), STATA (version 9, Statacorp LP, College Station, Tx) or Excel (Microsoft). MySql was used to retrieve basic summary statistics relating to the bioinformatics process, and bioinformatics data were further analysed and graphs were produced using STATA.

When observations were paired, for example comparing responses to pools of peptides versus response to media alone within a single mouse, or comparing different experimental methods using the same PBMC samples, the Wilcoxon matched pairs test was used. When observations were not paired, the Mann- Witney test was used, for example comparing responses between different groups. Friedman test with Dunn's multiple comparison test was used to compare IFN-γ responses between DMSO concentrations of 0.3%, 0.2% and 0.1%. $P>0.05$ was taken as significant. Median values were shown on all graphs using a horizontal line, and the range and interquartile ranges were also shown on box and whisker plots.

The Wilcoxon matched pairs test was used to compare responses to peptide pools with response in the unstimulated well; $P>0.05$ was considered a significant difference. ELISpot and proliferation assay cut-offs were calculated using the Median Absolute Deviation (MAD). This is a measure of statistical dispersion which is less influenced by outlying variables in a dataset compared to, for example, standard deviation. Cut offs were defined as the median plus 3 MADs [141].

Spearman correlation was used to study the relationship between response to ESAT-6/ CFP-10 peptide pool and NTM-specific peptide pools in healthy South Africans. Linear regression was used to produce a line of best fit for illustration.

3 Bioinformatics

3.1 Introduction

Knowledge of antigens which are present in common NTM but absent from the *M. tuberculosis* complex would allow human exposure to NTM to be studied in more detail. Currently, most studies of NTM exposure rely on measuring responses to PPD (or on comparing immune responses to PPD from different species of mycobacteria); both strategies are limited by the large number of proteins which are shared between different species of mycobacteria. The availability of an antigen (or antigens) specific to NTM may allow better testing of the hypothesis that NTM exposure in the tropics contributes to the lack of efficacy of BCG in these regions.

This chapter describes the processes leading to the definition of families of proteins which are present in common species of NTM and are absent from the *M. tuberculosis* complex, and involves the following steps:

- Defining NTM which are prevalent in the two study sites; the UK and South Africa
- Downloading publically available genome and protein sequence information from these NTM strains
- Predicting protein sequences from genome sequence information when protein sequence information was not available
- In the case of one species, using next generation sequencing to obtain basic sequence information, which involved:
    - Obtaining and establishing clonal cultures of a clinical isolate of the organism

    - Optimising DNA extraction from mycobacteria and verifying the quality of the DNA for sequencing
    - Using Illumina Genome Analyzer 11x (GA11x) technology to obtain sequence information which was assembled into contigs (contiguous DNA sequences) to be used in the pipeline above
- Defining proteins which are present in common NTM and absent from the *M. tuberculosis* complex using the protein Basic Local Alignment Search Tool (BLASTP, NCBI)
- Verifying the absence of these proteins from members of the *M. tuberculosis* complex using TBLASTN (which searches translated nucleotide queries)
- Ensuring that there was no short sequence identity between peptide sequences from these proteins and all bacterial reference proteins using BLASTP

All bioinformatics work was done with the help of Dr David Wyllie, postdoctoral fellow, Jenner Institute, Oxford University. Next generation sequencing was performed by the Wellcome Trust Centre for Human Genetics (Oxford University, UK). The *M. fortuitum* sequence was assembled by Dr Tanya Golubchik (Experimental Statistics Laboratory, Oxford University).

3.2 Results

3.2.1 NTM species selected for inclusion in the pipeline

NTM common to the study areas were defined as: *Mycobacterium abscessus, M. avium, M. avium paratuberculosis, M. fortuitum, M. gordonae, M. intracellulare* and *M. kansasii,* based on studies of clinical and environmental isolation in the UK and South Africa [121, 128-134]. The *M. tuberculosis* complex was defined as: *M. tuberculosis, M. africanum, M. bovis* and BCG.

NTM can be broadly divided into slow and fast growing species, with fast growing NTM appearing on solid culture media within 7 days, and slow growing NTM becoming visible after more than 7 days of incubation. Phylogenetic trees contructed using sequence polymorphisms in genes such as 16S ribosomal RNA, RPOB, and the heat-shock protein 65 (hsp-65) confirm this grouping, and further outline relationships between species (Figure 1.1).

Of the NTM species defined as 'common to the study areas' (above), *M. abscessus* and *M. fortuitum* are fast growing and *M. avium, M. avium paratuberculosis, M. gordonae, M. intracellulare* and *M. kansasii* are slow growing NTM. Members of the *M. tuberculosis* complex of organisms are also slow growing mycobacteria. All of the above NTM are capable of causing disease in human subjects, although they rarely do so unless the individual concerned is either immunosuppressed or has significant lung pathology, for example bronchiectasis or cavities from previous pulmonary TB.

M. abscessus is a common species of NTM infecting individuals with cystic fibrosis, a condition in which severe bronchiectasis predisposes to repeated and polymicrobial chest

infections. *M. avium, M, avium paratuberculosis* and *M. intracellulare* are very closely related to each other, and are members of the 'mycobacterium avium complex' (MAC) of mycobacteria. Infections with this group of micro-organisms were particulartly common in individuals with advanced HIV infection prior to the development of effective anti-retroviral therapy [133], and are still the most common NTM causing clinical infections [134, 142]. *M. kansasii* commonly causes pulmonary infection, and has been reported in HIV-negative miners, particularly those with silicosis or previous TB therapy [129], and also in individuals with a history of previous TB therapy [142]. *M. fortuitum* [128, 133] and *M. gordonae* infections also occur, although less commonly than the above [133].

3.2.2 Available NTM genome and protein sequence information

From the NTM species of interest and *M. tuberculosis* complex, complete and annotated genome sequences were available for *M. avium, M. avium paratuberculosis, M. abscessus, M. gordonae, M. tuberculosis, M. bovis* and BCG. Un-annotated genome contigs (contiguous DNA sequence reconstructed from a set of overlapping DNA segments) were available from incomplete genome sequencing projects of *M. intracellulare, M. kansasii and M. africanum*. There was no genome sequence information available for *M. fortuitum* or *M. gordonae.* An annotated genome sequence is one on which an automated gene prediction programme has been applied, resulting in the 'annotation' of the sequence with open reading frames and predicted protein sequences.

20 completed and annotated mycobacterial genome sequences (Appendix Two, Section 8.2) were downloaded, along with approximately 102,000 associated reference proteins. These are summarised in Table 3.1. A reference protein is a curated protein sequence in the NCBI reference sequence database. Reference protein sequences are often defined using

automated annotation programmes, and the existence of the protein is not always experimentally validated (see later). Three incomplete mycobacterial sequencing projects were downloaded from the NCBI and Sanger Centre websites (Appendix Three, Section 8.3). Sequence information from the next generation sequencing of the *M. fortuitum* genome was also loaded into the database. Mycobacterial species within the database were classified into 3 groups: group 1= NTM species of interest, group 2= *M. tuberculosis* complex, group 3= all other mycobacterial species (Table 3.1).

Table 3-1: Genome sequences downloaded

Group 1: NTM of interest	**Group 2: *M. tuberculosis* complex**	**Group 3: Other mycobacteria**
M. abscessus	*M. africanum*[1,6]	*M. gilvum*
M. avium 104	*M. bovis* AF2122/97	*M. leprae*[8]
M avium paratuberculosis K10	*M. bovis* BCG[7]	*M. marinum*
M. fortuitum[4]	*M. tuberculosis*[5]	*M. smegmatis*
M. intracellulare[1,2]		*M. ulcerans*
M. kansasii[1,3]		*M. vanbaalenii*
		Mycobacterium JLS
		Mycobacterium KMS
		Mycobacterium MCS

Genome sequences downloaded from the NCBI and Sanger websites. Sequences are divided into three groups: NTM species of interest to the project, the *M. tuberculosis* complex of mycobacteria and NTM sequences not of interest to the project.
Unless stated, sequences were downloaded from the NCBI website [143]
1. incomplete sequence, 2. *M. intracellulare* ATC 13950, 3. *M. kansasii* ATCC 12478, 4. next generation sequencing, Oxford University, 5. *M. tuberculosis* CDC1551, *M. tuberculosis* F11, *M. tuberculosis* H37Ra, *M. tuberculosis* H37Rv sequences downloaded, 6. Sanger website [144] 7. *M. bovis* BCG str. Pasteur 1173P2, *M. bovis* BCG [Fiocruz - FAP], 8. *M. leprae* Br4923, *M. leprae* TN

3.2.3 Next generation sequencing of *M. fortuitum* genome

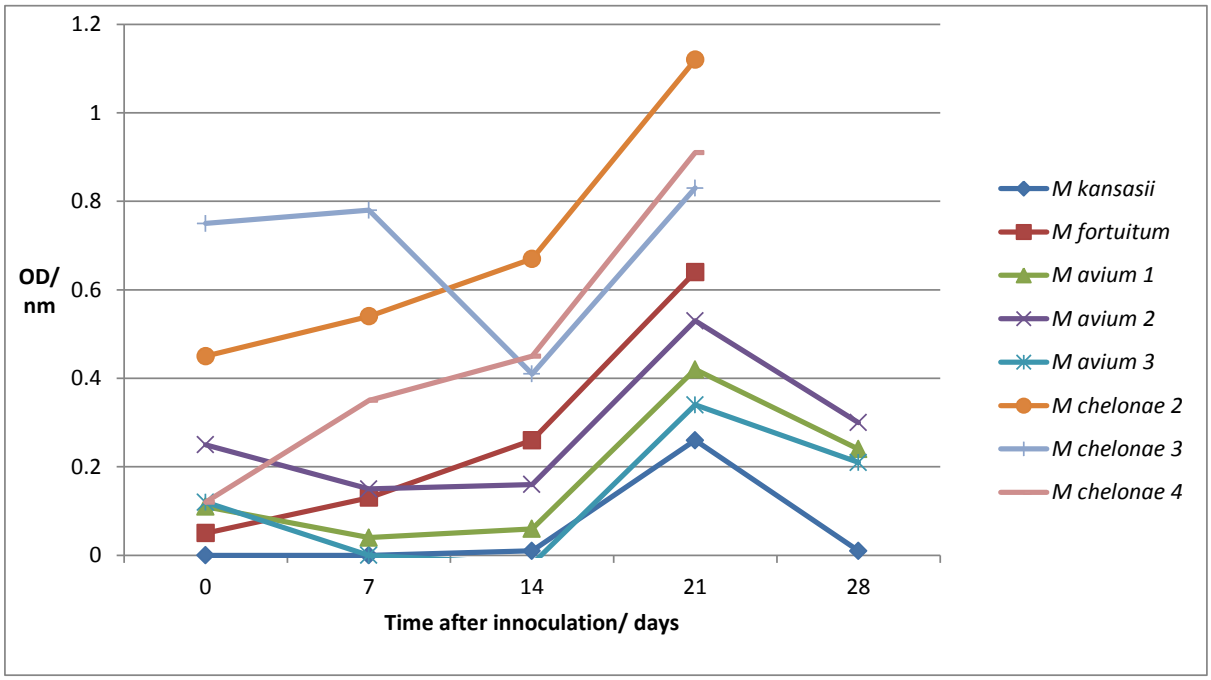
Since there was no genome sequence information available for *M. fortuitum*, it was decided to use the new next generation sequencing facility available at the Oxford Wellcome Trust Centre for Human Genetics to obtain sequence information for this species.

3.2.3.1 Mycobacterial culture

A clinical *M. fortuitum* isolate was obtained from the Department of Microbiology, John Radcliffe Hospital, Oxford, courtesy of Dr Ian Bowler, and plated onto Middlebrook 7H11 plates. Single colonies were picked, and established in liquid culture. Bacteria were harvested at mid-log phase (Figure 3.1), the presence of acid-fast bacilli was confirmed using the TB quick stain kit (Figure 3.2) and DNA was extracted using a modified bead-beater method (Method one, Section 2.2, Chapter 2, Methods).

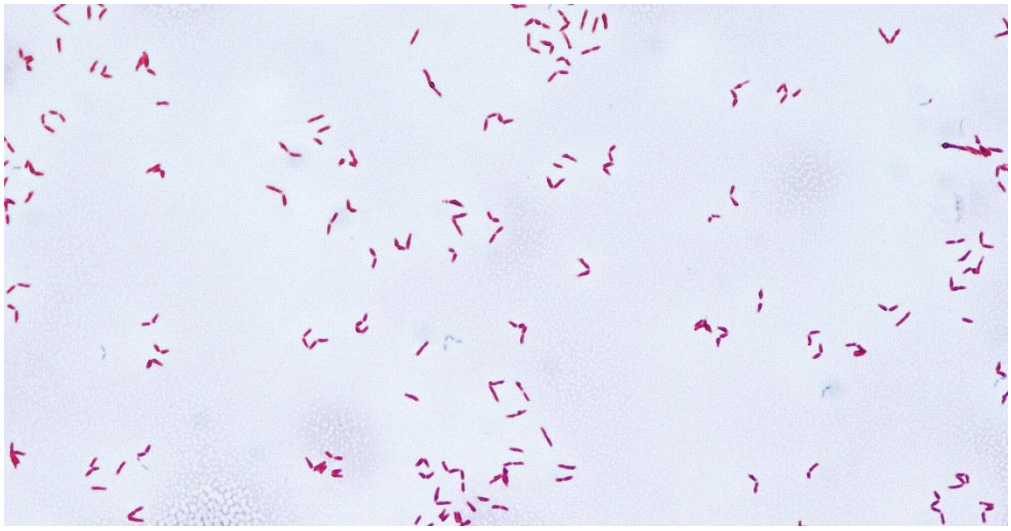
Isolates of other NTM species were also obtained from the Department of Microbiology, John Radcliffe Hospital, Oxford and also from the TB Reference Laboratory, Royal London Hospital, London, courtesy of Dr Malcolm Yates. These isolates were obtained as part of a separate project which is not described here, but some of the results presented include these isolates.

Figure 3-1: Mycobacterial growth



Graph showing optical densities of cultures of different species of NTM. An OD of one was taken as mid-log phase of growth (600nm). Other species of mycobacteria were obtained for a part of the project not described here, and are shown together with *M. avium* and *M. fortuitum.*

Figure 3-2: *M. fortuitum* microscopic appearance



Microscopic appearance of *M. fortuitum* isolate under oil immersion, X60 magnification. Stained with TB Quick Stain (based on the Kinyoun stain, using carbolfuchsin and methylene blue).

3.2.3.2 DNA extraction

Standard DNA extraction methods cannot be used on mycobacteria due to the impermeable cell wall, so three adapted methods were compared, numbered one to three (Table 3.2). All methods incorporated a high temperature incubation step, and methods one and two also used mechanical disruption. Methods one and two were based on the Qiagen DNA extraction kit (tissue). Method one consisted of a modification of the Qiagen DNeasy protocol for Blood and Tissue, using mechanical disruption with a bead-beater, followed by a 95^{o}C incubation step and lysozyme digestion. Method two was similar, but involved a three day incubation step at 56°C with vortexing. Method three involved a very simple chloroform extraction procedure, with an incubation step at 80°C. The results are shown in Table 3.2. Method one was selected for further DNA extractions, as it provided good yields of DNA. Method 3 was discarded despite the high yields of DNA due to concerns that the likely low purity of the extracted DNA might have affected the reliability of the sequencing. This method is routinely used as a basis for the Hain test in mycobacterial speciation at the National TB Reference Laboratory (M. Yates, National TB Reference Laboratory, personal communication), but the effect of using DNA from this method on the accuracy of next generation sequencing was not known.

Table 3-2: Comparison of DNA extraction methods

Method	**Summary**	**DNA yield (ng/µl)/ *M. fortuitum***	**DNA yield (ng/µl)/ *M. chelonae***
Method 1 [a]	Modified Qiagen kit (tissue) 2 min beadbeater 60 min 95°c incubation Lysozyme digestion	154	90
Method 2 [b]	Modified Qiagen kit (blood) 72h incubation at 56°c Daily vortexing	13	30
Method 3[c]	Chloroform Vortex 80°c 50 min Freeze, use aqueous phase	350	60

Three different methods of extracting DNA from mycobacteria were compared, and the DNA yields following extractions from *M. fortuitum* and *M. chelonae* are shown.
a: A Minassian, unpublished data, b: C Sander, unpublished data, C: M Yates, personal communication

3.2.3.3 DNA quality assessment

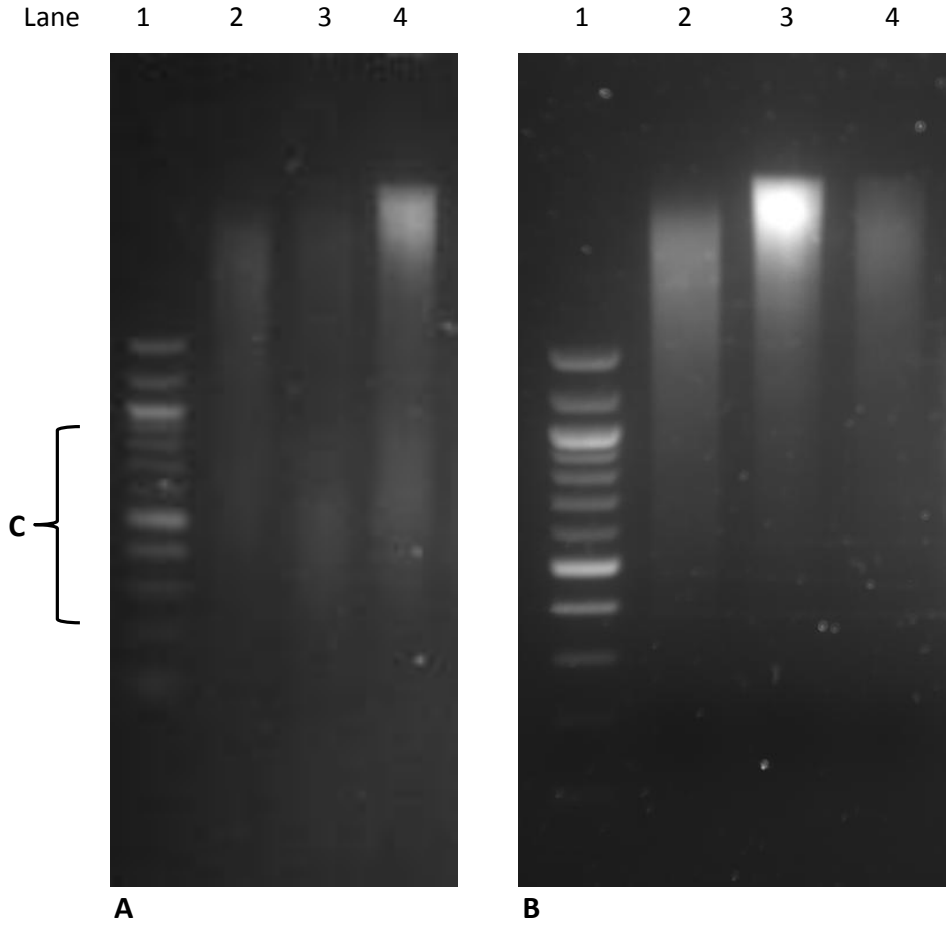
Prior to sequencing using the Illumina GA11x, DNA was fragmented using a Covaris S220 system [145]. This uses focussed, high frequency acoustic energy waves to shear DNA in an unbiased manner into fragments of a uniform and specified size (200 base pairs (bp) in this case). There was a concern that, prior to this step, using the bead-beater method to extract the mycobacterial DNA may have resulted in DNA fragmentation, with possible bias in the specific coding regions fragmented. If this was the case, the Covaris system could have

further fragmented these already short sequences of DNA, resulting in their exclusion from the sequencing process on the basis of small size.

This possibility was tested by running the extracted DNA on a 1.5% agarose gel prior to fragmentation with the Covaris, to ensure that it did not contain low molecular weight fragments (Figure 3.3). Faint staining was visible at approximately 500bp, which might either have indicated the presence of RNA or of short fragments of DNA, so the sample was incubated with RNAase for 30 minutes at 37^{o}C and run again on the gel. This resulted in removal of the high molecular weight staining, suggesting that the 500bp staining was caused by RNA contamination of the sample, and that the DNA extracted was of a high molecular weight.

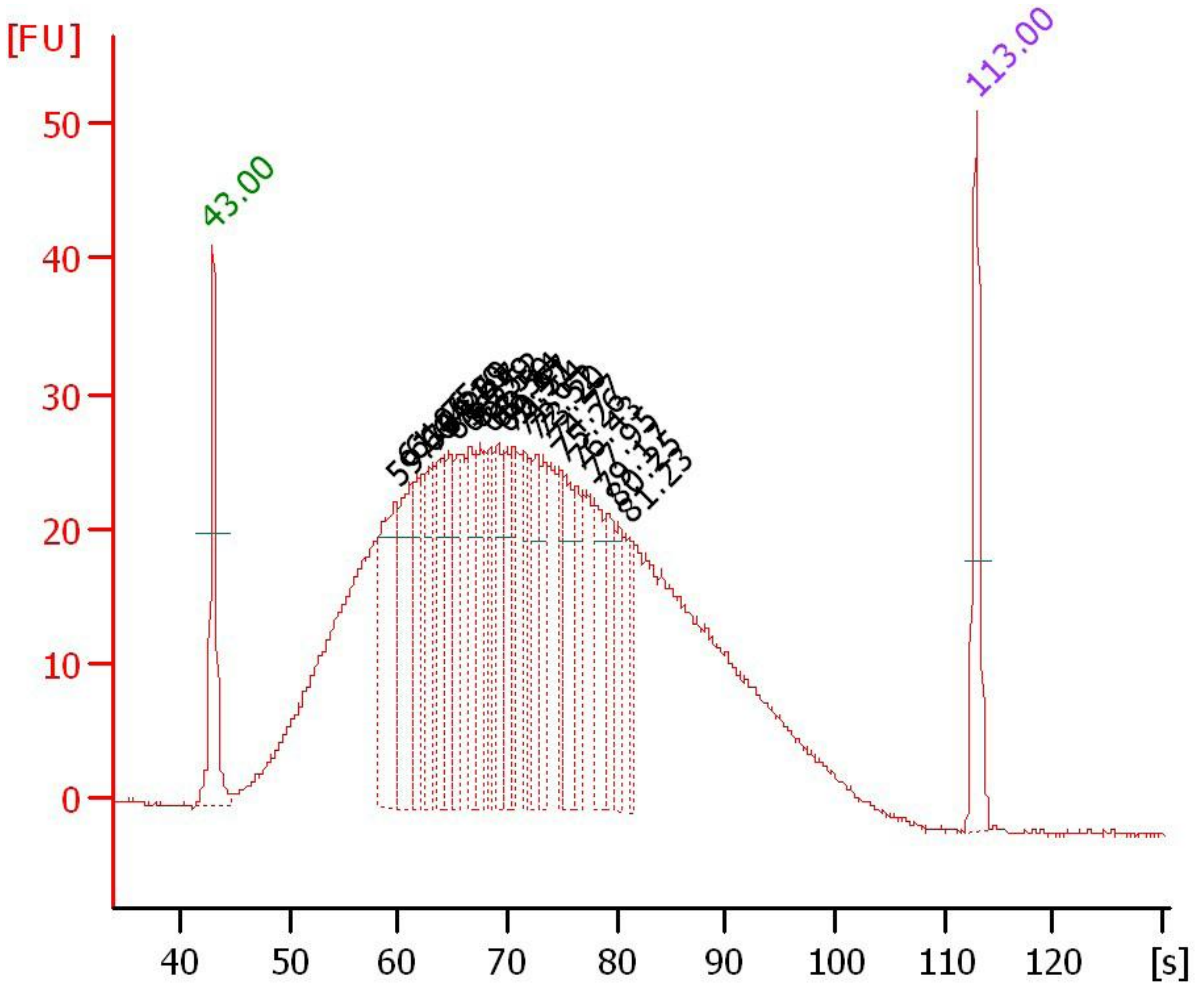
5µg DNA from *M. fortuitum* was fragmented using the Covaris S220 (Oxford Wellcome Trust Centre for Human Genetics), and it was confirmed on an Agilent 2100 tube gel electrophoresis (TGE) bioanalyzer that the peak molecular weight of the DNA fragments was 200 bp (Figure 3.4).

Figure 3-3: DNA separation using 1% agarose gel



1.5% agarose gel of extracted DNA, before (A) and after (B) RNAase digestion. *M. chelonae* isolates shown in addition to *M. fortuitum* as the ~500 bp staining can be seen more clearly in these samples.
Gel A: DNA extracted from clinical isolates of *M. fortuitum* and *M. chelonae*
Lane 1: 100bp ladder, lane 2: *M. fortuitum* DNA, lane 3: *M. chelonae* DNA (clinical isolate 2), lane 4: *M. chelonae* DNA (clinical isolate 3)
Gel B: DNA samples after incubation with RNAase for 30 minutes at 37^{o}C
Lane 1: 100bp ladder, lane 2: *M. fortuitum* DNA, lane 3: *M. chelonae* DNA (clinical isolate 3), lane 4: *M. chelonae* DNA (clinical isolate 4)
C: Region of staining of low molecular weight fragments of nucleic acid

Figure 3-4: DNA fragmentation plot

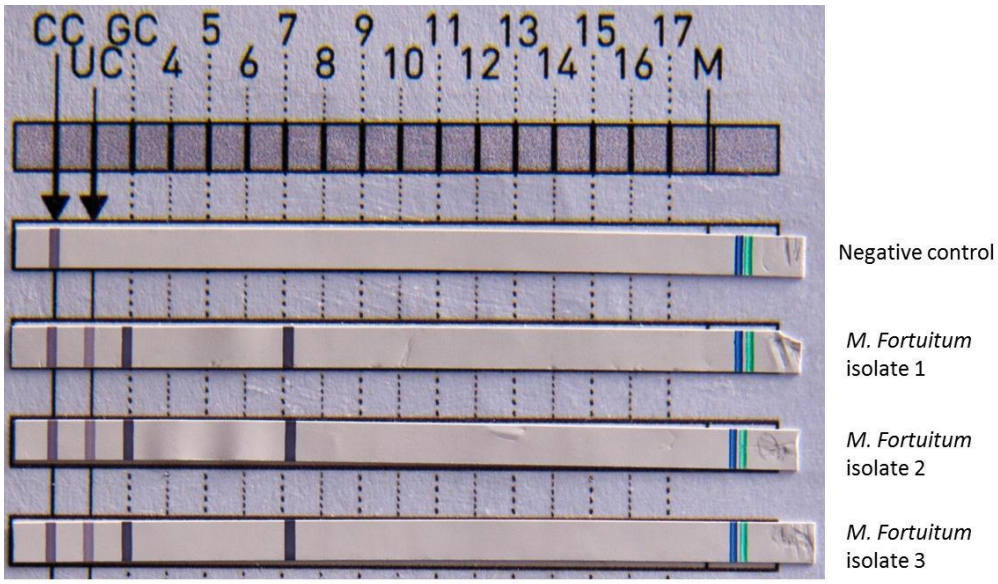


DNA fragmentation plot of fragmented *M. fortuitum* DNA produced using a Covaris S220 fragmentation system. Data are expressed as FU (fluorescence units) on the Y axis, and on the X axis DNA separation based on number of seconds. 60 seconds is roughly equal to 150 bp and 70 seconds to 200 bp. Peak DNA size is therefore about 200bp. Markers are 15bp (43 seconds) and 1500bp (113 seconds) in size.

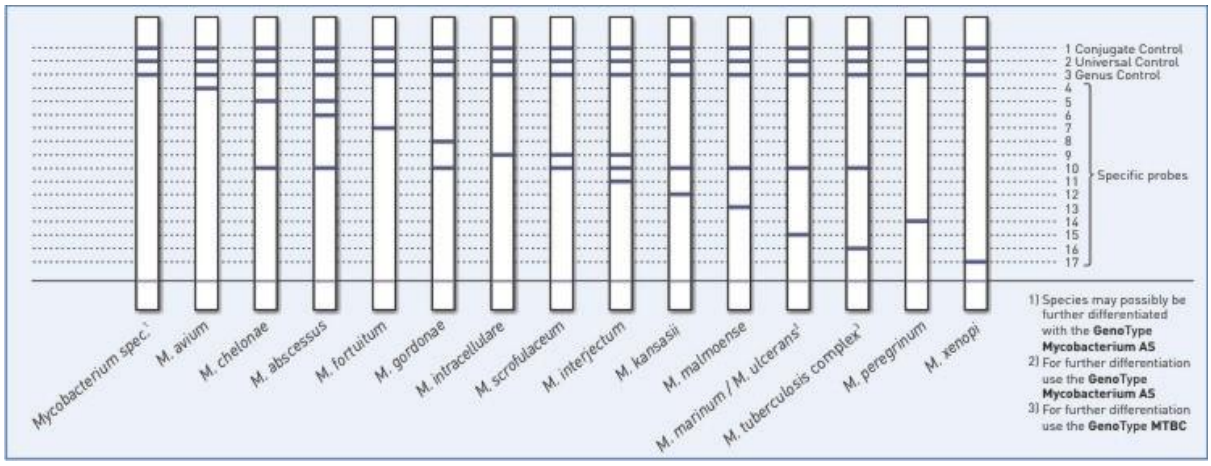
The Hain test (CM, Hain Lifesciences) was used to confirm that the identity of the isolate was *M. fortuitum* (Figure 3.5). This is a PCR-based test in which mycobacterial DNA is amplified, then hybridised with primers specific for a range of mycobacteria bound onto a filter paper in a series of rows. Bound single stranded DNA is then visualised via an alkaline phosphatase-based colourimetric reaction (http://www.hain-lifescience.de/en/technologies

/dnastrip.html).

Figure 3-5: Hain test results demonstrating the identity of *M. fortuitum*



A



B

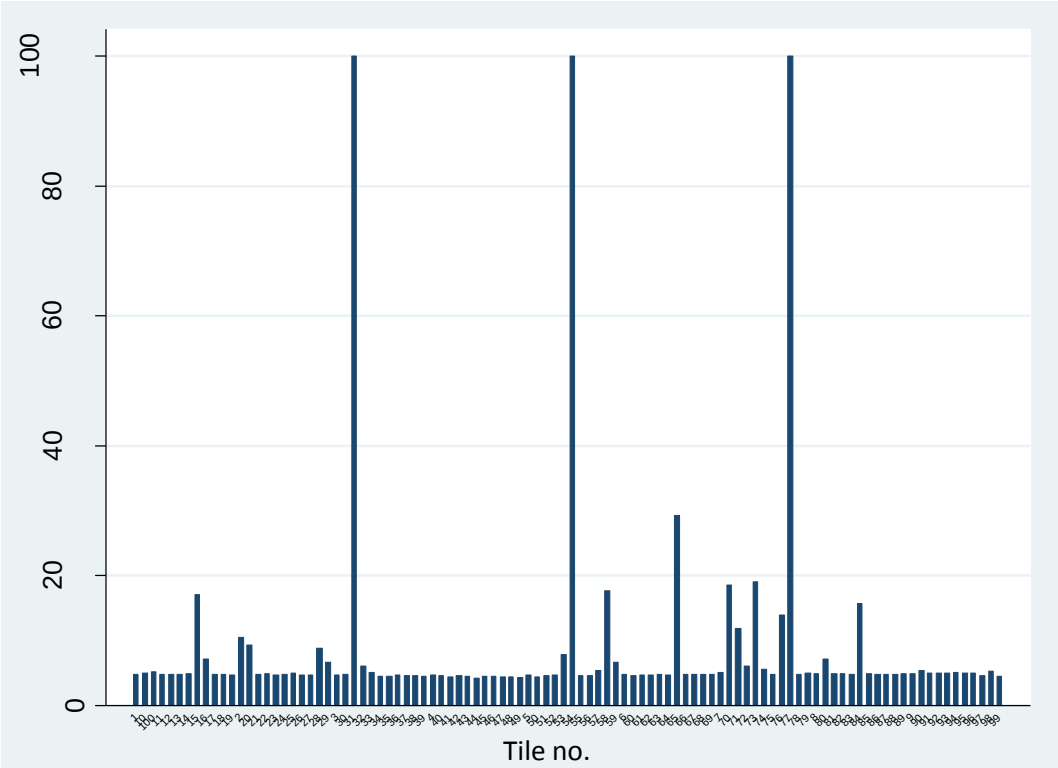
A: Hain test ('CM') results from the identification of three isolates of *M. fortuitum*
B: Key to interpretation of Hain test ('CM') results

3.2.3.4 Next generation sequencing of M. fortuitum genome

Next generation sequencing was carried out on the fragmented *M. fortuitum* DNA using an Illumina GA11x sequencing machine at the Oxford Wellcome Trust Centre for Human Genetics. Single end 72bp sequencing was performed on 2µg DNA, using one lane of the sequencer. Estimated coverage was 40-fold.

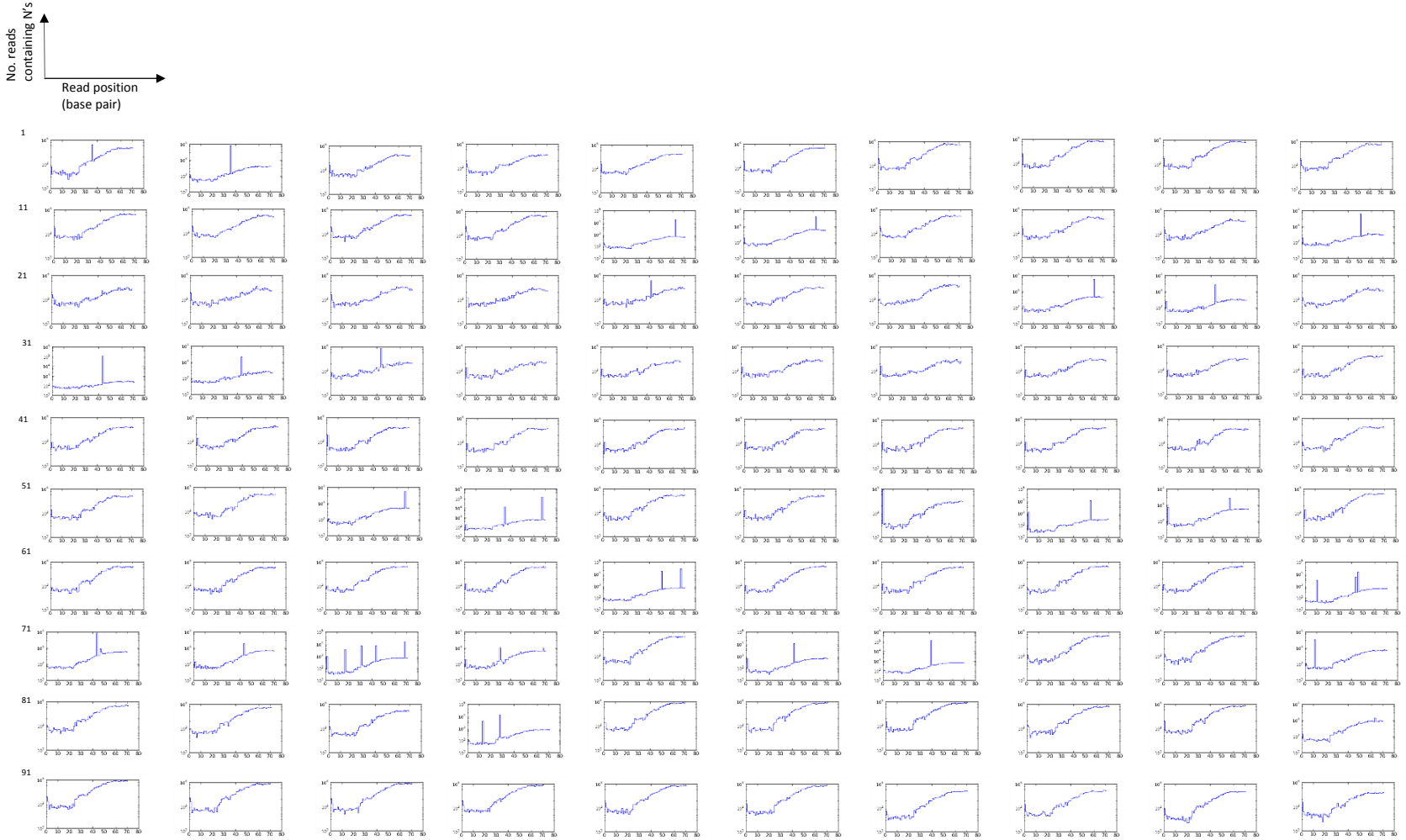
12,523,740 reads were produced, and were assembled into contigs using the tool Velvet, by Dr Tanya Golubchik (Experimental Statistics Laboratory, Oxford University). K-mer size was 31, with coverage cut-off of 5.19 and expected coverage of 140 (see discussion for an explanation of these terms). There was a high percentage of N's (or uncertain base calls) in the sequencing data, which made sequence assembly difficult, and read quality was observed to deteriorate towards the end of each read (Figure 3.6B). Prior to assembling the sequence information using Velvet, reads were trimmed to filter low quality sequencing data, using a strategy that preserved a minimum read length, irrespective of the sequence quality for each read. Everything was trimmed (from the first base onwards) where the quality represented the minimum quality for that read, as long as this left at least 35 bp of the read. For example, if the lowest quality base call was at the 40^{th} position in the read, this is where the read would be cut off. Using this read trimming strategy followed by genome assembly with a k-mer size of 31, coverage cut-off of 5.10 and expected coverage of 140 resulted in a genome assembly with 1043 contigs with an N50 of 15,810 bp, largest contig 39,343 bp (Figure 3.7). N50 is the contig length such that it and all contigs larger than it constitute half the bases of the assembly.

Figure 3-6: Distribution of reads containing Ns



A

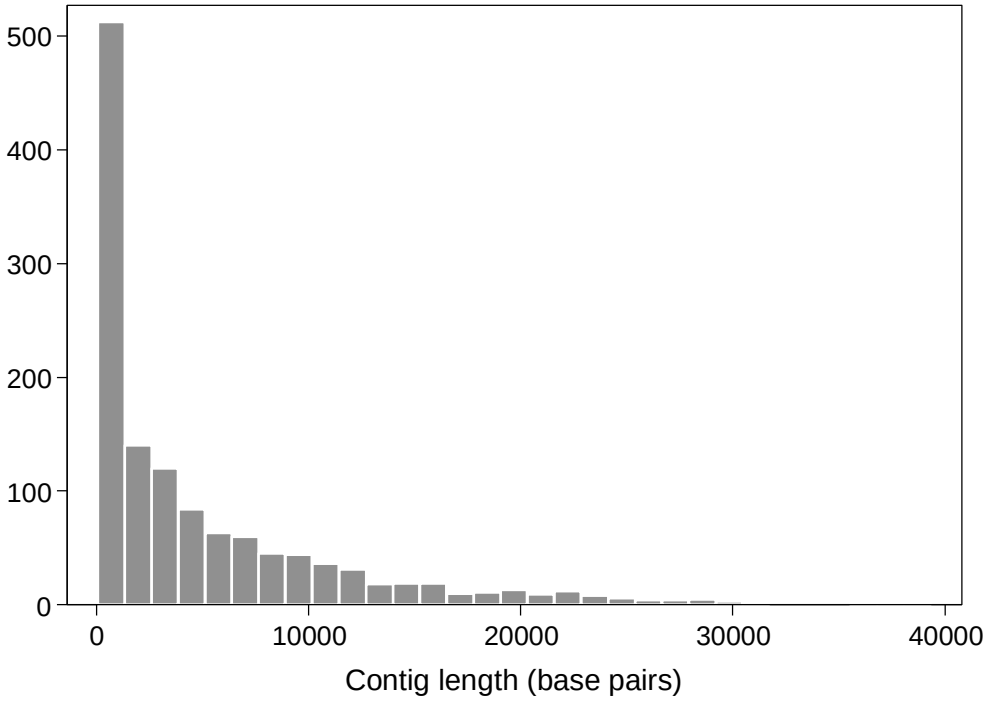
Figure continued over page



B

A: Percentage of reads containing at least one N, by tile number. B: Number of reads with N's in (y axis) per position in the read (x axis). Tiles are listed in numerical order from left to right. The scale is Y axis: 10^1-10^3 N's, X axis: base pair number 0-80.

Figure 3-7: Distribution of contig length from M. fortuitum sequencing

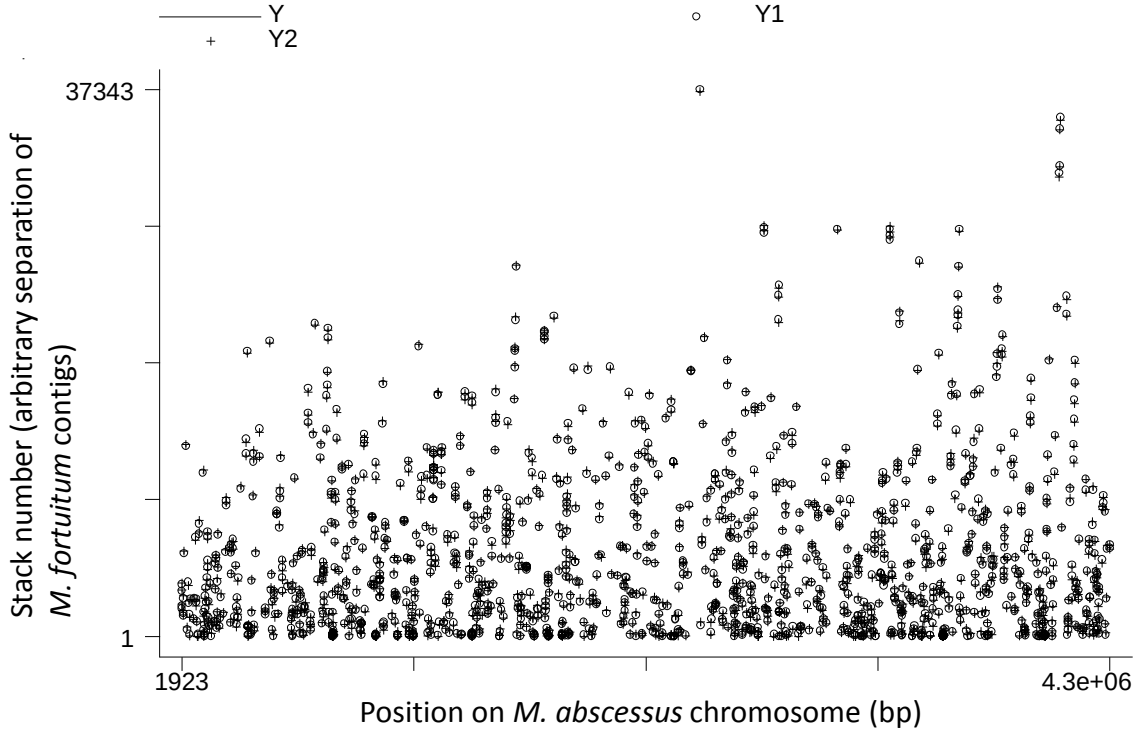


Histogram showing the distribution of contig lengths following sequence assembly using Velvet. There were 1043 contigs with an N50 of 15,810 bp, largest contig 39,343 bp.

It is only possible to accurately assess genome coverage by aligning contigs to a reference genome of the same species, and this cannot be done if there is no genome sequence information available for the species. Instead, the approximate coverage over the length of the genome was estimated by performing a nucleotide BLAST (BLASTN) on the genome sequence from the related fast growing mycobacterium *M. abscessus*, and noting the location on the *M. abscessus* chromosome of each hit from an *M. fortuitum* contig. It can be seen from Figure 3.8 that hits from contigs from the sequencing of *M. fortuitum* are equally distributed across the length of the genome of *M. abscessus.* Therefore, it was concluded that homologues of many *M. abscessus* genes were present in *M. fortuitum*, and that sequencing coverage appeared to be spread evenly across the whole *M. fortuitum* genome. BLAST searches of specific antigens (e.g. members of the Antigen 85 family) from different

mycobacterial species were also undertaken, and homologues were identified within the *M. fortuitum* contigs (data not shown).

Figure 3-8: Distribution of hits from M. fortuitum from on M. abscessus genome



Graph of location on the *M. abscessus* chromosome of hits from the *M. fortuitum* contigs following nucleotide BLAST of all *M. fortuitum* contigs, plotted using STATA (version 9, Statacorp LP, College Station, Tx). Demonstrates that the *M. fortuitum* contig hits are roughly equally spread out across the *M. abscessus* genome, suggesting good coverage of sequencing of the *M. fortuitum* genome.
Y1: start of hit on *M. abscessus* chromosome, Y2: end of hit on *M. abscessus* chromosome, Y: connecting line.

3.2.4 Genome annotation (open reading frame prediction)

For the un-annotated genomes, (*M. intracellulare, M. kansasii, M. africanum* and *M. fortuitum*) the Prodigal (v1.1) open reading frame prediction tool was used on all available contigs [135]. We opted to use this tool based on the reported good performance of the

tool in high GC content genomes such as mycobacteria. In a comparison of different annotation tools in experimentally validated gene sets, Prodigal predicted the 3' ends of genes in *M. tuberculosis* H37Rv with an accuracy of 100%, and the entire gene with an accuracy of 93.6% [135] (http://compbio.ornl.gov/prodigal/). On unannotated genomes, median predicted protein length was 243, 235, 275 and 185 amino acids for the *M. intracellulare*, *M. kansasii*, *M. africanum* and *M. fortuitum* genomes respectively (median contig lengths 7888, 9278, 98,860 and 2528 respectively); in total 25,130 protein sequences were predicted.

3.2.5 Choice of probable secreted proteins present in common NTM and absent from the *M. tuberculosis* complex

NCBI BLASTP was used to compare all protein sequences against themselves; 4048 proteins hit at least 4 proteins in group 1 genomes (NTM species of interest), and none in group 2 (*M. tuberculosis* complex), using the using NCBI default parameters including an expect value of 10 and word size of 3 [137].

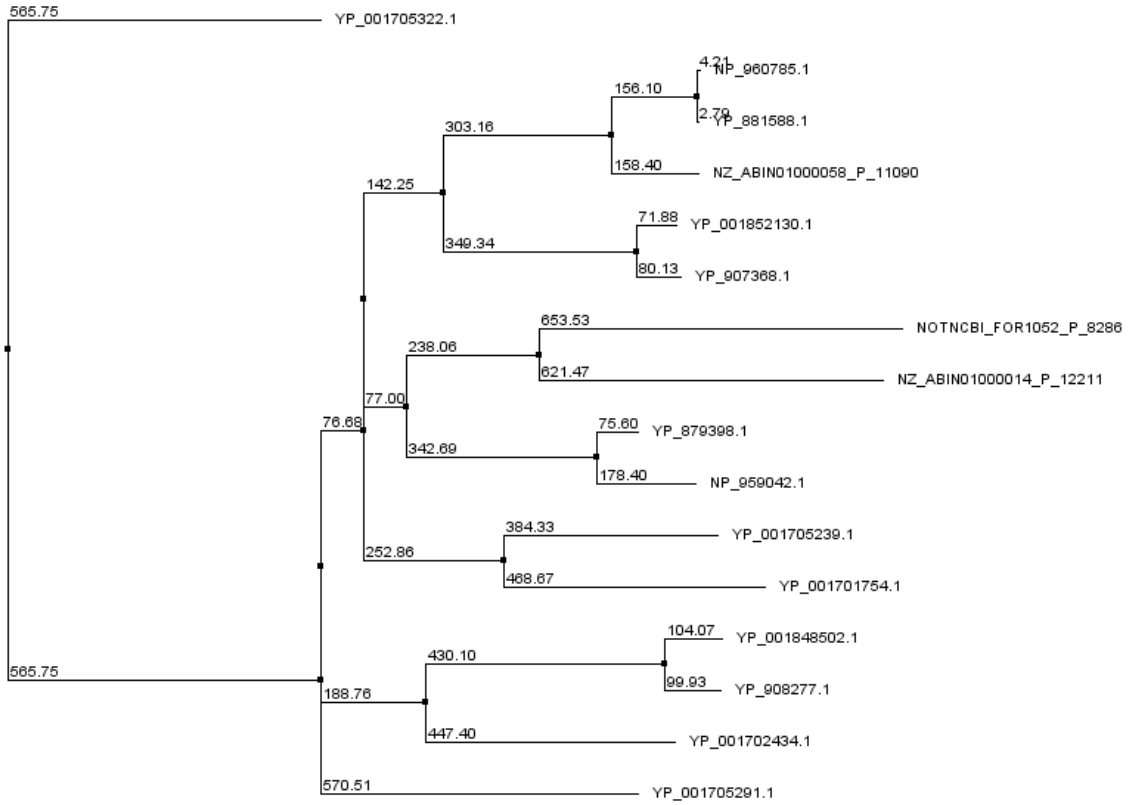
These 4048 proteins were selected, and all others discarded. To test that the pipeline was functioning correctly, we used the same methodology to compare the genomes of *M. tuberculosis* and BCG, and identify which proteins were present in *M. tuberculosis* but absent from BCG. As predicted, this resulted in the identification of proteins known to be present in *M. tuberculosis* but not in BCG, such as members of the EsxA gene family, including ESAT-6 and CFP-10 (data not shown).

3.2.6 Clustering

BLASTClust (NCBI) was used to assign these proteins into related families. BLASTClust uses a 'hard' (dichotomous) clustering method based on percentage identity of a particular protein length. Clustering cut-offs of $\geq$50% identity over $\geq$50% length were used. Clusters were then realigned using ClustalW and visualised in ClustalX (both European Molecular Biology Laboratory (EMBL)/ European Bioinformatics Institute (EBI)). Variations in percentage identity and percentage length of match were tried, and none produced such consistent arrangement in related families as the 50% identity/ 50% length values. During development, this method was initially tested on the Antigen 85 gene family (data not shown).

817 clusters were produced (an example is shown in Figure 3.9). As an additional specificity check, cluster members were compared against *M. tuberculosis* complex genomes using TBLASTN. Any clusters with significant homology were excluded. After this, only 78 clusters were left, varying in size from 2-17 proteins. The median size of clusters was 5 proteins (standard deviation (sd) 3.3). From these, 12 clusters were selected for experimental testing (Table 3.3). 11 were prioritised on the basis that they contained predominantly secreted proteins. One cluster (Mce family proteins, shown in Figure 3.9) was selected on the basis of experimental evidence of immunogenicity [146]. Median selected cluster size was 7 (sd 3.5).

Figure 3-9: Phylogenetic tree of the MCE family of proteins



Phylogenetic tree showing the relationships between the different Mce family proteins from different mycobacterial species. The tree was produced in Jalview (http://www.jalview.org/) with the Neigbour Joining (NJ) method, using BLOSUM 62. Numbers indicate the distances between the sequences, and are a sum of BLOSUM 62 scores for the residue pair at each aligned position.

YP_001705322.1: putative Mce family protein [Mycobacterium abscessus ATCC 19977], YP_881588.1: mce related protein [Mycobacterium avium 104], YP_001852130.1: Mce protein, Mce5A [Mycobacterium marinum M], YP_907368.1: Mce protein, Mce5A [Mycobacterium ulcerans Agy99], YP_879398.1: mce related protein [Mycobacterium avium 104], YP_001705239.1: putative Mce family protein [Mycobacterium abscessus ATCC 19977], YP_001701754.1: putative MCE family protein [Mycobacterium abscessus ATCC 19977], YP_001848502.1: MCE-family protein Mce6A [Mycobacterium marinum M], YP_908277.1: MCE-family protein Mce6A [Mycobacterium ulcerans Agy99], YP_001702434.1: putative Mce family protein [Mycobacterium abscessus ATCC 19977], YP_001705291.1: putative Mce family protein [Mycobacterium abscessus ATCC 19977]. Other numbers relate to protein sequences predicted using the programme Prodigal. NOTNCBI_FOR1052_P_8286: *M. fortuitum* predicted protein, NP_960785.1 and NP_959042.1: *M. intracellulare* predicted proteins, NZ_ABIN01000058_P_11090 and NZ_ABIN01000014_P_12211: *M. kansasii* predicted proteins

Table 3-3: Protein clusters

Cluster id	Protein family	No. proteins/ cluster	Predicted cellular location			
			Secreted	Lipid anchored	N-anchored	Intracellular
A	Mce family protein	16			12	1
B	hypothetical protein	13	7	1		1
C	hypothetical protein	12	4		2	
D	hypothetical protein	8	4		1	
E	hypothetical protein	8	4			
F	hypothetical protein	7	4			
G	hypothetical protein	7	4			
H	hypothetical proteins	7	3		1	
I	Tat-translocated enzyme	6	3	1		
J	27 kDa lipoprotein antigen	6	3			
K	hypothetical protein	6	3	3		
L	hypothetical protein	5	3	2		

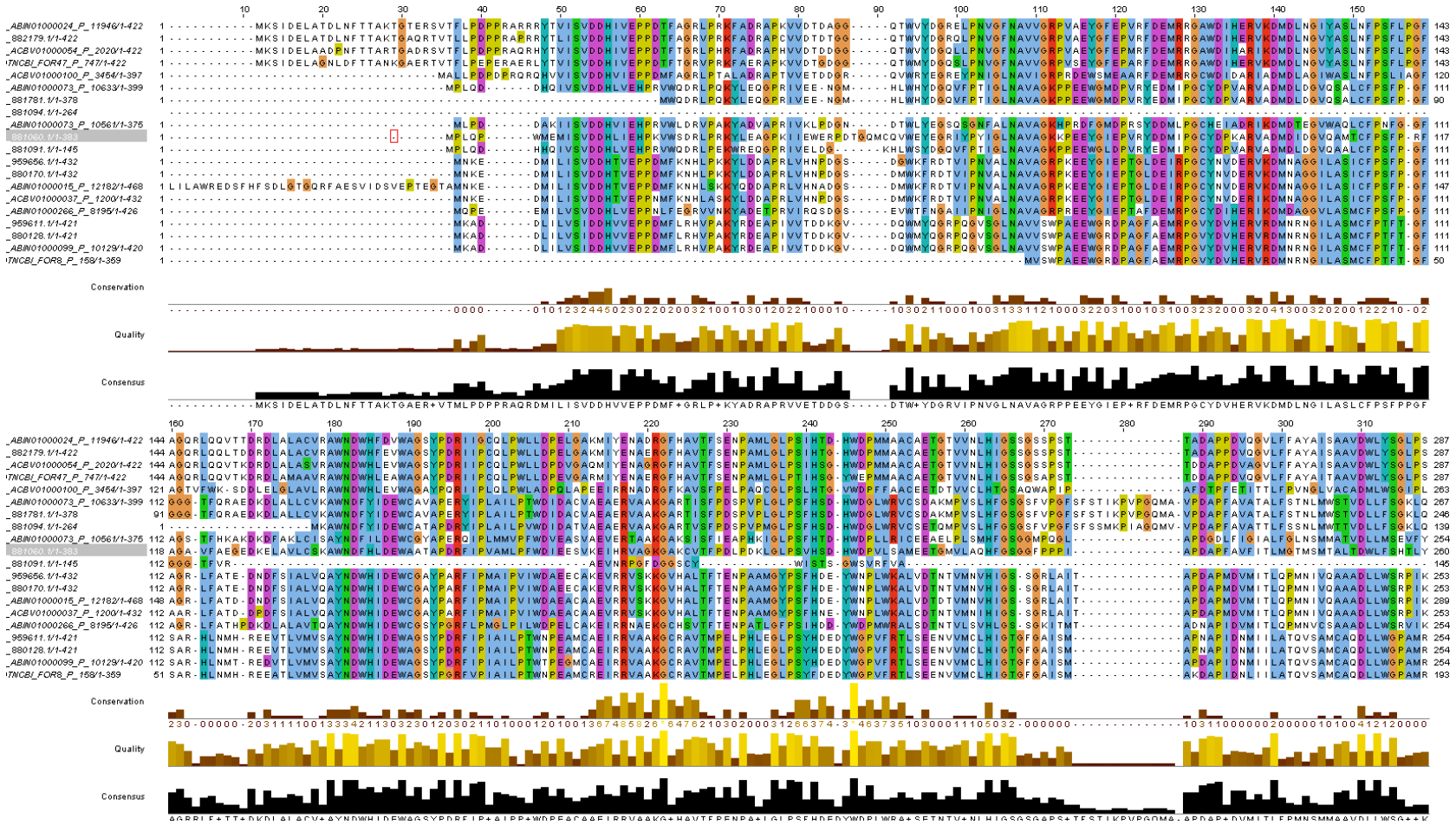
Summary of the clusters selected for the generation of peptides for experimental testing.
Cluster id: identification number allocated to cluster, Protein family: function of proteins within the cluster (if known), Predicted cellular location: predicted cellular location according to the LocateP database. Note, number of cellular location predictions may be less than the number of proteins in a cluster as the LocateP database only contains predictions to reference proteins, so the protein sequences generated by Prodigal will not be included in this process.

An example of a ClustalX alignment is shown in Figure 3.10, and the contribution of the different NTM species to the make-up of the clusters is shown in Figure 3.11. Out of the 6 NTM in the selection procedure, 50-100% species contributed to all clusters (Figure 3.11B). The predicted location of the proteins is shown in Figure 3.12. 98% of the selected proteins were predicted to be secreted or cell surface associated.

3.2.7 Selection of peptides

20-mer peptide sequences overlapping by 10 amino acids spanning the selected proteins were computationally defined. In order to exclude peptides with a high sequence similarity to common bacteria, BLASTP was performed on all peptide sequences against all published bacterial reference protein sequences, with BLAST parameters modified for short sequence length. 25,636 hits were generated (Figure 3.13). Not surprisingly, the majority of these hits were against the genus *Mycobacterium*, followed by hits against the related genera *Nocardia* and *Rhodococcus.*

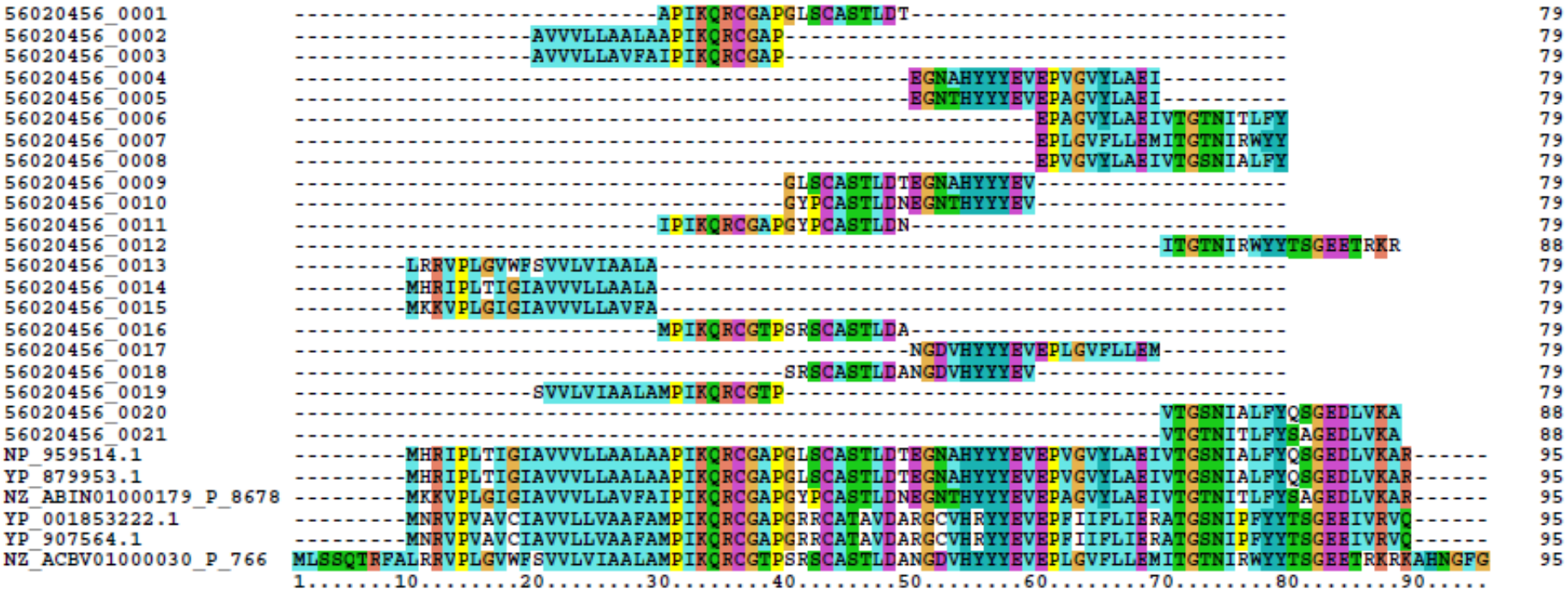
Figure 3-10: ClustalX alignment of protein clusters



A

Figure continued over page

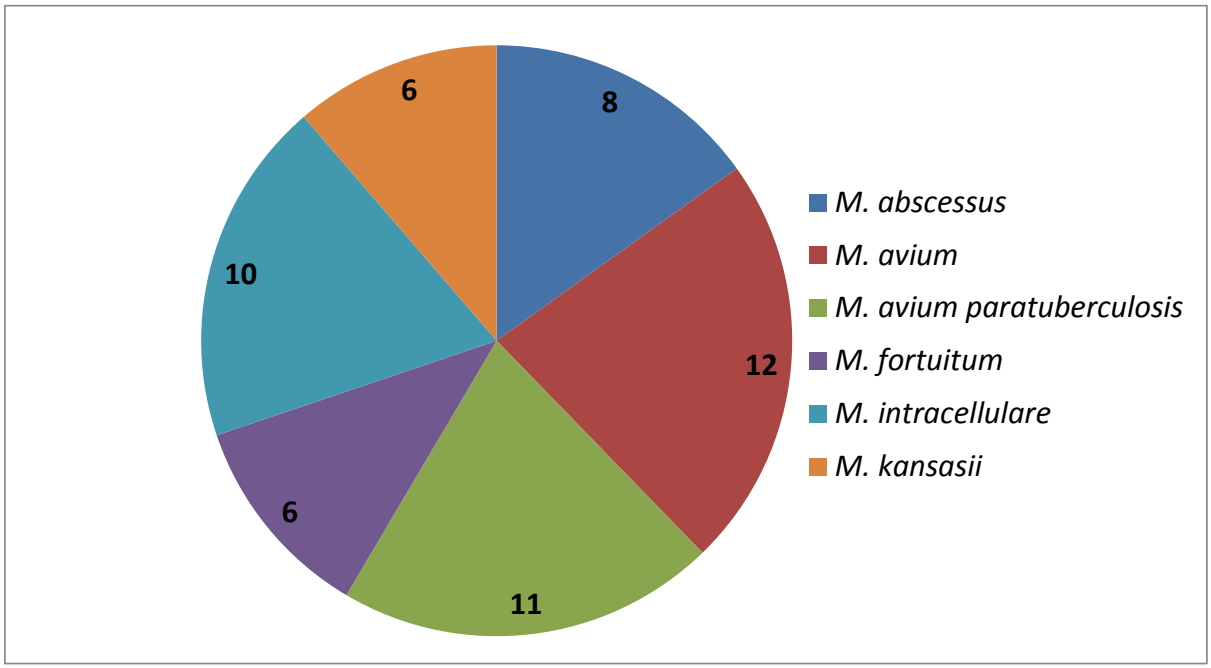
Figure 3.10 ctd.



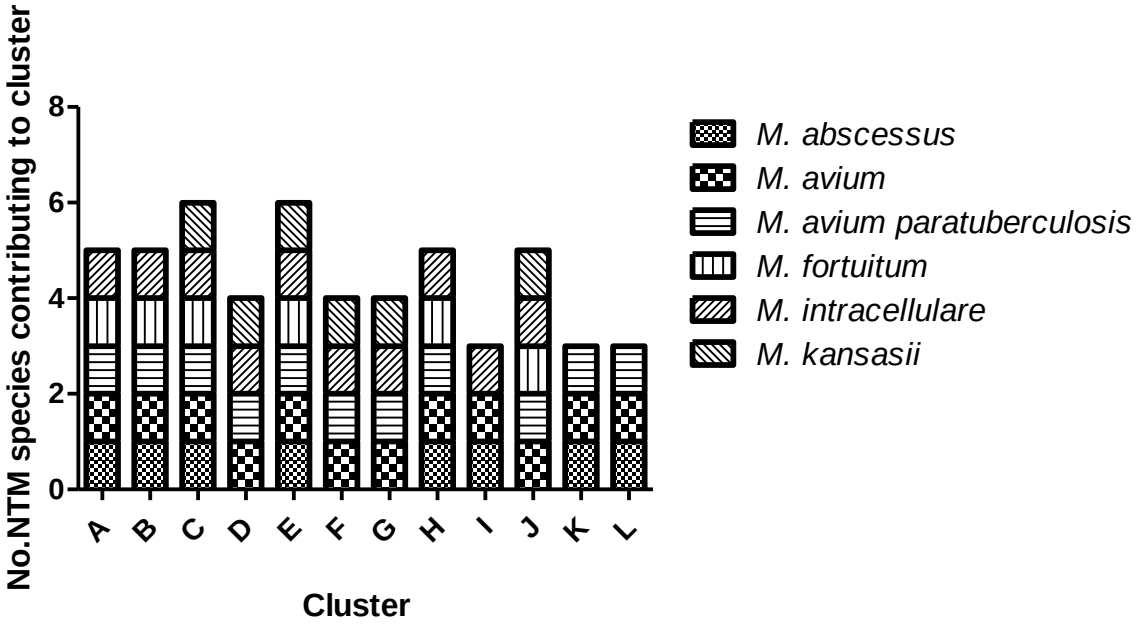
B

Two examples of clusters alignment.
A: the Mce cluster of proteins. Text down the left hand side: NCBI accession number or (for proteins predicted *ab initio* by the Prodigal tool) an internally generated identification number. Black histogram indicates the degree of consensus between the sequences. Amino acid colour codes represent different amino acid characteristics (eg acidic/ basic, hydrophilic/hydrophobic, etc).
B: A small protein cluster (Cluster G) with associated peptides: 20-mer peptides overlapping by 10 amino acids were selected to cover all variations in sequence in proteins within each cluster

Figure 3-11: Representation of different NTM species in clusters



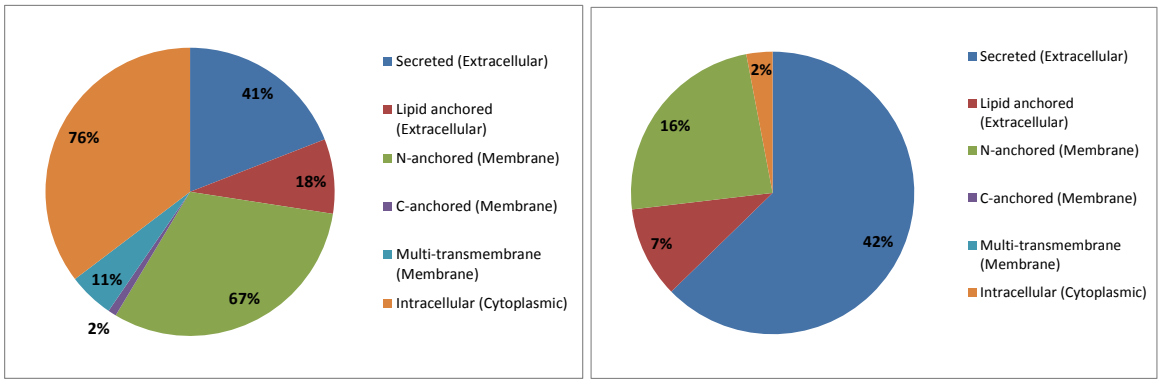
A



B

Representation of the different species of NTM within the clusters.
A: within all clusters; there is a roughly equal distribution of species from which proteins are derived, with a slight predominance of proteins originating from *M. avium* and *M. avium paratuberculosis.*
B: Representation of different NTM species shown for each cluster. Clusters contain proteins from 50-100% of species involved in the process.

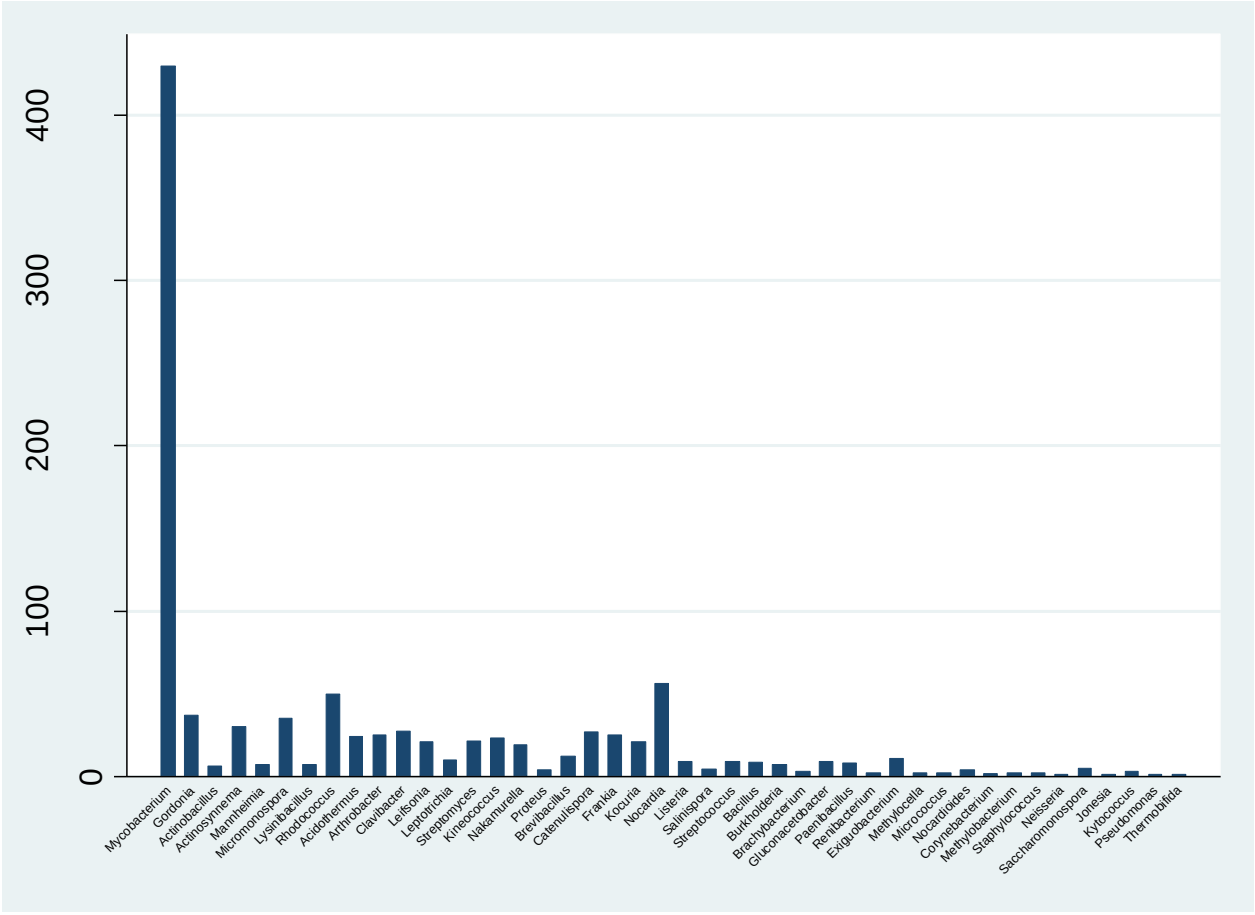
Figure 3-12: Predicted locations of clustered proteins



A **B**

A: Predicted locations of all proteins in clusters containing greater than one protein
B: Predicted locations of proteins in selected clusters only

Figure 3-13: Peptide 'hits' on bacterial reference proteins following BLASTP



Bacterial reference protein genus

BLASTP was carried out of all peptide sequences against all bacterial reference proteins. The resulting hits are shown by bacterial genus. Y axis: the median number of peptide hits across the different species within the genus

Peptides with any hits to non-mycobacterial predicted proteins under these BLAST conditions were eliminated from the process and 1699 peptides were synthesised. No additional, unexpected hits were obtained on manual entry of randomly selected proteins or peptides into the online NCBI BLAST server.

3.3 Discussion

A bioinformatic pipeline was constructed in order to compare the proteomes of selected common NTM with those of the *M. tuberculosis* complex, and with a high degree of specificity select proteins which were common to 4 species of NTM and not present in the *M. tuberculosis* complex. The possibility that any peptide sequence had unexpected similarity with peptide length sequences from the *M. tuberculosis* complex or from other bacterial species was then excluded. This methodology was designed to be stringent, aiming to prevent unexpected cross-reactivity at the stage of immunological testing.

3.3.1 Choice of NTM

Common NTM were selected based on studies both of clinical and environmental isolation from South Africa and the UK [121, 128-134]. Clinical isolation studies will identify micro-organisms which humans are, by definition, exposed to. However, in order to be isolated from clinical specimens, an organism must usually be able to replicate in the human host. Humans are probably exposed to a far wider range of micro-organisms than just those that can replicate in humans; hence, environmental isolation studies may be less biased. So a combination of both was used.

M. africanum is a common cause of TB in West Africa [147, 148] and was included so the results of the study would also be applicable to trials of MVA85A and other candidate TB

vaccines in this region. It has a high level of similarity to *M. tuberculosis* and causes TB in 11-51% cases in West Africa [148], although it is rare in South Africa [149].

3.3.2 Steps to ensure the specificity of the pipeline

A large part of the pipeline was based on different BLAST operations. BLAST is a heuristic programme (provided by NCBI) which produces local alignments of nucleotide, protein and translated nucleotide sequences, based on similarity and identity. It uses a 'word'-based approach, in which an initial search is made for a 'word' consisting of a certain number of amino acids or nucleotides. Once an identical sequence of this length is found, the programme attempts to extend the match in both directions, using both a gapped (if there are non-matching nucleotides or amino acids) and a gap-free approach. Having done this with all possible combinations of words, it then assembles this information for all queries and assigns a bit score, which indicates how good the alignment is, and an expect (e) value, which indicates the statistical probability that each particular hit arose as a result of chance (similar to a P value). This takes into account the alignment of similar and identical residues, and the presence of gaps between the sequences. Sequence 'similarity' refers to the number of amino acid or nucleotide substitutions present between two sequences, and is measured in terms of number of substitutions and the degree of difference in the translated sequence which results. An identical sequence is literally identical.

There were a number of features incorporated in this pipeline which aimed to ensure that none of the final peptides would cross-react with sequences present in any members of the *M. tuberculosis* complex. The first BLASTP step produced a subset of proteins which were present in 4 NTM taxa of interest but not in members of the *M. tuberculosis* complex. This is a step which has been used in many similar antigen discovery projects [150, 151].

Its accuracy, however, depends entirely on the accuracy of the genome annotation which produced the protein sequences which were deposited as reference sequences, and on which BLASTP is based. In the case of this thesis it also depends on the quality of the genome annotation which was performed using the tool Prodigal. Of note, the average predicted protein length from the *M. fortuitum* contigs was significantly lower than from the *M. intracellulare, M. kansasii* and *M. africanum* contigs, which is not surprising given the shorter length of the contigs. It is likely that some of these short contig lengths may have resulted in the prediction of incomplete or fragmented protein sequences.

Protein predictions resulting from genome annotations are not always experimentally validated, and the reliability varies depending on the tool used [152]. This is illustrated by a comparison of two strains of *M. tuberculosis* H37rv and CDC1551. A published comparison of these indicates that they differ by a few single nucleotide polymorphisms (SNPs) and one gene[153]; by contrast, *M. tuberculosis* CDC1551 has 4189 reference proteins (http://www.ncbi.nlm.nih.gov/sites/genome?Db=genome&Cmd=ShowDetailView&TermToSearch=181) whereas *M. tuberculosis* H37rv only has 3988 (http://www.ncbi.nlm.nih.gov/sites/genome?Db=genome&Cmd=ShowDetailView&TermToSearch=135). Junglbut *et al* described 5 genes present in *M. tuberculosis* H37*Rv* which had not been predicted by genome annotation [154]. Using BLASTP alone and relying on predicted protein sequences would risk identifying proteins which appeared to be unique to NTM, but which in fact also existed in members of the *M. tuberculosis* complex, but had not been correctly annotated as proteins. Interestingly, pipelines which have relied on BLASTP alone have suffered from significant unpredicted cross-reactivity (see later).

The second stage of the pipeline involved TBLASTN, which compares protein sequences against six frame translations of the target DNA sequences. This makes it independent of genome annotation, so it is not affected by inaccuracies in genome annotation. It is, however, computationally much more demanding; this particular process took an entire weekend to run.

The analysis of the selected proteins in clustered families of related proteins allowed rational choices to be made about very large numbers of proteins. The default clustering parameters of $\geq$50% identity over $\geq$50% length resulted in the identification of a number of multi-gene families, such as the Mce family, the DoxX and the 27kDa lipoprotein antigen. Clusters typically contained some proteins for which structure and function had been experimentally defined, some for which it had been predicted, and some where there was no knowledge available (so called 'hypothetical proteins'). Analysing families of proteins together in this way allowed assumptions about the structure and function of hypothetical proteins to be made.

The LocateP database was used to predict protein subcellular location. It should be borne in mind, however, that this protein location prediction was only available for reference proteins. This underlines the importance of the clustering process, as clusters contained a mixture of reference proteins, where this information was available, and predicted proteins, where it was not. The process of clustering therefore enabled assumptions to be made regarding protein location on all proteins within a cluster, as the selection or rejection of a cluster was performed based on the ratio of predictions for secretion versus other cellular locations.

Secreted proteins were preferentially selected, on the basis that this group of proteins is consistently associated with the induction of strong immune responses in studies of mycobacteria [155-158]. This criterion was based on the need to select a subset of proteins for functional testing. Not all mycobacterial antigens are secreted, but secreted proteins make up a large proportion of currently defined mycobacterial antigens. It is likely that several factors determine the immunogenicity of mycobacterial proteins after infection, including subcellular location. Others may include bacterial factors such as the abundance of the protein, posttranslational modifications, and also host factors such as the different major histocompatibility complex alleles within the antigen [151].

3.3.3 Other similar pipelines

Bioinformatic pipelines have been used for the purposes of identifying antigens for the diagnosis of *M. bovis* infection in cattle [151], and leprosy [159] and *M. ulcerans* [160] in humans. The *M. bovis* pipeline consisted of a genome BLAST step using the NCBI and Tuberculist servers, but did not incorporate any secondary BLAST such as TBLASTN or BLASTP of peptide sequences. The investigators discovered a significant degree of unpredicted cross-reactivity, with responses both in cattle infected with *M. bovis* and in those exposed to BCG, which may be partly explained by this. They went on to investigate cross-reactive peptides using BLASTP against proteins from the *M. tuberculosis* proteome, and found similar peptides in *M. tuberculosis* which explained the cross-reactive responses found. Similarly, in the *M. ulcerans* pipeline, 11 out of 34 protein clusters identified using BLASTClast were found on PCR to contain previously undetected homologues in strains of *M. marinum.* This finding was thought to have occurred because of the high degree of genetic variability between different strains of *M. marinum,* most of which have not been

sequenced. The leprosy pipeline compared the *M. leprae* genome sequence with the genome sequences of other published mycobacteria, and in addition protein sequences of *M. leprae* with publically available proteomic data, using FASTA and BLASTP. Recombinant proteins were recognised by T cells from patients infected with *M. tuberculosis* as well as by those with leprosy [150], again suggesting cross-reactivity.

3.3.4 Next generation sequencing

3.3.4.1 Illumina sequencing

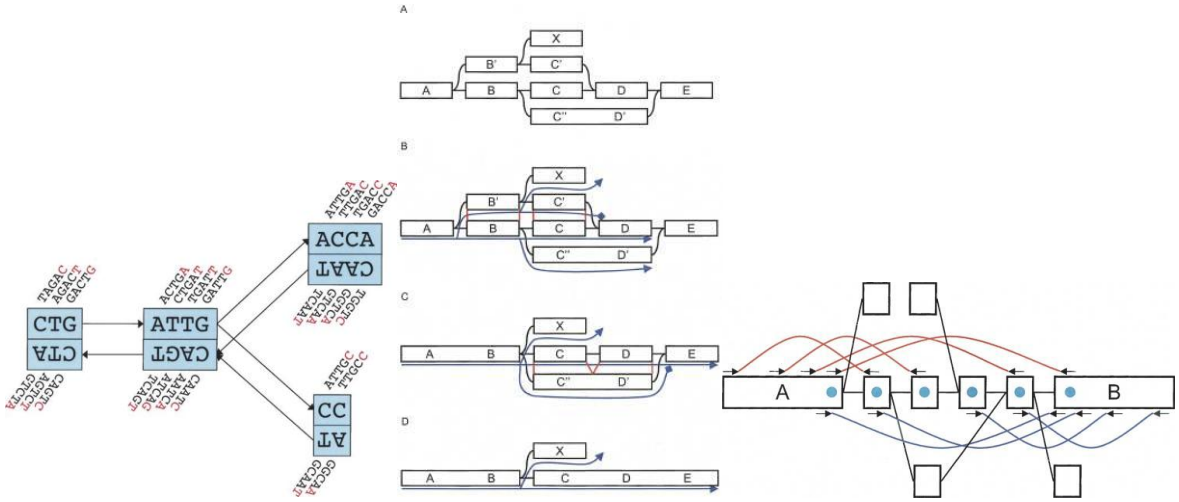
It was decided to perform next generation sequencing of the *M. fortuitum* genome as this organism is a fast-growing mycobacterium, and other than this species, *M. abscessus* was the only fast growing mycobacterial species included in the process. We chose to use Illumina GAIIx technology based on a number of factors, including the lower cost compared to the alternative (Roche 454 sequencing), and the fact that Illumina sequencing is better suited than Roche 454 to runs of G or C homo-polymers, which are characteristic of mycobacterial genomes.

Illumina sequencing is also termed 'sequencing by synthesis'. DNA is fragmented into specific sized fragments, which are ligated to adapters at both ends and which bind to immobilised primers. Each fragment is PCR-amplified to generate clusters of 'polonies' of fragments. Sequencing is carried out by sequential introduction of a mixture of all 4 nucleotides labelled with fluorescent reversible chain terminators. Highly sensitive cameras detect the colour of nucleotide incorporation at each polony location when the terminator is enzymatically cleaved. The output is read as a ratio of signal intensity of the greatest intensity signal in the region with the sum of the greatest and the second greatest intensity signals, which generates a quality score of the validity of the read [161].

3.3.4.2 Velvet sequence assembly

Traditional methods of genome sequence assembly (such as Atlas [162]) joined up series of overlapping reads, comparing differences between duplicate reads to eliminate error. Next generation sequencing generates vast quantities of very short reads of sequence, with too much redundancy to be able to use this method, so Velvet uses a different approach (Figure 3.14) [163, 164]. Series of short sequences of defined length, called k-mers, are progressively lined up with an overlap of all base pairs except one, so the sequence progressively elongates by one base pair at a time. This produces a 'node', consisting of the sequence of the final nucleotides from each k-mer, which is paired with its twin node, consisting of the reverse complement of the sequence, to produce a 'block'. Reads (known as 'roadmaps') are mapped as paths along nodes, and after the production of reads, various error correction methods are used to remove sequence errors and duplications. Duplicate sequences are merged, and various 'topological' sequence features (termed 'tips'- nodes which are disconnected at one end- and 'bubbles'- redundant duplications of sections of sequence, Figure 3.14B) are removed. On a test assembly of the genome of *S. suis*, 470 contigs (median size of 8661bp) were produced, with a coverage of 97% and an error rate of 0.02% [163].

Figure 3-14: Pictorial representation of Velvet sequence alignment



A B C

From: Velvet: Algorithms for de novo short read assembly using de Bruijn graphs, Zerbino DR *et al*, Genome Res. 2008 18: 821-829[163]

A: series of short k-mers joined into nodes, with the sequence of the node in the blue box, nodes joined with complementary nodes to produce blocks and blocks joined with arcs.

B: duplicate sequences are progressively merged and removed using the 'Tour Bus' algorithm

C: finally, the breadcrumb algorithm is used to join unconnected parts of sequence (A and B) based on connections, or 'arcs', between these and the various interconnecting sequences.

3.3.4.3 Discussion of choice of sequencing methodology

The main application of Illumina technology is in re-sequencing, or sequencing strains of an organism which has already been sequenced [163]. Then, the 'scaffold' sequence which already exists can be used as a structure on which to base the assembly of the new genome. The main alternative next generation sequencing technology, Roche 454, has also been used in this way. Roche 454 sequencing produces longer read lengths (about 250- 450 bp compared to about 36- 100bp in the case of Illumina at the time), so sequencing information may be easier to assemble. However, 454 sequencing has more difficulty in resolving long homopolymer repeats, which are problems in the CG-rich genomes of mycobacteria, and 454 sequencing data is less accurate, with more indels (insertion and deletion errors).

We chose to use single rather than paired end reads on the basis of the lower cost, but using paired end reads would probably have resulted in a better genome assembly. Paired end sequencing is when both the forward and reverse strands of a DNA fragment are sequenced in a single read. This has the advantage of providing known sequence distance information, allowing more easy sequence assembly. Velvet has additional parameters for assembling paired end reads (e.g. scaffolding an insert size, where absolute distances between sequences can be computed). Illumina compared assemblies of both the *E. coli* genome and human chromosome 20 using paired and single end reads. N50 contig size for *E. coli* was 132,476 bp (paired end) and 68,659 bp (single end reads) and for chromosome 20 was 70,744 bp (paired end) and 2,320 bp (single end reads). Our N50 contig length was 15,810 bp, which is relatively low. This is probably because of the high number of N's in our sequencing data.

Until recently, no genome had been sequenced and fully assembled *de novo* using only next generation techniques, since assembly of the multiple short reads without a genome to align them on had proved such a challenge. Najaran *et al* have recently used a novel strategy combining sequencing information from both Illumina and 454 sequencing of *Geobacter sulfurreducens* to assemble an entire genome *de novo* [165]. They combined information from both technologies at a very early stage of assembly, combining the advantages of the longer reads of the 454 and the greater depth of coverage and hence accuracy of the Illumina sequencing.

The original intention had been to sequence the genome of *M. gordonae* once the *M. fortuitum* genome had been completed, but in view of the difficulties that we encountered in the *M. fortuitum* sequencing project, and the development of a large back-log at the

Wellcome Trust sequencing facility which would have entailed considerable delay, it was decided to proceed without sequencing this organism.

3.3.5 Protein selection

Most of the proteins identified consisted of families of hypothetical proteins (Table 3.3), as is the case in other, similar bioinformatic pipelines [151, 166]. However, the Mce (mammalian cell entry) and 27kDa lipoprotein antigen families were identified, as well as a group of Tat (twin-arginine translocated) proteins. The Mce family of proteins are virulence factors which are involved in mycobacterial entry and survival in macrophages [167, 168]. Mce knockout mutants of *M. tuberculosis* have reduced virulence in mice [169], *M. bovis* Mce4 induces an inflammatory response in bovine macrophages [146], and antibodies have been demonstrated to Mce proteins in *M. tuberculosis* -infected humans [170]. The 27kDa lipoprotein antigen (or MPT51) is related to the Antigen 85 family [171], and is associated with humoral immune responses in humans [172]. Tat (twin-arginine translocated) proteins are secreted proteins exported via the Tat pathway; other than being secreted proteins, no further comment on their function can be made. Of course, the evidence of immunogenicity presented above related to proteins within the *M. tuberculosis* complex of mycobacteria. The NTM proteins defined through the pipeline will have very low levels of homology with the *M. tuberculosis* complex proteins (had they been similar, they would have been excluded during the process). However, it is still likely that these groups of proteins have similar functions in NTM compared to their equivalents in the *M. tuberculosis* complex.

3.4 Conclusion

In conclusion, antigens which were predicted to be present in common NTM and absent from members of the *M. tuberculosis* complex were defined. Those that were predicted to be secreted were selected, and those that had no detectable homology with other bacterial species at a peptide level were synthesised for experimental testing of immunogenicity.

4 Basic Immunology

4.1 Introduction

This chapter focuses on the screening of murine and human samples for T cell IFN-γ responses to pools of NTM-specific peptides using *ex vivo* ELISpot assay. This was initially done using splenocytes from mice infected with *M. avium* or BCG. subsequently PBMC samples from humans were screened, comparing those with a history of definite or probable exposure to NTM, cord blood samples which have an extremely low risk of NTM exposure, and healthy South African volunteers with a presumed high probability of NTM exposure.

The *ex vivo* gamma interferon ELISpot assay was chosen as the initial screening method for a number of reasons. Firstly, it is an efficient way to screen large numbers of pools of peptides, and is cheaper than alternative methods such as flow cytometry and cytometric bead arrays. Secondly, IFN-γ is considered to be an essential component in the immune response against *M. tuberculosis* [5], and is a stable and easily measured cytokine.

Screening of the peptide pools was initially performed using a mouse model in the hope that a subset of immunodominant antigens would be identified, allowing the testing of smaller numbers of peptides on the human samples. The mouse model consisted of a single intravenous administration of 10^5 CFU *M. avium* (clinical isolate, John Radcliffe Hospital, Oxford) or 10^5 CFU BCG (SSI strain) to C57Bl6 mice, followed by analysis of specific responses by IFN-γ ELISpot in splenocytes 6 weeks later. It was based on a murine model of *M. avium* exposure developed by Clare Sander, a previous DPhil student in Prof McShane's group ([173] Jenner Institute, Oxford University).

The human subjects consisted of three clinical groups; individuals from the UK and US from whom NTM had been isolated from sputum (considered definitely exposed and very low TB risk), healthy South Africans with a known negative response to QuantiFERON®-TB Gold In-Tube test (Cellestis) (considered probably NTM exposed, probably TB unexposed), healthy South Africans with a known positive response to QuantiFERON®-TB Gold In-Tube test (Cellestis) (considered probably NTM exposed, probably TB exposed), and cord blood samples from the UK (considered NTM and TB unexposed).

Here, optimisation of the mouse model and of the experimental methods relating to the murine and human testing of the NTM-specific peptide pools are discussed, and the results of these screening experiments.

4.2 Results

4.2.1 Mouse model of *M. avium* exposure

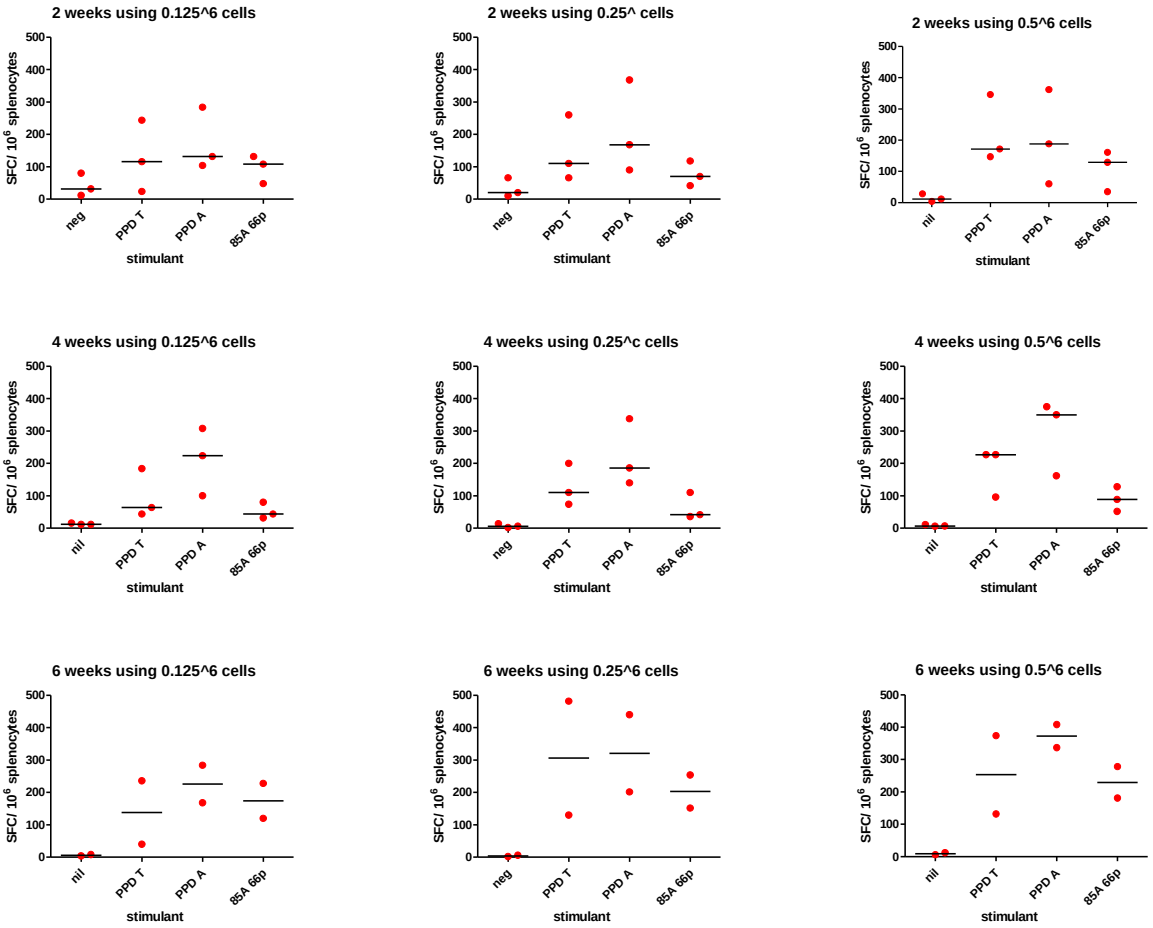
Prior to the arrival of the NTM-specific peptides, the kinetics of the murine immune response to intravenous *M. avium* infection was determined. This was done using various pools of peptides covering known highly immunogenic proteins from *M. tuberculosis* (ESAT-6, CFP-10, Antigen 85A and Antigens TB10.3 or TB10.4). Responses to PPD from *M. tuberculosis* (SSI) were also measured, although responses to the peptide pools were preferentially used in assay optimisation, since the NTM-specific proteins would also be tested as peptides.

4.2.1.1 Measuring splenocyte responses 6 weeks after i.v. exposure results in optimal detection of responses to Antigen 85A peptides

In order to determine the optimum timing of exposure to *M. avium,* mice were given a dose of $1x10^5$ CFU *M. avium* or of BCG intravenously, culled at 2, 4 and 6 weeks following intravenous exposure, and splenocyte responses to Antigen 85A and PPD from *M. tuberculosis* and *M. avium* were measured. In addition, whether the number of splenocytes plated onto the ELISpot plate influenced the final read-out of SFC per million PBMC was assessed by plating $0.125x10^6$, $0.25x10^6$ and $0.5x10^6$ cells for each time point.

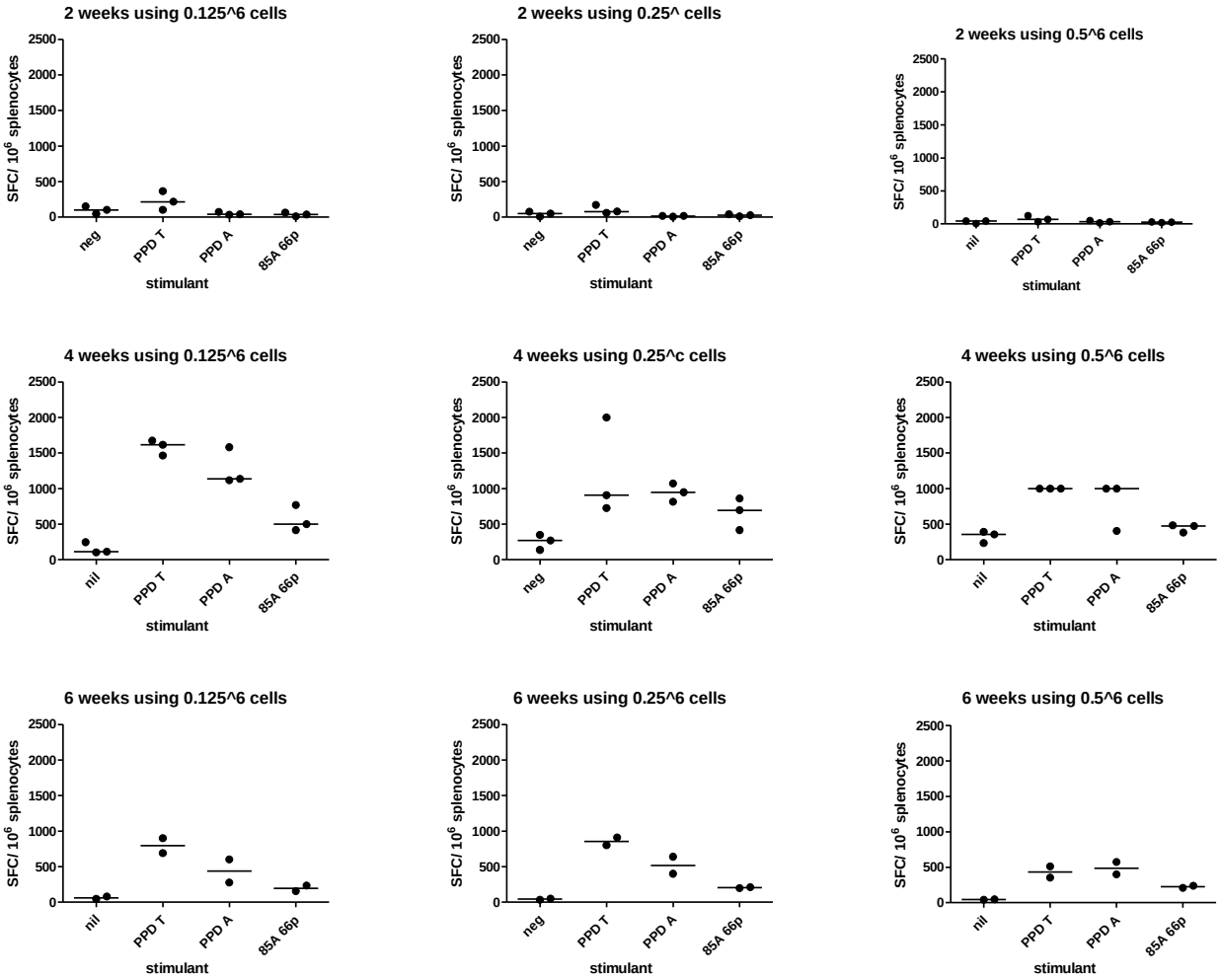
The results from this experiment are shown in Figure 4.1. There were greater responses to all antigens in the BCG-exposed mice compared to the *M. avium*-exposed mice. Focusing on the *M. avium* exposed mice, the greatest responses to the pool of Antigen 85A peptides were seen in the mice which had been exposed 6 weeks previously, although this difference did not reach statistical significance (Mann Witney test, comparing Antigen 85A response at week 6 with Antigen 85A response at each of weeks 2 and 4). This time point was selected for use (though it was noted that there was a greater response at 4 weeks in the BCG-exposed mice). There was no difference in results depending on the number of splenocytes plated on the ELISpot plate; therefore $0.5x10^6$ cells per well was selected for future experiments.

Figure 4-1: Investigating the kinetics of the IFN-γ response following *M. avium* or BCG infection, and comparison of number of splenocytes plated onto an ELISpot plate



A

Figure continued over page



B

Comparison of immune responses induced by i.v. *M. avium* or BCG; A: *M. avium*-exposed mice, B: BCG-exposed mice. There were three mice per group in all groups except for at 6 weeks, in which there were two. The greatest response to Antigen 85A peptides in *M. avium* exposed mice was seen 6 weeks after *M. avium* exposure (differences not significant, P>0.05, Mann Witney test).

4.2.1.2 Low IFN-γ responses are seen to NTM-specific peptides in the mouse model of M. avium exposure

Pools of overlapping NTM-specific peptides were created, where possible ensuring that no more than one cluster of related proteins was represented in a single pool of peptides (Table 4.1). There were 43 pools in total, containing a median of 39 peptides (range 33-49).

Splenocyte IFN-γ responses to *M. tuberculosis* PPD (PPD-T), *M. avium* PPD (PPD-A), a pool of overlapping peptides covering *M. tuberculosis* Antigen 85A and pools of NTM-specific peptides were analysed by overnight *ex vivo* ELISpot assay. In both *M. avium* and BCG-infected mice, strong responses to PPD-T and PPD-A were seen, as well as to pooled Antigen 85A peptides. For example, median SFC response to PPD-T in mice infected with *M. avium* was 492 SFU/10^6 PBMC. There were, however, no statistically significant responses to any of the NTM-specific peptide pools either in the mice infected with *M. avium* or in those infected with BCG ($P>0.05$, Wilcoxon matched pairs test). Results are summarised in Figure 4.2. Two out of five *M. avium*-infected mice made responses (of 83 and 369 SFC/ million PBMC) to pool 16, which corresponded to cluster 0D, a cluster of hypothetical proteins. Responses to pools 41 and 42 were of borderline statistical significance (pool 41, $P=0.0625$, pool 42, $P=0.0579$, Wilcoxon matched pairs test, comparing response to peptide pool with response in unstimulated well) in the BCG-infected mice. These pools covered clusters 1A-1L, which contained proteins from all of the pools which had hit proteins from the *M. tuberculosis* family with a low affinity. It was therefore concluded that these proteins were likely to be less specific than those in clusters where there were no 'hits' with the *M. tuberculosis* family (with the prefix 0), and these pools were not investigated further.

Table 4-1: Arrangement of peptides from different clusters into peptide pools

Cluster id	Protein family	No. proteins/ cluster	Predicted cellular location				Peptide pool no.
			Secreted	Lipid anchored	N-anchored	Intracellular	
0A	Mce family protein	16			12	1	1-8
0B	hypothetical protein	13	7	1		1	9-12
0C	hypothetical protein	12	4		2		13-15
0D	hypothetical protein	8	4		1		16-17
0E	hypothetical protein	8	4				18-19
0F	hypothetical protein	7	4				20-22
0G	hypothetical protein	7	4				23-24
0H	hypothetical proteins	7	3		1		25
0I	Tat-translocated enzyme	6	3	1			26-29
0J	27 kDa lipoprotein antigen	6	3				30-31
0K	hypothetical protein	6	3	3			32
0L	hypothetical protein	5	3	2			32

Table continued over page

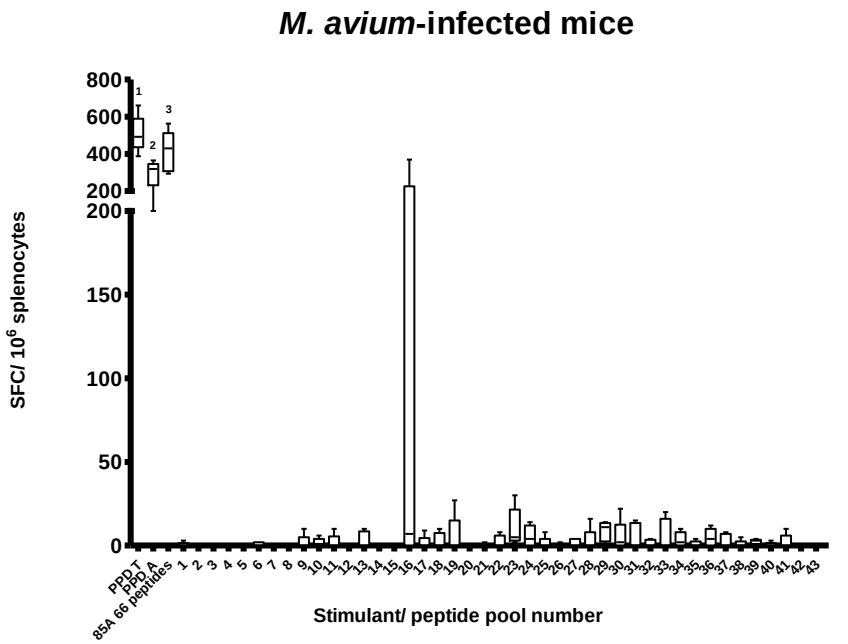
Table 4 1 ctd

Cluster id	Protein family	No. proteins/ cluster	Predicted cellular location				Peptide pool no.
			Secreted	Lipid anchored	N-anchored	Intracellular	
1A-1L	see above						39,41,42
2A-2L	see above						33-38
3A-3L	see above						43

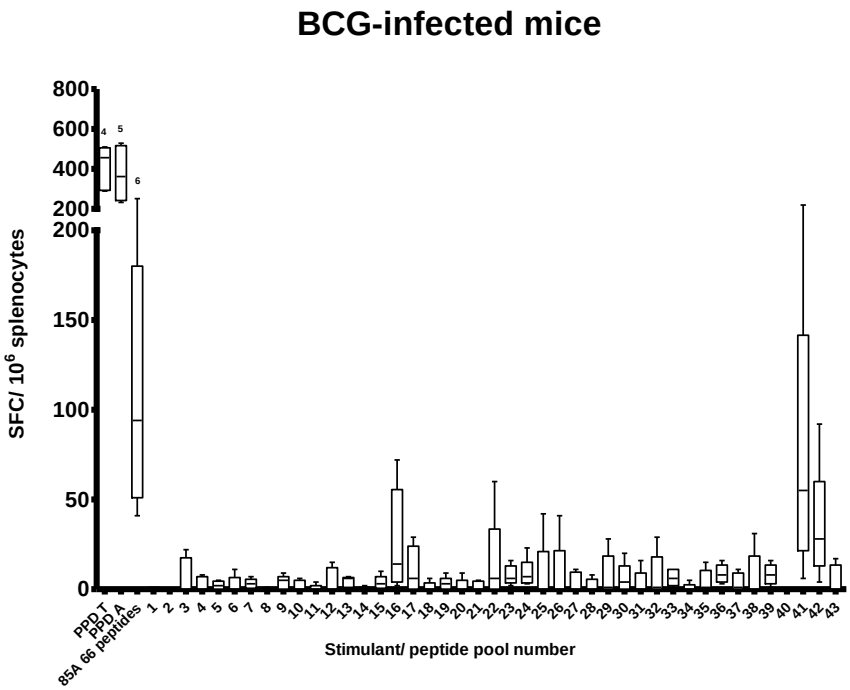
Summary of the clusters selected for the generation of peptides for experimental testing, and arrangement in pools for testing on human samples.

Cluster id: identification number allocated to cluster, protein family: predicted function of proteins within the cluster, predicted cellular location: predicted cellular location according to the LocateP database, peptide pool no.: the number of the peptide pool which clusters of proteins were tested in. Peptide pools 33-43 consisted of peptides which hit bacterial reference proteins with a low affinity on Protein BLAST. These peptides came from all clusters, and were tested separately from the others as there was a concern they may be less specific. Note, number of protein location predictions is not always equal to number of proteins in a cluster as the LocateP database only contains predictions to reference proteins; the proteins annotated using Prodigal will not have associated location predictions.

Figure 4-2: Responses to pools of NTM-specific peptides in mice infected with *M. avium* or BCG



A



B

Ex vivo ELISpot responses to NTM-specific peptide pools in mice infected with either *M. avium* (A) or BCG (B). Significant responses were seen to the positive control antigens PPD-T, PPD-A and Antigen 85A peptide pool, but no responses to NTM-specific peptides were statistically significant. Box and whisker plot showing medians and range of responses. 1: P=0.0079, 2: P=0.079, 3: P=0.0079, 4: P=0.0079, 5: 0.0079, 6: 0.0160, all Wilcoxon matched pairs test, comparing response to control antigen with unstimulated well. There were five mice per group.

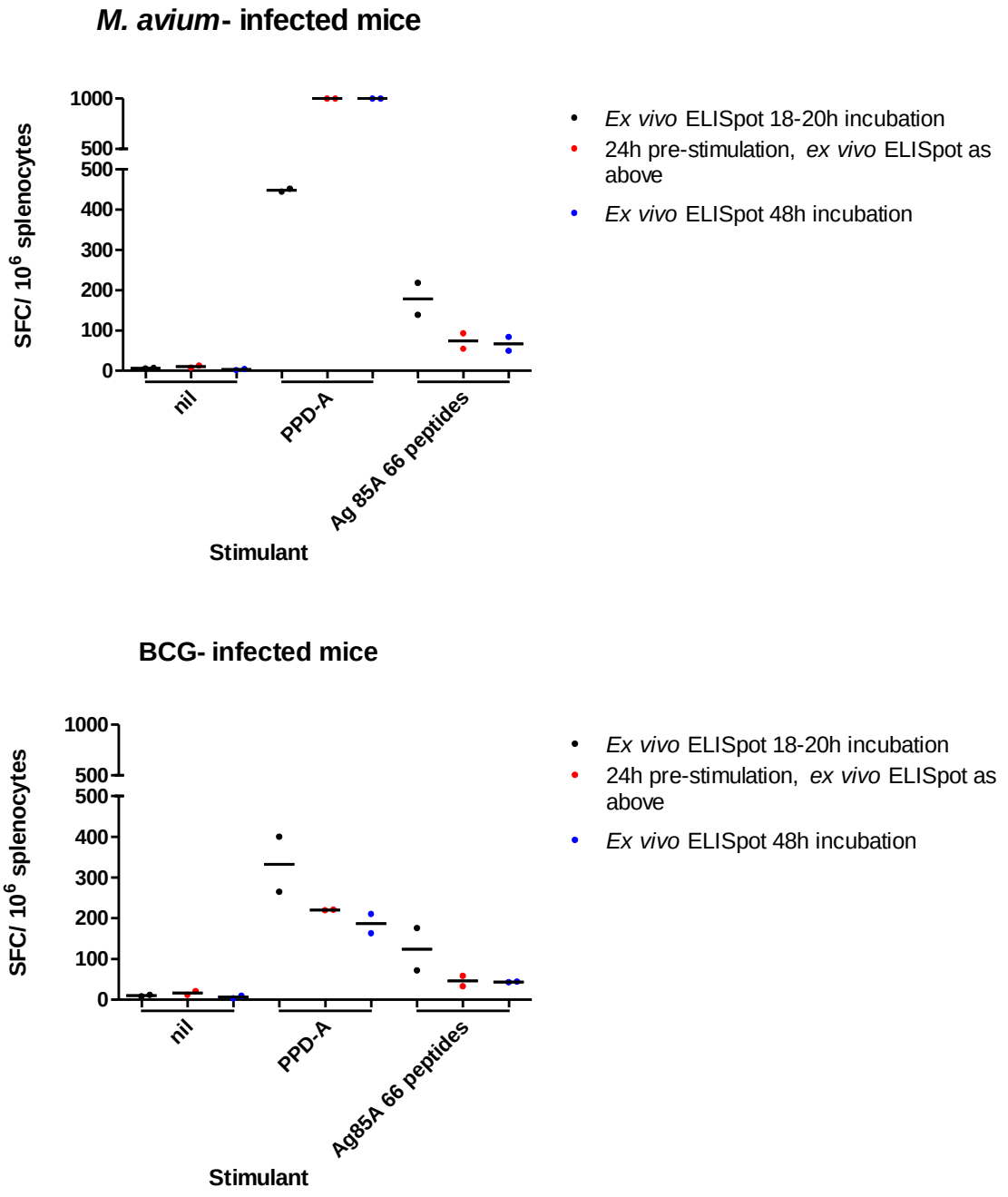
4.2.1.3 Neither pre-stimulation nor increased ELISpot incubation time increases sensitivity of IFN-γ detection in mice

A possible explanation for the low responses to NTM-specific peptides is that the *ex vivo* ELISpot assay was insufficiently sensitive to detect them. This seems unlikely given the robust responses which were detected to the Antigen 85A peptide pool and PPD, but it was decided to undertake a brief comparison of modifications to the ELISpot assay that could potentially increase its sensitivity.

Smith *et al* compared differences between ELISpot protocols currently in common use in four different laboratories performing TB vaccine trials, including one assay with a 'pre-stimulation step', in which PBMC were incubated with antigen for 24 hours before plating on the ELISpot plate [174]. This pre-stimulation step increased ELISpot responses to PPD, but the effect on responses to peptides was variable. A second protocol in use at one of the laboratories surveyed (which he did not investigate) involved incubation of PBMC and antigen on the ELISpot plate for 48 hours rather than for 18-20 hours.

A comparison of (1) standard ELISpot, (2) 24 hour pre-stimulation in a 96 well U bottomed plate followed by standard ELISpot and (3) 48 hour ELISpot was undertaken, with a limited number of replicates (Figure 4.3). Responses to PPD-A and a pool of Antigen 85A peptides were compared with media only.

Figure 4-3: Comparison of pre-stimulation and 48 hour ELISpot incubation with standard ex vivo ELISpot



Comparison of *ex vivo* ELISpot protocol modifications with standard *ex vivo* ELISpot, using *M. avium* and BCG-infected mice. Responses to PPD-A and to a pool of overlapping peptides covering *M. tuberculosis* Antigen 85A were compared with media only.

An increase in response to PPD-A was seen in the *M. avium*–exposed mice following both protocol modifications, but this was not also seen in the BCG-exposed mice. A decrease in the response to peptide pools was seen in both groups of mice with both protocol modifications. Since the protocol modifications did not appear to increase responses to peptides, it was decided to proceed directly to peptide screening on samples from human subjects, the target species.

As a result of the lack of responses to NTM-specific peptides in the mouse model, it was not possible to select a subset of peptides for further testing on the human samples. Instead, larger peptide pools were made in order to be able to test all the peptides on the limited number of PBMC available.

4.2.2 Volunteer recruitment

Four groups of individuals were recruited. 14 individuals who met the selection criteria for NTM exposure and low TB risk were recruited from the chest clinic at the Churchill Hospital, Oxford, and PBMC were isolated and cryopreserved. In addition, cryopreserved PBMC were donated by Oxford Immunotec, from individuals recruited from Portland VA Medical Center, United States. Inclusion and exclusion criteria were the same (see Section 2.7.2, Chapter 2, Methods). Samples from the UK cohort were evaluated in this chapter, for responses to NTM-specific peptides measured by *ex vivo* ELISpot. Samples from both the UK and the US cohorts were used in the proliferation experiments and in the multiplex cytokine bead arrays, described in Chapter 5 (Further Immunology). Table 4.2 summarises the clinical characteristics of the volunteers.

Table 4-2: Clinical characteristics of individuals exposed to NTM

No.	Age/ years	Sex	Country	Relevant diagnoses	Organism isolated	Treated?	Steroids?
1	21	M	UK	Cystic fibrosis	*M. kansasii, M. abscessus*	Y[a]	Y
2	41	F	UK	Bronchiectasis	*M. fortuitum*	N	Y
3	22	F	UK	Cystic fibrosis	*M. chelonae, M. abscessus*	Y[a]	Y[b]
4	88	M	UK	Bronchiectasis	*M. gordonae*	N	N
5	28	M	UK	Cystic fibrosis	*M. chelonae*	Y	N
6	59	F	UK	Bronchiectasis	*M. chelonae, M. xenopi*	N	N
7	28	M	UK	Cystic fibrosis	*M. avium*	N	N
8	37	F	UK	Cystic fibrosis	*M. abscessus, M. chelonae*	N	N
9	72	M	UK	Bronchiectasis	*M. chelonae, M. gordonae, M. fortuitum*	N	Y[c]
10	72	F	UK	Bronchiectasis	*M. intracellulare*	N	N
11	57	F	UK	Bronchiectasis	*M. avium*	Y	N
12	75	M	UK	COPD[d]	*M. malmoense*	N	Y
13	27	M	UK	Bronchiectasis	*M. intracellulare*	N	N
14	76	M	UK	Bronchiectasis	*M. avium*	Y	N
MB10	76	F	US	Bronchiectasis	*M. avium intracellulare*	Y	N
MB11	65	F	US	Bronchiectasis	*M. avium intracellulare*	Y	Y
MB16	64	F	US	Bronchiectasis	*M. avium intracellulare*	Y	N
MB40	78	F	US	Corticosteroid use	*M. avium intracellulare*	N	Y
MB41	81	F	US	Bronchiectasis	*M. avium intracellulare*	N	N

Summary of the clinical characteristics of individuals exposed to NTM. M= male, F= female, Y = yes, N = no, Steroids = individual taking steroids currently or within past 6 months. a. low level of adherence to prescribed treatment, b. very recent steroid therapy (2 doses only), c. very low dose (1mg prednisolone per day), d. chronic obstructive pulmonary disease.

Cord blood was collected from 12 healthy neonates, and cord blood cells were isolated and cryopreserved.

9 healthy South Africans with previous negative responses to QuantiFERON®-TB Gold In-Tube were recruited, and 7 healthy South Africans with known positive responses to QuantiFERON®-TB Gold In-Tube. PBMC from these individuals were used fresh in *ex vivo* ELISpots.

4.2.3 A pre-stimulation step prior to *ex vivo* ELISpot does not result in the detection of enhanced IFN-γ responses in humans

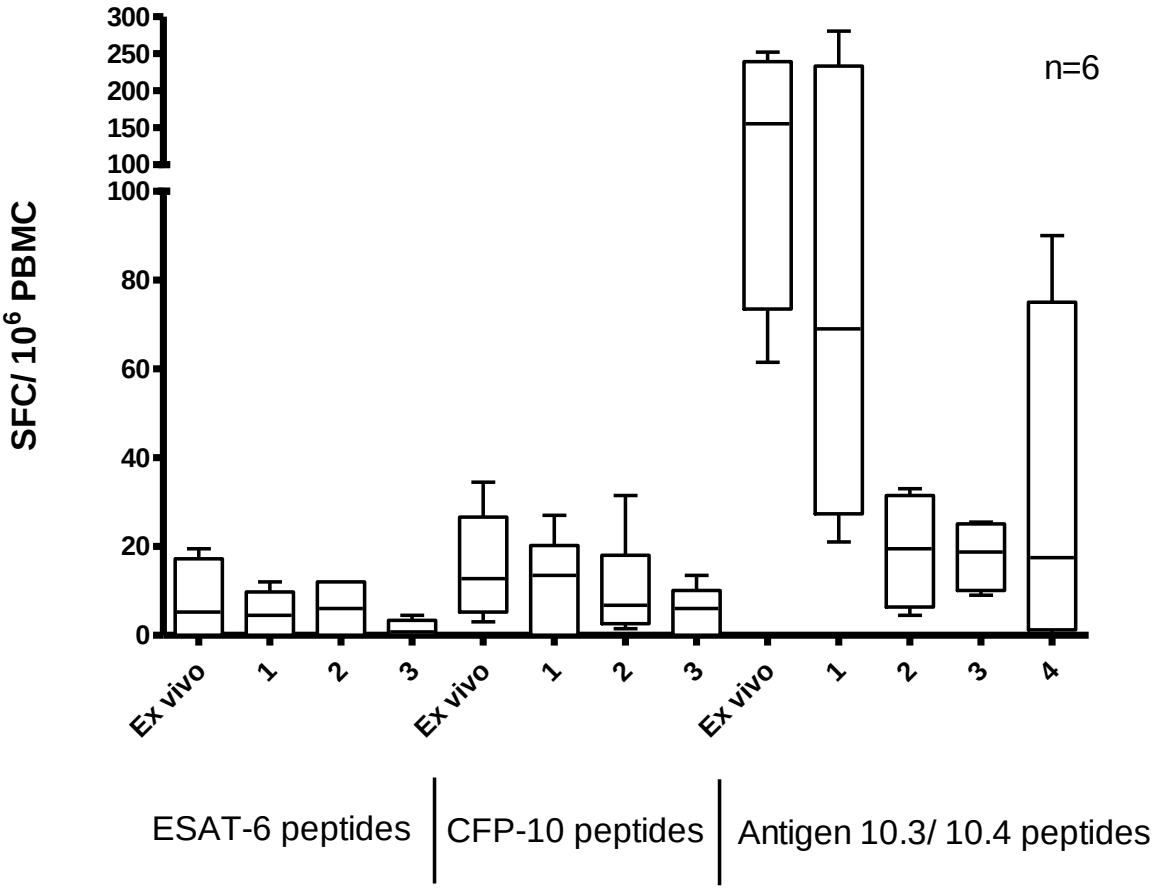
Given the lack of immune responses to NTM-specific peptides detected in the mouse model, it was decided to evaluate a pre-stimulation step in the human ELISpot protocol prior to testing the NTM-specific peptides. It was hoped that this might amplify the magnitude of responses detected.

Four different pre-stimulation protocols were compared, as outlined in Section 2.8.3.5, (Chapter 2, Methods). They all included a 24 hour pre-stimulation step in which PBMC were incubated with antigens at 37°C (Figure 4.4). Protocol one consisted of 24 hours pre-stimulation with transfer of PBMC plus supernatant to ELISpot plate, protocol two 24 hours pre-stimulation and transfer of washed PBMC with the addition of antigen, protocol three 24 hours pre-stimulation and transfer of washed PBMC with no additional antigen and protocol four 24 hours pre-stimulation in polypropylene tubes with transfer of washed PBMC alone, no antigen added. The washing steps were included in protocols two to four since it was noted in a preliminary experiment comparing the standard *ex vivo* ELISpot with protocol two, that where there were moderate or large responses in the *ex vivo* ELISpot, a

'black out' appearance resulted in the corresponding wells in the pre-stimulation ELISpot (example shown later, Figure 4.5). It was concluded that this was most likely due to the presence of IFN-γ in the pre-stimulation culture supernatant which bound diffusely to the ELISpot membrane. Therefore, modifications were incorporated in which PBMC were washed to remove cytokines present in the supernatant. In case of loss of antigen during this washing step, another modification included the addition of antigens together with washed PBMC to the ELISpot plate (which could be considered a re-stimulation).

The results of this comparison are shown in Figure 4.4. None of the differences were statistically significant compared to the standard *ex vivo* ELISpot ($P>0.05$, Wilcoxon matched pairs test), but there was a non-significant decrease in response in all of the protocol modifications compared to the standard *ex vivo* ELISpot technique.

Figure 4-4: Comparison of the ability of different short term culture methods to increase the sensitivity of ex vivo ELISpot

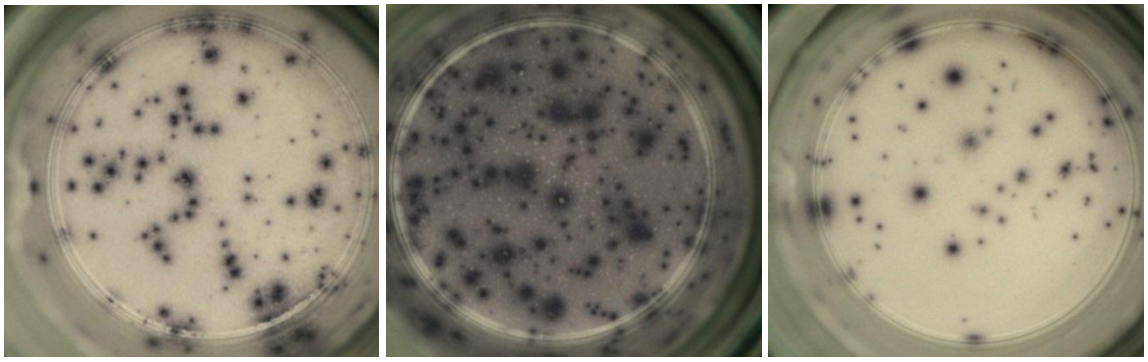


Four methods were compared: *Ex vivo:* standard 18-20 hour *ex vivo* ELISpot, 1: 24h pre-stimulation with transfer of PBMC plus supernatant to ELISpot plate, 2: 24h pre-stimulation in 96 well plate and transfer of washed PBMC with the addition of peptides, 3: 24h pre-stimulation and transfer of washed PBMC alone, no peptides added and 4: 24h pre-stimulation in FACS tubes with transfer of washed PBMC alone, no peptides added. Differences were not statistically significant (Wilcoxon matched pairs test). Box and whisker plot showing medians, interquartile range and range of responses.

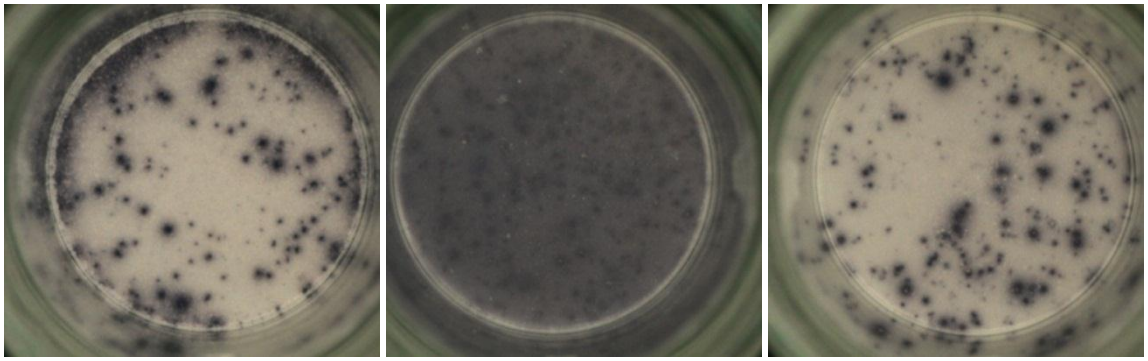
As described above, where a moderate response was detected in the *ex vivo* ELISpot experiment, the equivalent wells in the pre-stimulation experiment (protocol one) had a 'black out' appearance, making it hard for the automated ELISpot counter to differentiate spots from background. Examples are shown in Figure 4.5.

It was concluded that none of the modifications on the standard *ex vivo* ELISpot protocol resulted in improved detection of responses.

Figure 4-5: Appearance of ELISpot wells after modifications on the standard ex vivo ELISpot protocol



A B C



D E F

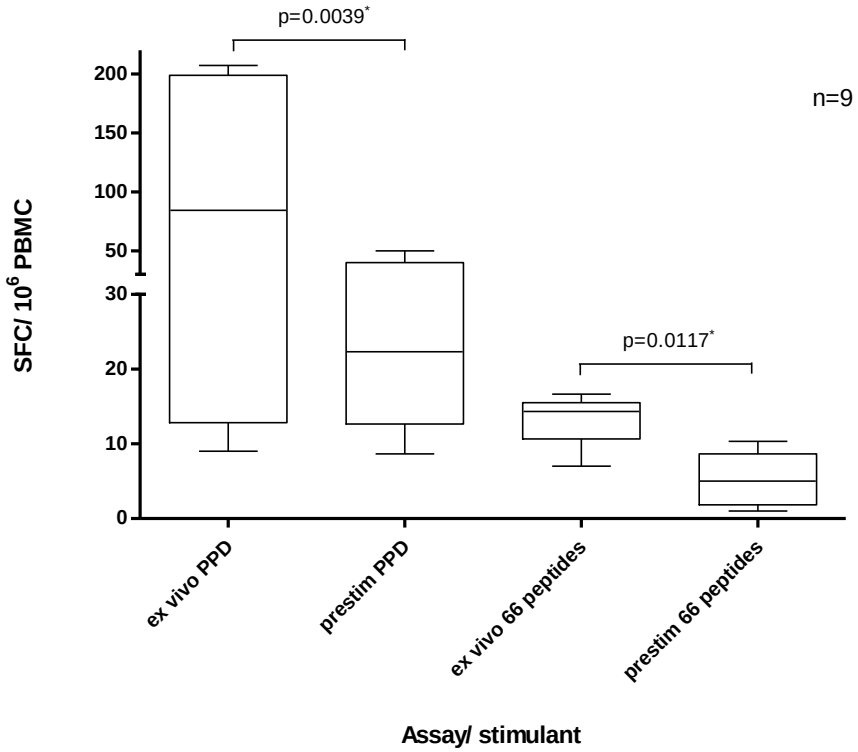
The appearance of ELISpot wells in the standard *ex vivo* ELISpot and in two of the modifications.24 hour pre-stimulation without PBMC washing resulted in a black out appearance (B, E), and washing the PBMC prior to plating on the ELISpot plate resulted in an appearance similar to the original *ex vivo* ELISpot (washing: C, F, *ex vivo* ELISpot: A, D).

4.2.4 A pre-stimulation step in the ex vivo ELISpot protocol does not result in the detection of enhanced responses using cryopreserved cells

A comparison of standard *ex vivo* ELISpot and protocol three, which involved a pre-stimulation step followed by PBMC washing, was undertaken on thawed cryopreserved

PBMC (Figure 4.6). This was done because the NTM-specific peptides would be screened on cryopreserved cells, and incorporating a pre-stimulation step might result in a lower level of background signal, which can occur when using cryopreserved cells due to the presence of dead cells [174]. Protocol three was selected because of the problems with 'black out' wells using protocol one, and there was no appreciable difference between re-stimulating with peptides (protocol two) or not (protocol three). There was a significant reduction in the responses detected, both to PPD-T (P=0.0039) and to a pool of peptides covering *M. tuberculosis* Antigen 85A (P=0.0017, Wilcoxon matched pairs test, Figure 4.6). Backgrounds were low in the *ex vivo* ELISpot, and were not significantly reduced with a pre-stimulation step (data not shown). Based on these results, it was decided not to include a pre-stimulation step, although this would have had the advantage of providing supernatant which could be used in further assays of other cytokine levels.

Figure 4-6: A pre-stimulation step does not increase the size of response to PPD-T or Antigen 85A pooled peptides using cryopreserved cells



A pre-stimulation step followed by PBMC washing resulted in the detection of reduced responses to PPD and to a pool of Antigen 85A peptides, using cryopreserved cells. Box and whisker plot showing medians and range of responses.*Wilcoxon matched pairs test

4.2.5 Altering DMSO concentration between 0.3% and 0.1% does not alter the response detected by *ex vivo* ELISpot or by *ex vivo* ELISpot incorporating a pre-stimulation step.

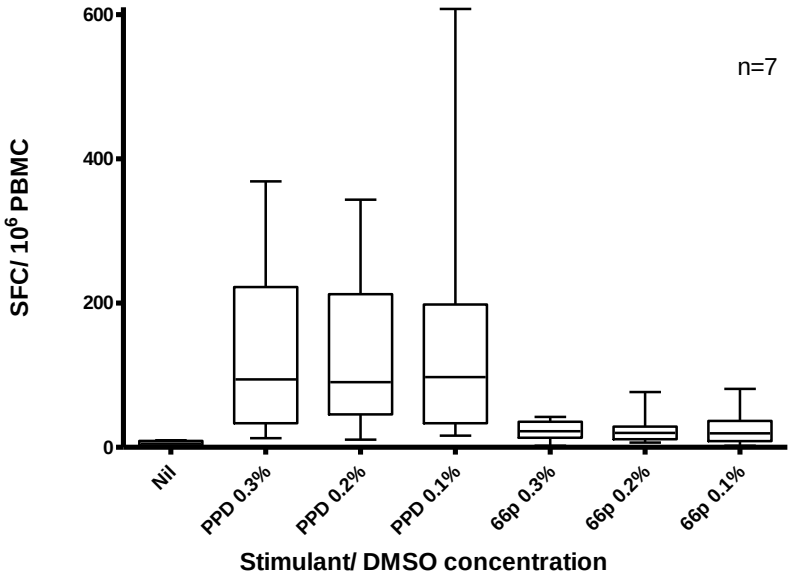
The number of peptides available for testing necessitated the use of large pools of peptides, which could result in high DMSO concentrations in the pools, with the possible effect of reducing the magnitude of responses detected [175]. The highest DMSO concentration used was 0.3% (range 0.2-0.3%). This could potentially have a greater effect if a pre-stimulation

step was incorporated, as a result of more prolonged contact of PBMC with DMSO. Therefore, the effect of varying the DMSO concentration was investigated in an experiment in which *ex vivo* ELISpot and *ex vivo* ELISpot with a pre-stimulation step were compared, studying responses to PPD-T and to a pool of peptides covering Antigen 85A. In this experiment, a 48 hour ELISpot incubation was used, in combination with a 24 hour pre-stimulation period, in a further attempt to increase the sensitivity of the assay (Figure 4.7)

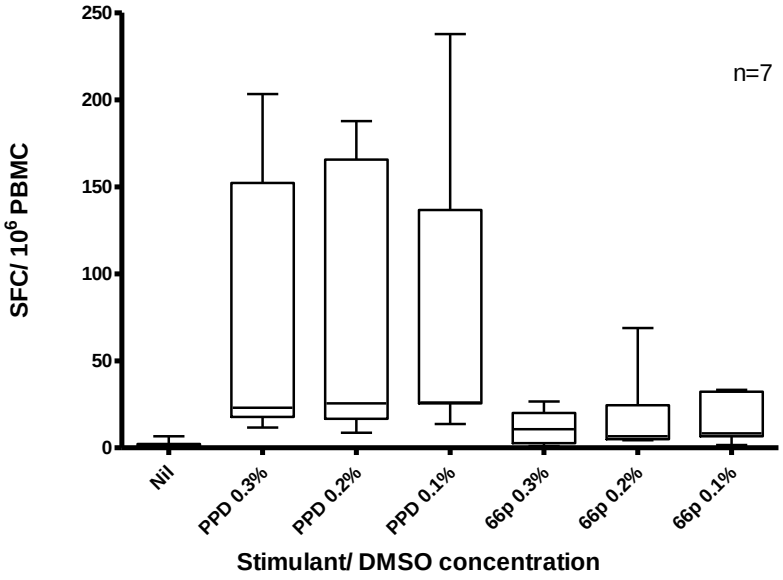
There was no difference in response to either PPD or to pool of Antigen 85A peptides at any of the tested concentrations of DMSO, both with *ex vivo* ELISpot and with 48 hour ELISpot with pre-stimulation (Figure 4.7, $P>0.05$, Friedman test, Dunn's multiple comparison test). There was also no difference in response comparing the two different methods at each DMSO concentration ($P>0.05$, Wilcoxon matched pairs test). Additional experiments comparing a range including 1%, 0.1% and 0.01% DMSO concentration also showed no detectable difference in immune response (data not shown).

Therefore for future experiments, pools of peptides with a DMSO concentration up to 0.3% were used.

Figure 4-7: The effect of varying DMSO concentration on the magnitude of ELISpot responses



A



B

Comparison of the effect of varying DMSO concentration on the IFN-γ response detected by *ex vivo* ELISpot, without (A) or with (B) a pre-stimulation step. Box and whisker plot showing medians and range of responses. Differences were not significant (Friedman test, Dunn's multiple comparison test).

4.2.6 Peptide screening: UK and US NTM-exposed and non-exposed populations

Responses to pools of NTM-specific peptides were analysed in humans; in a group of individuals recruited from the chest clinic at the Churchill Hospital, Oxford, who had had NTM isolated from their sputum on at least two occasions (section 4.2.2), and also in cord blood from normal healthy deliveries, as a negative control. The arrangement of peptides in larger pools is shown in Table 4.3. Results are shown in Figure 4.8. ELISpot positive and negative controls were the same as in the mouse experiments, apart from two differences. Firstly, a pool of peptides covering the ESAT-6 and CFP-10 proteins was substituted for the Antigen 85A pool. ESAT-6 and CFP-10 are highly immunogenic proteins present in *M. tuberculosis*, and absent from most species of NTM (other than *M. kansasii*, *M. marinum, M. szulgai* and *M. leprae* [7-9]). This peptide pool was introduced in order be able to exclude individuals with unexpected evidence of exposure to *M. tuberculosis*. Secondly, only PPD-T was used, and not PPD-A. This was done because there were no differences in response to PPD-T and PPD-A between mice exposed to *M. avium* or BCG (Figure 4.2). PPD-T will from here on be referred to as PPD.

Two individuals had low level responses to the ESAT-6/ CFP-10 pool (21.7 and 15 SFC/10^6 PBMC respectively) and were excluded from further analysis. Responses to PPD ranged from 0 to 296 SFC/10^6 PBMC; 50% of responses were above the cut-off for assay positivity (3 Median Absolute Deviations (MADs)). Responses to the NTM peptide pools were low, and none were statistically significantly above the negative control well (Figure 4.8A, $P>0.05$, Wilcoxon matched pairs test).

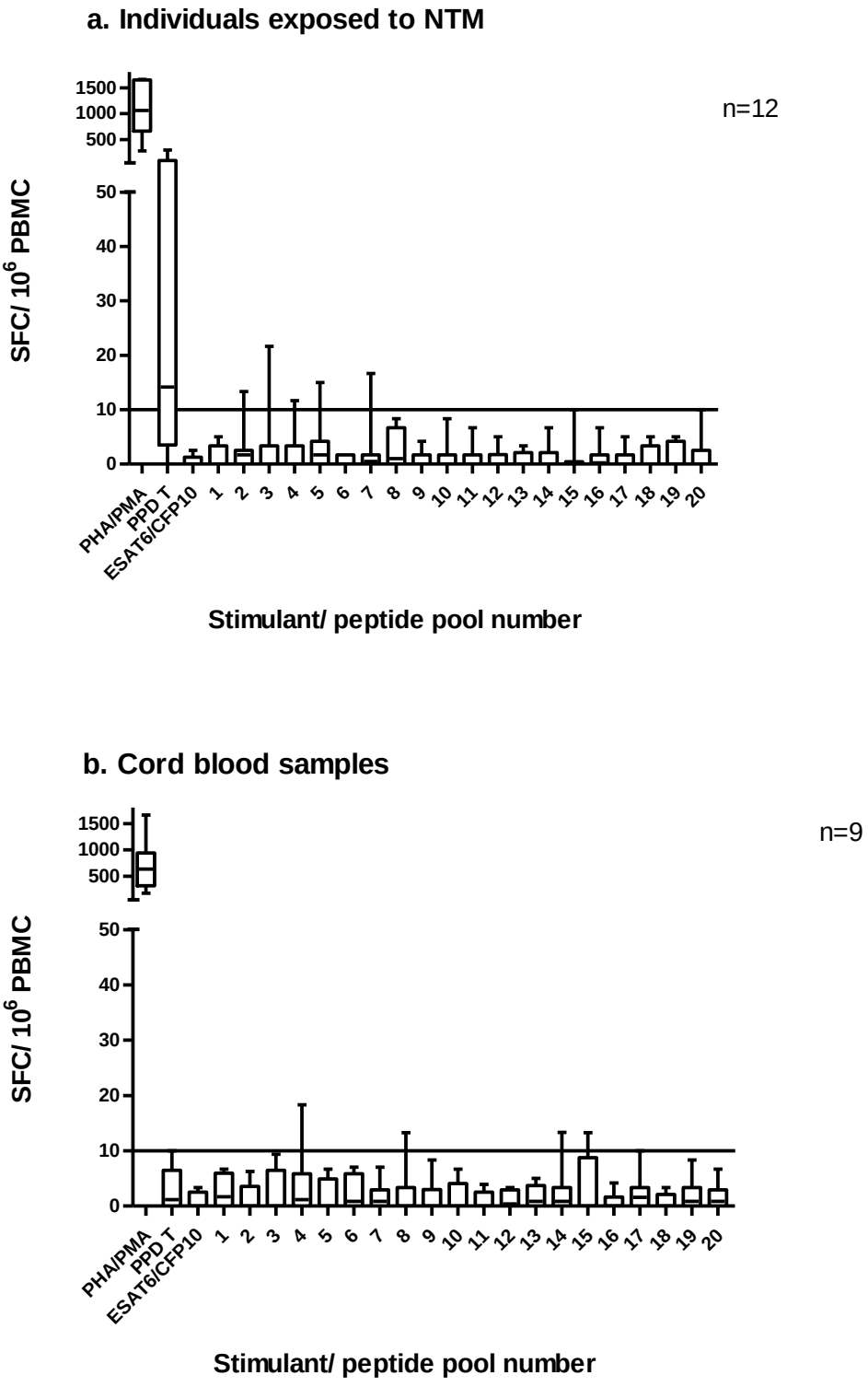
As expected, there were no responses to PPD in cord blood samples, nor were there statistically significant responses to the peptide pools (Figure 4.8B).

Table 4-3: Arrangement of peptides from different clusters into peptide pools

Cluster id	Protein family	No. proteins/ cluster	Predicted cellular location				Peptide pool no.
			Secreted	Lipid anchored	N-anchored	Intracellular	
0A	Mce family protein	16			12	1	1-5
0B	hypothetical protein	13	7	1		1	6-7
0C	hypothetical protein	12	4		2		8,11
0D	hypothetical protein	8	4		1		9
0E	hypothetical protein	8	4				10
0F	hypothetical protein	7	4				11,12
0G	hypothetical protein	7	4				13
0H	hypothetical proteins	7	3		1		17
0I	Tat-translocated enzyme	6	3	1			14,15
0J	27 kDa lipoprotein antigen	6	3				16
0K	hypothetical protein	6	3	3			17
0L	hypothetical protein	5	3	2			17

Summary of the clusters selected for the generation of peptides for experimental testing, and arrangement in pools for testing on human samples.
Cluster id: identification number allocated to cluster, protein family: predicted function of proteins within the cluster, predicted cellular location: predicted cellular location according to the LocateP database, peptide pool no.: the number of the peptide pool which clusters of proteins were tested in. These peptides came from all clusters, and were tested separately from the others as there was a concern they may be less specific. Note, number of protein location predictions is not always equal to number of proteins in a cluster as the LocateP database only contains predictions to reference proteins; the proteins annotated using Prodigal will not have associated location predictions.

Figure 4-8: Peptide screening in 'NTM-exposed' and 'non-exposed' UK subjects

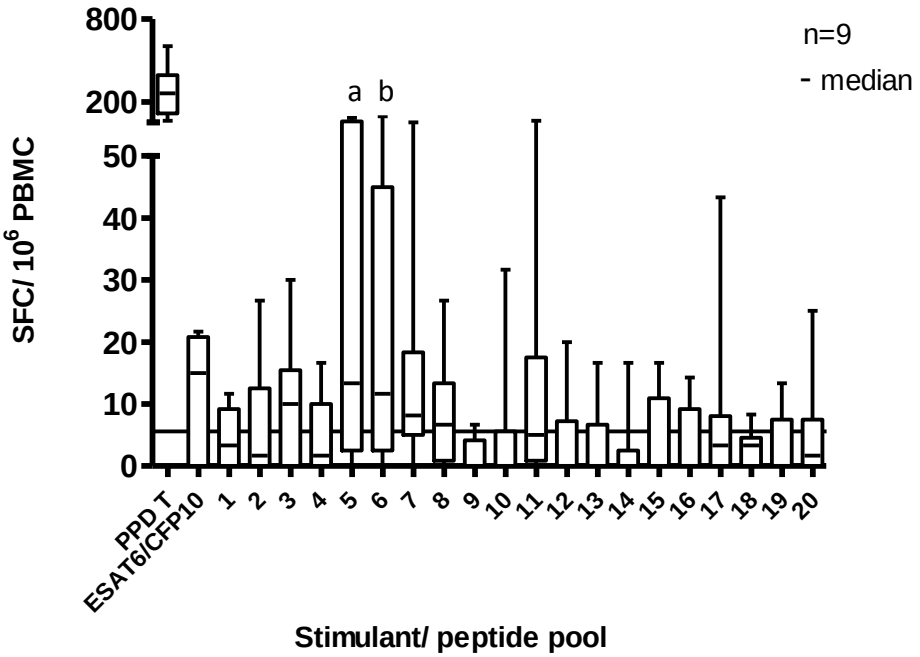


Ex vivo ELISpot IFN-γ responses to pooled peptides in individuals from whom NTM were isolated from sputum samples on at least 2 occasions and from cord blood samples. Peptides were tested in 20 pools containing between 58-85 peptides. Assay cut off = 3 Median Absolute Deviations (MADs). Box and whisker plot showing medians, interquartile range and range of responses.

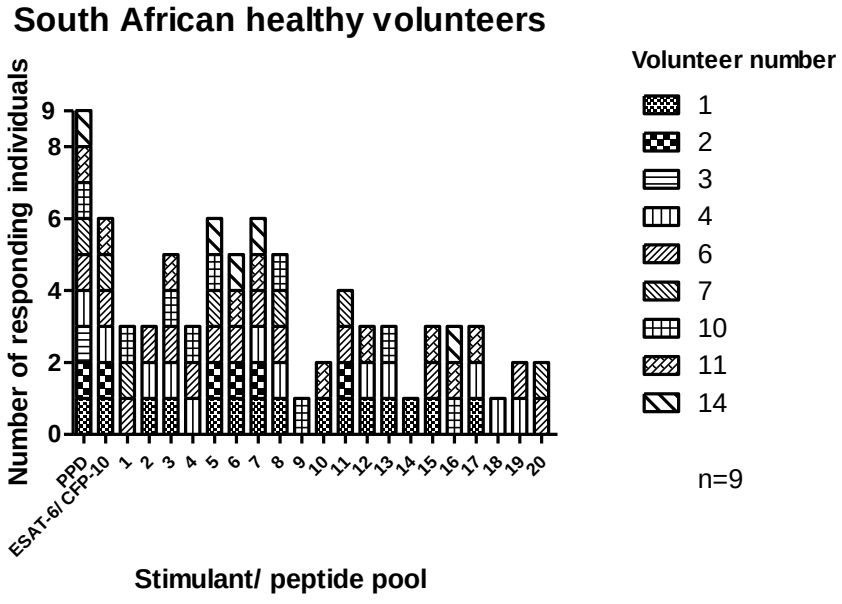
4.2.7 Peptide screening: healthy South Africans with no prior immunological evidence of TB exposure

PBMC IFN-γ responses were also determined by *ex vivo* IFN-γ ELISpot assay in 9 healthy South African donors with previous negative responses to QuantiFERON®-TB Gold In-Tube (Figure 4.9). Responses to PPD were universally strong, and some of these individuals had low level responses to the ESAT-6/CFP-10 peptide pool. Analysing responses to NTM-specific proteins, responses to peptide pools 5 and 6 were significantly increased above the negative control (pool 5: P=0.0065, pool 6: P=0.012, Wilcoxon matched pairs test). In addition, a broad range of individuals made responses over the cut off, in particular to pools 3, 5 (from cluster OA, the Mce family of proteins), 6, 7 (from cluster OB, a family of hypothetical proteins of unknown function) and pool 8 (from cluster OC, another family of hypothetical proteins of unknown function).

Figure 4-9: IFN-γ responses to pools of NTM-specific peptides in healthy South Africans



A



B

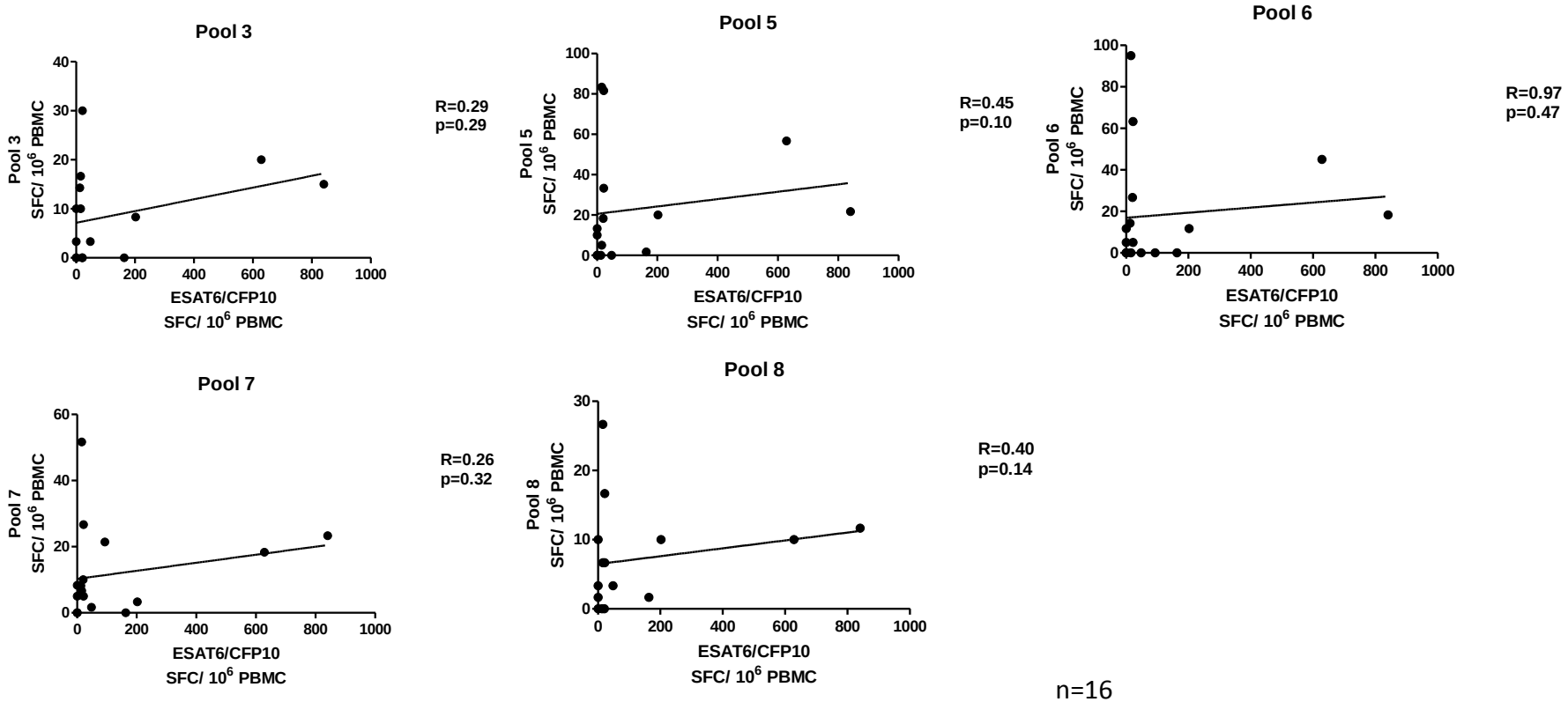
Responses to NTM-specific peptide pools in healthy South African volunteers with previous negative responses to QuantiFERON®-TB Gold In-Tube.

A: magnitude of IFN-γ responses. a: Pool 5: P=0.0065, b: pool 6: P=0.0162, Wilcoxon matched pairs test

B: number of individuals responding to each peptide pool.

Because of the low level of response to ESAT-6/CFP-10 peptides which was detected in this group, a second South African group of individuals with previous positive responses to the QuantiFERON®-TB Gold In-Tube test was also recruited, due to concerns that there might be cross-reactivity between the NTM-specific peptides and antigens present in *M. tuberculosis.* Analysing results from the two South African cohorts together (16 individuals), there was no significant correlation between response to any NTM-specific peptide pool and to the ESAT-6/ CFP-10 pool of peptides ($P>0.05$, Spearman correlation, Figure 4.10).

Figure 4-10: Correlation between response to ESAT-6/CFP-10 peptide pool and NTM-specific peptide pools



Spearman correlation between response to pool of ESAT-6/ CFP-10 peptides and NTM-specific peptide pools, in all healthy South African volunteers. Best fit lines generated using linear regression are shown for illustration.

4.3 Discussion

IFN-γ response to NTM-specific peptides was evaluated in an *ex vivo* ELISpot assay using mouse splenocytes from *M. avium* or BCG-infected C57BL/6 mice. It was expected that responses to certain pools of peptides might be detected in the *M. avium* infected mice, and that no responses would be detected in the BCG infected mice. Since BCG is a member of the *M. tuberculosis* family, infection with BCG should not prime immune responses which cross-react with the NTM-specific peptides. There were, however, no statistically significant responses to NTM-specific peptides in either group of mice.

There are a number of possible explanations as to why responses to the peptides were not detected in the mouse model. Firstly, the timing of mycobacterial infection and the time point of immunological assays may not have been be ideal; sub-dominant responses are known to be more readily detectable early in the time course of infection, with immunodominant responses becomingly increasingly prominent later on [176]. The time point was chosen based on the immune response to the Antigen 85A, which is a highly immunodominant antigen. It is possible that responses to the NTM-specific peptides may have been sub-dominant, and if so that they may have peaked earlier.

Secondly, the dose of mycobacterial exposure may have been inadequate, although the robust responses seen to both PPD and Antigen 85A peptides suggest that this is unlikely. The dose was chosen based on experiments carried out by Dr Clare Sander in the course of her DPhil in this laboratory [173]. However, it is possible that due to technical challenges some of the mice may have received some of the mycobacterial infectious dose subcutaneously. Because of the use of live *M. avium,* each experiment had to be performed

in quarantine in the animal facility, so the tail warmer could not be used. This made intravenous injections technically much more difficult, and some of them are likely to have been at least in part subcutaneous. Finally, *ex vivo* ELISpot may not have been a sufficiently sensitive assay to detect a response; if the peptides were not highly immunogenic, for example if they induced sub-dominant responses, as discussed above, it may be that a more sensitive assay may be required to detect these responses. The use of large peptide pools might also reduce the size of the immune response detected; this is discussed below.

Rather than spending time optimising the mouse model, I elected to move straight onto testing the peptides on the human samples which I had collected and stored, testing all of the peptides with the exception of those to which there were low level responses in the BCG-exposed mouse model (pools 1A-1L), in fewer, larger pools. It was decided that spending further time in developing the mouse model of screening was not justified, as it is likely that responding epitopes (and therefore peptides) between mice and humans would differ.

Interestingly, the pre-stimulation step modifications to the human *ex vivo* ELISpot protocol all resulted in reduced responses to peptide and protein antigens. This contrasts with the results of the study by Smith *et al*, in which responses to protein antigens were increased and those to peptides were unchanged [174]. It also contrasts with the small pilot performed in the mouse, in which increased responses to PPD were detected, and reduced responses to peptides, although numbers were too small to perform statistical analysis on this (Figure 4.3). It might be expected that a longer incubation period in which further antigen processing and presentation could take place might result in greater responses to protein antigens; this would not be expected to make a difference with peptide antigens

since they do not require further processing prior to binding the MHC complex on the antigen presenting cell. It is not clear why I did not find an increased response to PPD with a longer incubation period in the human samples, although it may be that incorporating a washing step after the pre-stimulation resulted in a reduction in the number of PBMC plated onto the ELISpot plate.

Ranges of DMSO from 0.1-0.3% did not adversely affect the results of ELISpot experiments, including those that incorporated a pre-stimulation step. This was reassuring, and supported the use of large pools of peptides. This finding is also supported by the literature; Kloverpris *et al* described decreased T cell cytokine production which was incubation time-dependent in the presence of DMSO concentrations above 1.5% (7 days culture) [177], and Suneetha *et al* showed decreased CD8 T cell proliferative capacity in the presence of 1% DMSO with a borderline decrease in proliferation at 0.5% DMSO and no detectable decrease at 0.1% (7 days culture) [175].

There are also theories that using peptide pools made up from very large numbers of peptides may reduce the chances of detecting significant responses, due to a reduced chance of peptide/ MHC binding, or some form of 'interference'. For example, it has been shown in studies of immune responses to peptides from malaria antigens that certain combinations of peptides, if co-presented, have the ability to inhibit the activation of cytotoxic T cells, although proliferative responses were not affected in this study [178].

It can be difficult to separate the effects of large numbers of peptides in a pool from a higher concentration of DMSO in those larger pools, as the two often co-exist. Winstone *et al* demonstrated high levels of CD4 T cell IFN-γ production from PBMC from HIV positive volunteers receiving a novel HIV vaccine, using pools containing 90 peptides and a DMSO

concentration of up to 0.45%. Mustafa *et al* carried out a thorough comparison of the effect of size of peptide pool on CD4 T cell IFN-γ production and proliferation, using no DMSO at all in their peptide pools, and showed that similar responses were detected using peptide pools up to 220 peptides in size, compared to pools of 6 peptides (although arguably larger responses should have been detected in the former) [179].

Screening NTM-specific peptide pools on PBMC from human volunteers, there were no significant responses to any pools in the UK volunteers with documented culture of NTM from sputum samples (and therefore with documented NTM exposure). Surprisingly, given that their exposure to NTM was highly probable, only 50% made significant responses to PPD. PPD consists of a very immunogenic mixture of proteins, many of which are shared between *M. tuberculosis* and NTM. Therefore a response rate to PPD closer to 100% would be expected in a cohort with definite exposure to NTM.

There are a number of explanations as to why responses to NTM-specific peptide pools were so low in this group. Firstly, the peptides may not elicit a significant IFN-γ response, either because they elicit no immune response at all, or because they elicit an immune response characterised by cytokines other than IFNγ, such as regulatory cytokines. This chapter is solely concerned with measuring IFN-γ responses, and it is possible that exposure to these peptides may stimulate an immune response that is not characterised by a Th1 response. As mentioned in Section 1.3.1 (Chapter One, Introduction), Hernandez *et al* demonstrated that exposing mice to different doses of killed *M. vaccae* resulted in the induction of dramatically different immune responses on *M. tuberculosis* infection, with a lower dose resulting in a Th1 immune response and a higher dose in a Th2 response [117].

The possibility that the response generated by stimulation with the NTM-specific peptides may be a Th2 response is investigated further in Chapter 5 (Further Immunology).

NTM are generally not highly pathogenic, unlike *M. tuberculosis,* and, as discussed previously, it may be that their most immunogenic epitopes are shared by members of the *M. tuberculosis* complex, and those that are unique to NTM are sub-dominant and hence harder to detect.

Alternatively, there may be specific problems with the 'NTM-exposed' PBMC samples. Firstly, exposure levels to NTM may have been low- individuals were selected on the basis of NTM isolation from sputum rather than on evidence of infection (the latter group would have been much more difficult to recruit). Although some individuals may have had long-term NTM colonisation, with on-going exposure, others may have been transiently colonised. In addition, NTM are generally relatively non-pathogenic, so exposure may not induce a robust immune response. Furthermore, the individuals from whom the blood was taken were suffering from chronic lung diseases, often cystic fibrosis, which is associated with a number of factors (poor nutritional status, repeated and frequent chest infections) which might be associated with immunosuppression [180]. A small proportion of patients were on steroids, which would be expected to depress the T cell response. This may explain the finding that 50% of individuals did not respond to PPD, so it is perhaps not surprising that responses to the pools of peptides (each representing a small subset of secreted proteins) are lower. Finally, experiments were performed on cryopreserved cells, from which responses may be reduced up to 50% compared to experiments using fresh cells [181].

Statistically significant but low level responses were seen to pools 5 and 6 of the NTM-specific peptide pools in the healthy South African cohort (Figure 4.9). These pools represent the Mce family of proteins (pool 5) and a cluster of hypothetical proteins (pool 6). Additionally, a range of low-level responses was seen to multiple other peptide pools. It is interesting that considerably stronger responses were detected in the South African cohort, both to PPD and to the NTM-specific peptides. There are several potential reasons for this. This was a healthy group of individuals, with no known causes of immunosuppression, and PBMC were used fresh. In addition, it is possible that they were exposed to higher levels of NTM than the UK NTM-exposed group. The lack of correlation between response to ESAT-6/CFP-10 pool of peptides and NTM-specific peptides makes *M. tuberculosis* exposure a less likely explanation. ESAT-6 and CFP-10 are present in *M. tuberculosis, M. kansasii, M. szulgai M. marinum,* and *M. leprae* to varying levels of homology, and are absent from all other sequenced NTM [7-9]. It was therefore assumed that absence of a response to ESAT-6/ CFP-10 peptides indicated that the individual concerned had not experienced significant exposure to *M. tuberculosis.*

However, it is impossible to define the extent of NTM exposure in this group, or indeed if individuals had been exposed at all. It is also possible that the individuals in whom no responses were detected to NTM-specific proteins had not been significantly exposed to NTM.

The immunogenic Mce (mammalian cell entry) family of proteins are virulence factors which are involved in mycobacterial entry and survival in macrophages [167, 168]. Mce knockout mutants of *M. tuberculosis* have reduced virulence in mice [169], *M. bovis* Mce4 induces an inflammatory response in bovine macrophages [146] and antibodies have been

demonstrated to Mce proteins in TB-infected humans [170]. It is likely, therefore, that the Mce family of proteins may also be virulence factors in NTM. It should be noted, however, that the cluster of NTM-specific Mce proteins we selected are not similar in primary sequence to those in the *M. tuberculosis* complex; if they had been, they would have been excluded from the pipeline. The remaining protein families against which immune responses were detected were hypothetical proteins of unknown function, which are the most frequent class of protein predicted by this, and by other similar pipelines [151, 166].

It is interesting that low-level responses were detected to such a wide range of peptide pools in the South African cohort. Such a pattern of responses, if present across the many antigens shared between NTM and *M. tuberculosis,* might contribute to a level of cross-protective immunity against *M. tuberculosis.*

It is promising that responses were detected to NTM-specific antigens. However, these responses were significantly lower than are routinely seen to immunodominant *M. tuberculosis*-associated mycobacterial antigens in individuals with latent *M. tuberculosis* infection. As discussed above, NTM are not highly pathogenic, and NTM-specific proteins may be less immunogenic than the proteins shared with *M. tuberculosis.* Secondly, in the absence of a gold standard test for NTM exposure we could not determine whether all South African volunteers had been exposed, nor how recently. The individuals in whom we detected positive responses may have been the only individuals with recent and significant exposure.

4.4 Conclusion

NTM-specific peptides were tested in pools on a mouse model of NTM exposure, then on human subjects with definite, probable or no NTM exposure and with probable or no *M. tuberculosis* exposure. Low level responses were defined to two pools of peptides in healthy South Africans, but not in individuals from the UK with definite exposure to NTM. Further studies therefore focussed on using more sensitive assays to investigate these responses, and on investigating the possibility that responses may be characterised by the production of cytokines other than IFN-γ.

5 Further Immunology

5.1 Introduction

Significant IFN-γ responses were detected by *ex vivo* ELISpot to the NTM-specific pools 5 and 6 in the South African cohort, and there were borderline responses to multiple other pools (Section 4.2.7, Chapter Four, Basic Immunology). Responses appeared to be specific to NTM exposure, with no correlation between response to any NTM-specific peptide pools and response to a pool of ESAT-6/ CFP-10 peptides. However, in the UK and US cohorts, in whom exposure to NTM was documented, no responses to the NTM-specific peptide pools were detected by *ex vivo* ELISpot (Section 4.2.6). There are a number of possible explanations for the lack of detection of immune responses in this cohort, outlined in Chapters Four and Six (Basic Immunology and Discussion).

The aim of this chapter was to determine whether using either more sensitive methods, or methods which do not rely on the detection of IFN-γ, might result in the detection of immune responses to NTM-specific peptide pools in the UK and US NTM-exposed cohorts. Immunological assays which involve longer-term culture, and therefore cellular expansion, are more sensitive than direct *ex vivo* ELISpot [182], so two culture-based methods were evaluated. These were a cultured ELISpot and a proliferation assay. In addition, a multiplex bead array was used to investigate the possibility that cytokines other than IFN-γ might be produced in response to the NTM-specific peptide pools.

The cultured ELISpot and proliferation assays were performed at the South African TB Vaccine Initiative (SATVI) laboratories, University of Cape Town, South Africa, and the multiplex assays at the Jenner Institute, Oxford University, UK.

Cultured ELISpot involves the culture of short-term cell lines for 9-12 days in the presence of antigen, with the addition of interleukin 2 (IL-2) [183-185] or IL-2 and IL-7 (Dr I Satti, personal communication). IL-2 facilitates T cell proliferation and, at certain concentrations, the development of memory T cells [186], and enhanced survival, the latter partly due to the up-regulation of the IL-7 receptor [187]. IL-7 is important in the maintenance of T cell homeostasis, and facilitates persistence and survival of both effector and memory T cells [187, 188]. The aim of the assay is to expand low frequency antigen-specific effector T cell populations, in order to facilitate detection. It should be borne in mind, however, that adding IL-2 and IL-7 is likely to alter the phenotype of the cells subsequently detected.

Proliferation assays, which generally involve slightly shorter culture periods of up to 6 days, do not require the addition of cytokines for cell survival. There are three main types of cellular proliferation assay; those that measure the incorporation of a substance into the DNA of the proliferating cell, those that measure the progressive dilution of an intracytoplasmic dye as cells divide and proliferate, and those that measure the expression of cell surface or nuclear markers which are expressed at particular stages of the cell cycle.

Early proliferation assays measured the incorporation of tritiated thymidine into the DNA of proliferating cells, and were highly sensitive and specific [189], but had the disadvantage of using radioactivity. Subsequent modifications to this method have included measuring incorporation of the non-radioactive thymidine analogue bromodeoxyuridine (BrdU) into the DNA of proliferating cells.

Assays measuring the concentration of a fluorescent dye (eg carboxyfluorescein diacetate succinimidyl ester, CFSE [190]) which is progressively diluted with subsequent rounds of cell division, have the major advantage that the number of cellular divisions can be quantified

by measuring intracellular CFSE concentration, allowing back-calculation of the original precursor frequency of proliferating cells. Furthermore, cells can be sorted depending on their intracellular CFSE concentration, and further studies undertaken on the populations of cells which have proliferated differing numbers of times. CFSE is, however, toxic to cells, reducing both cellular proliferation and the expression of cell surface activation markers [189]. The CFSE derivative Oregon Green carboxylic acid diacetate (Oregon Green) has been reported to be less toxic to cells and to stain more brightly than CFSE [191]; other dyes have also been used, and are summarised by Wallace *et al* [191].

The measurement of different cellular activation markers which are associated with cell turnover can also be used to assess proliferation levels, and this is associated with varying levels of sensitivity and specificity depending on the marker measured [189]. The nuclear marker Ki67 is expressed on cell division [192], with high levels of expression from 60 hours in culture with stimulating antigen [193]. It has a high sensitivity, even compared with tritiated thymidine incorporation [193]. All of these methods (with the exception of tritiated thymidine incorporation) allow co-staining with other cellular markers, so cellular phenotypic and functional characteristics (e.g. cytokine production) can also be determined [140].

Lastly, a multiplex assay was used to screen culture supernatants from all available PBMC samples, to investigate the possibility that the NTM-specific peptides generate an immune response which is characterised by the production of cytokines other than IFN-γ. It is conceivable that prior exposure to NTM followed by PBMC stimulation with pools of NTM-specific peptides could result in an immune response characterised by the production of

regulatory or Type 2 cytokines (as described by Hernandez *et al* when mice were exposed to a higher dose of *M. vaccae* [117]).

An 11-plex multiplex cytokine bead-based array and a TGF-β simplex bead array (both eBiosciences) were used to determine the concentrations of 12 different Th1 and Th2 cytokines in supernatants from stimulated PBMC. The eBiosciences multiplex bead-based assays are routinely used at the Jenner Institute, based on favourable sensitivity profiles and cost. Multiplex assays have the advantage of providing multiple cytokine measurements using very small volumes of sample, although they are expensive.

There are several multiplex assays available, all of which allow the simultaneous quantitation of large numbers of different cytokines. The eBioscience 11-plex fluorescent bead immunoassay uses beads which are coated with antibodies which bind each of the cytokines to be detected; a secondary biotin-conjugated antibody is then added, followed by Streptavidin-Phycoerythrin (Streptavidin-PE). The beads exist in two sizes, and beads measuring different cytokines have been dyed with fluorescent dye of differing intensity, allowing differentiation between the beads, and hence the cytokines, based on size and fluorescence intensity.

5.2 Results

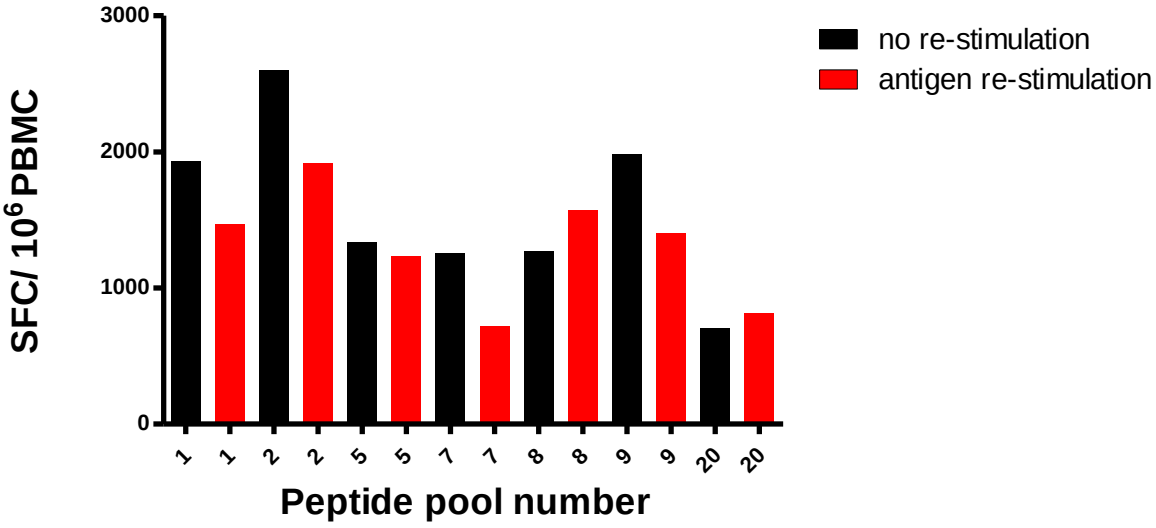
5.2.1 Cultured ELISpot

First, the cultured ELISpot method was evaluated (see Section 2.8.3.6, Chapter 2, Methods). Each sample was cultured with antigen in triplicate on day 0, and on day 12 a standard *ex vivo* ELISpot assay was plated out. In this, one of the triplicate samples was re-stimulated with the same antigen as on day 0, one with nil (‘un-stimulated’) and one with PHA.

Cryopreserved PBMC were used from a healthy South African donor with a known negative response to QuantiFERON®-TB Gold In-Tube, and were rested for four hours after thawing and prior to use in the assay.

The results are shown in Figure 5.1; un-stimulated samples were cultured in the presence of peptide pool from day 0, but not re-stimulated with further peptide on day 12. High levels of background IFN-γ response were seen in all of the un-stimulated samples. PHA responses are not shown; median PHA response was 5617 SFC/ 10^6 PBMC (range 2567- 9033 SFC/ 10^6 PBMC).

Figure 5-1: Cultured ELISpot responses to NTM-specific peptide pools

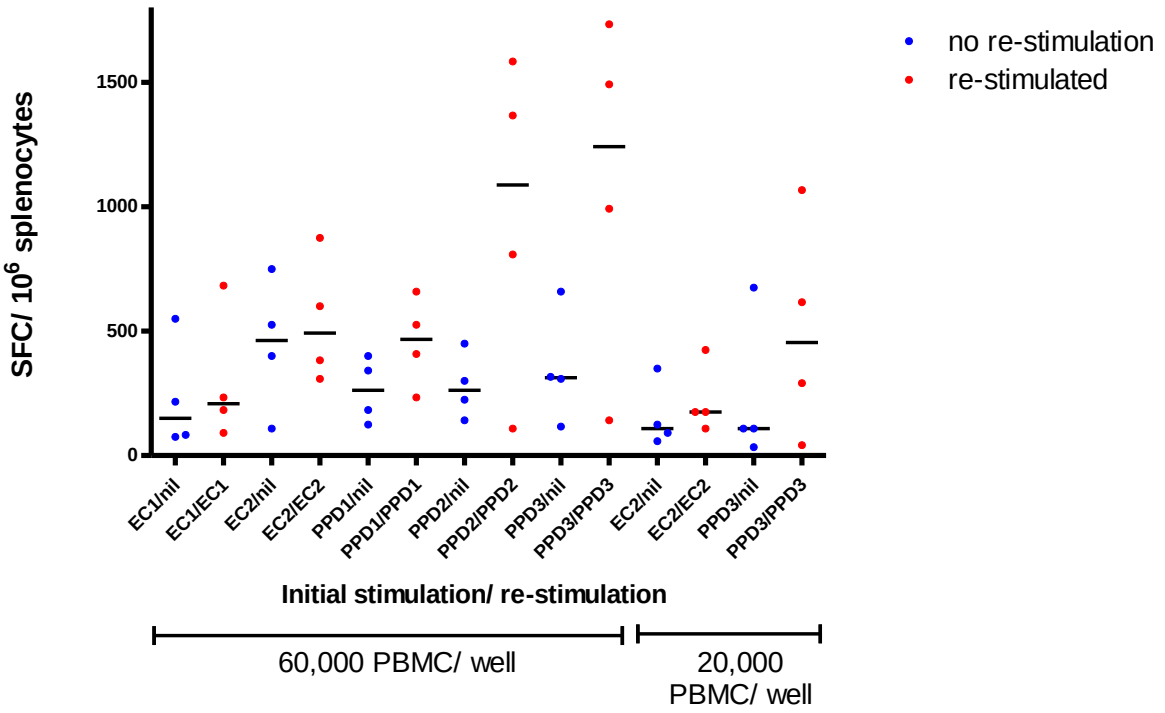


Cultured ELISpot responses to NTM-specific peptide pools, showing background responses (black bars, sample stimulated with NTM-specific peptide pools on day 0 and not re-stimulated on day 12) and NTM-specific peptide pool responses (red bars, sample stimulated with NTM-specific peptide pools on day 0 and again on day 12). PBMC from one individual were tested.

A second attempt was made to perform the cultured ELISpot assay without such high levels of background, resting PBMC overnight prior to their use in the hope that this would reduce background proliferation. In this experiment, the concentration of ESAT-6/ CFP-10 peptide

pool and of PPD was also titrated to determine the optimal concentration (Figure 5.2). Levels of background response were lower than in the previous experiment, but remained unacceptably high. Greater responses (per million PBMC) were detected if 60,000 PBMC rather than 20,000 were plated into a well, although this difference was not statistically significant.

Figure 5-2: Cultured ELISpot titration of ESAT-6/CFP-10 and PPD concentrations



Cultured ELISpot: all samples were stimulated with antigen on day 0. Results shown in blue were left un-stimulated on day 12 ('background'); those shown in red were stimulated with the same antigen on days 0 and 12. EC1: ESAT-6/CFP-10 2µg/ml, EC2: ESAT-6/CFP-10 1µg/ml, PPD1: PPD 5µg/ml, PPD2: PPD 2µg/ml, PPD3: PPD 1µg/ml. Results are expressed as SFC per million PBMC; the results of plating 60,000 and 20,000 PBMC on the final ELISpot plate are compared. Group size = 4. There was no significant difference between results using 60,000 and 20,000 PBMC per well (P>0.05, Wilcoxon matched pairs test).

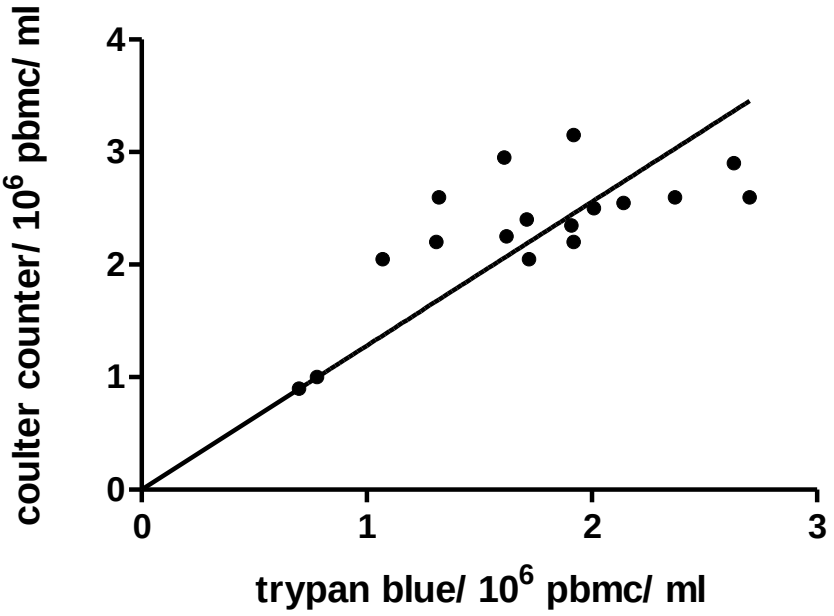
A proliferation assay was also evaluated, and due to superior results without the same problem with background responses, and also due to the greater ease with which the

method could be used for the screening of large numbers of samples, it was decided to use this method instead (see later).

5.2.2 Coulter counter and Trypan Blue counts of thawed cryopreserved PBMC correlate well

For the cultured ELISpot, manual cell counting had been performed using a microscope and Trypan Blue staining. In order to perform large numbers of proliferation assays, an automated method would be preferable, so Trypan Blue counting was compared with automated Coulter Counter counting. 17 samples of thawed PBMC (from adult and from cord blood samples) were counted manually and using the Coulter Counter, and a close correlation of results was found (R=0.6222, P=0.0077, Spearman's correlation, Figure 5.3). The Coulter Counter was therefore used from here on. This method appeared to have advantages over Trypan Blue counting of cord blood cells, as these contain nucleated red blood cells, which are difficult to distinguish from lymphocytes using Trypan Blue, but which can be readily distinguished using the Coulter Counter.

Figure 5-3: Comparison of coulter counter and Trypan Blue methods of counting PBMC numbers



Spearman correlation between counting PBMC using Trypan Blue staining and a microscope and an automated Coulter Counter. R=0.6222 (95% confidence intervals 0.187-0.853), P=0.0077 (2-tailed). A line was fitted for illustration using linear regression.

5.2.3 Staining with Oregon Green results in PBMC death

Following review of the available PBMC proliferation methods, it was decided to compare Oregon Green and Ki67 staining; both of these were methods in current use in the laboratory at the South African TB Vaccine initiative, and are robust, sensitive and specific. Oregon Green has been shown to be less toxic to cells compared to CFSE, so should also have less effect on proliferation, cell surface marker expression and cytokine production [191]. Staining for Ki67 is straightforward, as it involves no pre-staining step as is required in the Oregon Green assay.

A comparison of the two methods was therefore undertaken, to establish which was more sensitive in detecting cellular proliferation. A similar comparison using fresh PBMC had

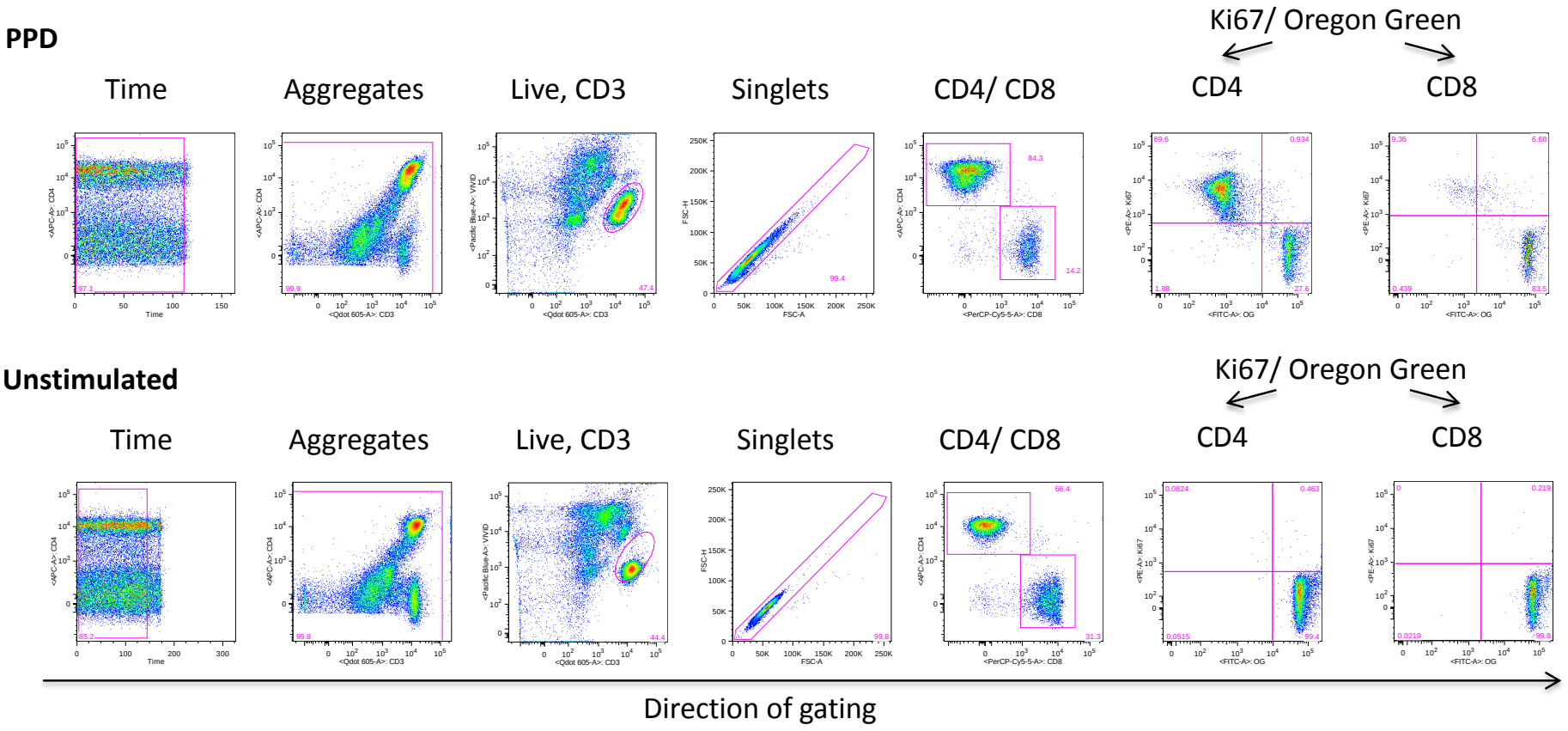
already been undertaken in the SATVI laboratory in Cape Town, and had concluded that the two methods had similar sensitivity [140], but there was concern from unpublished data that the Ki67 staining method may be associated with greater background levels of staining in un-stimulated cells when using cryopreserved cells (T Scriba, SATVI, University of Cape Town, personal communication).

Fresh and cryopreserved PBMC were used from healthy South African volunteers with known positive responses to QuantiFERON®-TB Gold In-Tube. The gating strategy used is outlined in Figure 5.4; cells were initially gated on time (to exclude changes in cell flow rate), aggregates (to exclude antibody aggregates), live CD3 positive cells (live-dead marker ViViD low) and singlets. Subsequent gates were CD4 or CD8, followed by Oregon Green (in the FITC channel) and Ki67. Note, proliferating cells would be expected to be Oregon Green low (due to dilution from cell division) but Ki67 high (due to cellular up-regulation).

Percentage proliferation was 70.5% (Ki67) compared to 71.5% (Oregon Green) in the PPD-stimulated samples, and 0.5% (Ki67) compared to 0.1% (Oregon Green) in the un-stimulated samples. Percentage concordant results were 97% (PPD, CD4), 92.9% (PPD, CD8), 99.5 (Unstimulated, CD4) and 99.7% (Unstimulated, CD8). When the experiment was repeated, background was lower in the Ki67 stained samples compared to the Oregon Green-stained samples (data not shown).

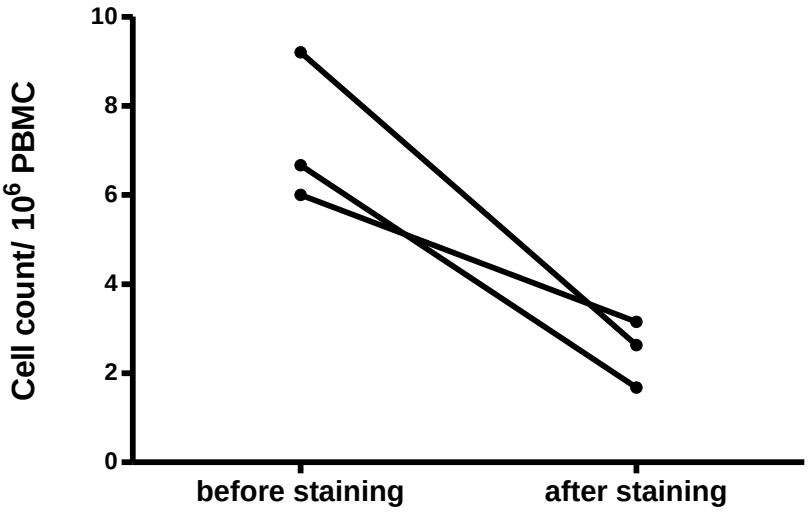
It was also noted that the Oregon Green staining process was associated with a marked decrease in cell count (Figure 5.5). Given that the anticipated problems with background proliferation did not occur in the Ki67 samples, and given the high levels of cell loss during the Oregon Green staining procedure, the Ki67 method was selected.

Figure 5-4: Gating strategy for comparison of Ki67 and Oregon Green staining in the detection of PBMC proliferation



Gating strategy for comparison of Ki67 and Oregon Green staining. Time gates (excluding fluctuations in fluorescence), antibody aggregate exclusion gates and forward scatter/ side scatter gates (selecting singlets) were followed by gating on live CD3 positive cells, then CD4 or CD8 positive cells, then Ki67 positive or negative and Oregon Green positive or negative cells. Note that proliferating cells are Ki67 high but Oregan Green low (as a result of dye dilution following proliferation). There is good agreement between the two methods, with percentage concordant results of 97% (PPD, CD4), 92.9% (PPD, CD8), 99.5 (Unstimulated, CD4) and 99.7% (Unstimulated, CD8).

Figure 5-5: Comparison of PBMC cell counts before and after staining with Oregon Green



PBMC cell counts are shown before and after staining with Oregon Green; there was a non-significant reduction in cell count following staining (P>0.5, Wilcoxon matched pairs test).

5.2.4 Using RPMI/ AB serum results in lower levels of background proliferation compared to AIM-V media

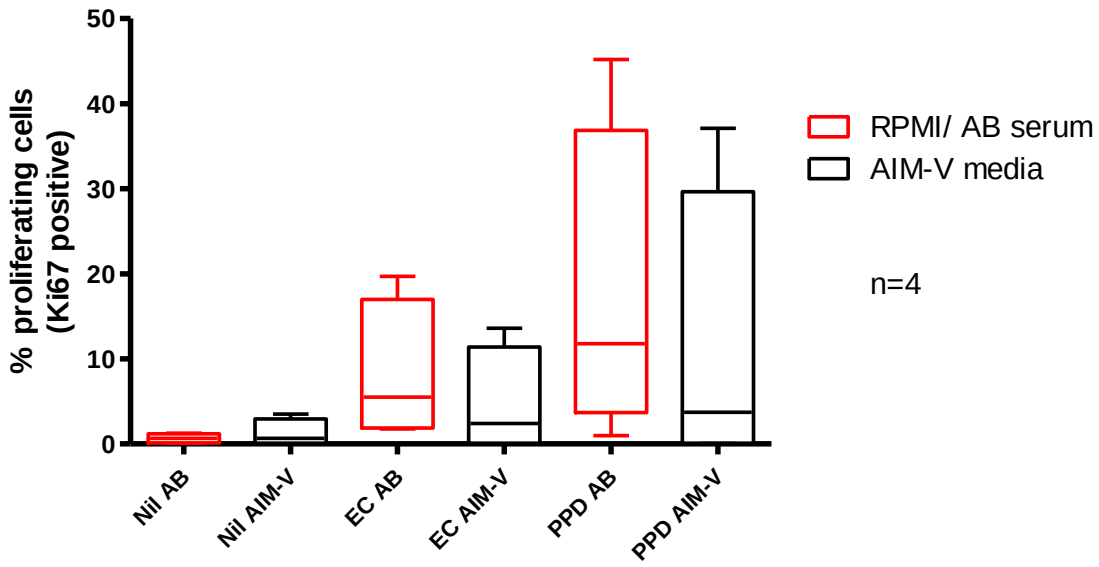
During the course of the comparisons of Oregon Green and Ki67 above, it was noted that levels of background proliferation were sometimes unacceptably high, reaching levels of 10% in some experiments (data not shown).

Certain media may be associated with higher levels of background proliferation than others, for example R10, which contains fetal calf serum, has been observed to result in higher levels of background signal compared to the synthetic AIM-V media, both in the context of *ex vivo* ELISpots [174] and 12 day cultured ELISpot (Dr I Satti, Jenner Institute, Oxford University, personal communication). A comparison of the standard media used in the SATVI laboratory, RPMI with human AB serum (12.5%) (RPMI/ AB), with AIM-V was therefore

undertaken using cryopreserved PBMC from laboratory volunteers with known positive responses to QuantiFERON®-TB Gold In-Tube (Figures 5.6, 5.7).

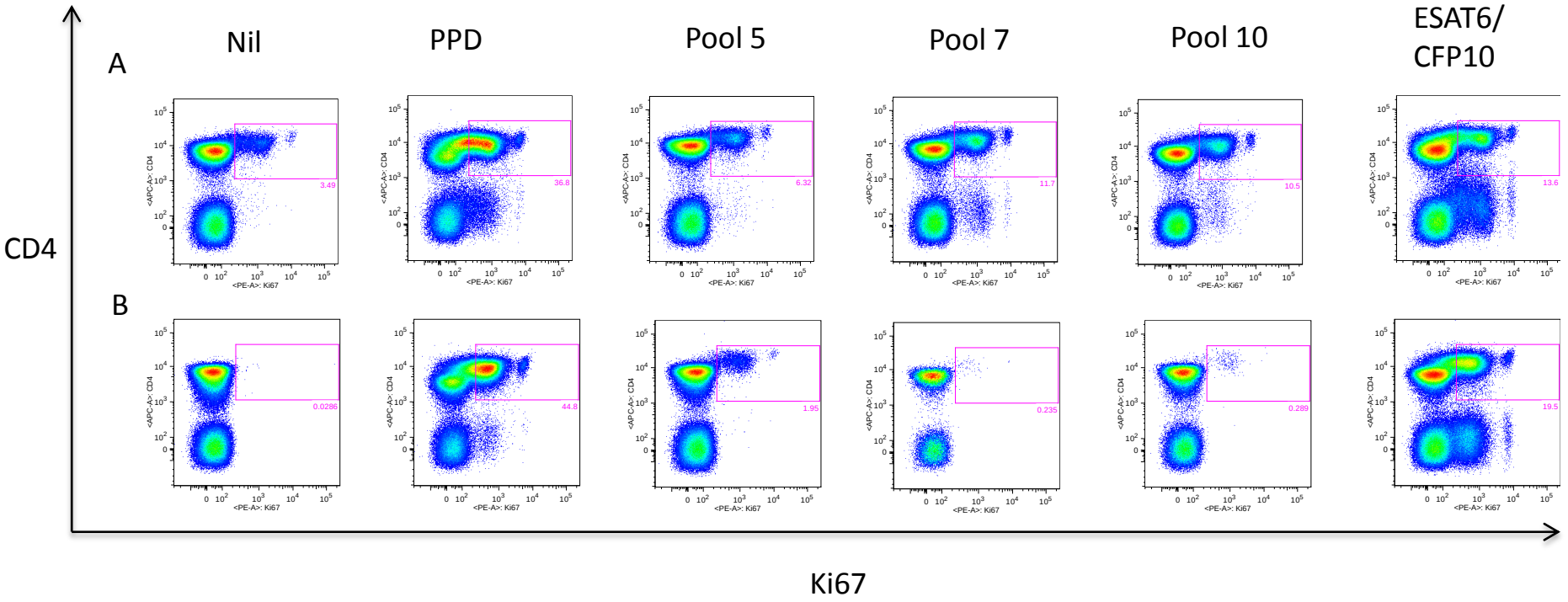
It was concluded that RPMI/ AB resulted in lower levels of background proliferation and greater responses to antigens detected, though none of the differences were significant (P>0.05, Wilcoxon matched pairs test). RPMI/ AB was therefore used in subsequent experiments.

Figure 5-6: The effect of using RPMI/ AB serum or AIM-V media on proliferation levels in un-stimulated PBMC



The effect of different media types on background proliferation levels was assessed. RPMI with 12.5% human AB serum was compared with AIM-V media. Background proliferation (no antigen), ESAT-6/CFP-10 pooled peptides at 1µg/ml and PPD at 1µg/ml were compared. Box and whisker plot showing medians, interquartile range and range of responses. None of the differences between AIM-V media and RPMI/ AB were significant (P>0.05, Wilcoxon matched pairs test).

Figure 5-7: Representative FACS plot of comparison of AIM-V and RPMI media with human AB serum

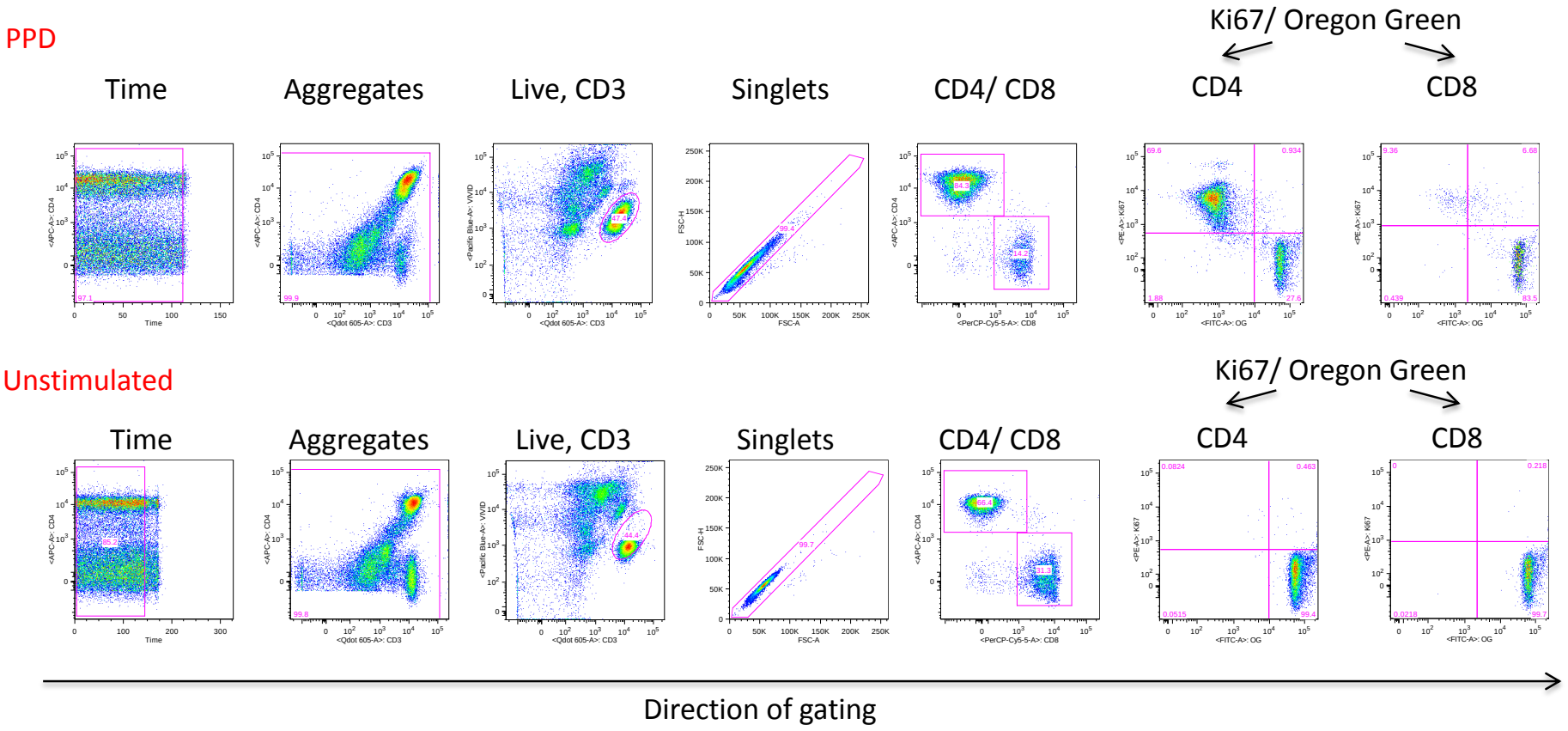


Representative example of proliferation assay using either AIM-V media (A) or RPMI/AB serum (B). Ki67 expression in the negative control well was 3.49% in the AIM-V well compared to 0.03% PBMC in the RPMI/ AB serum well.

5.2.5 Proliferation assay results

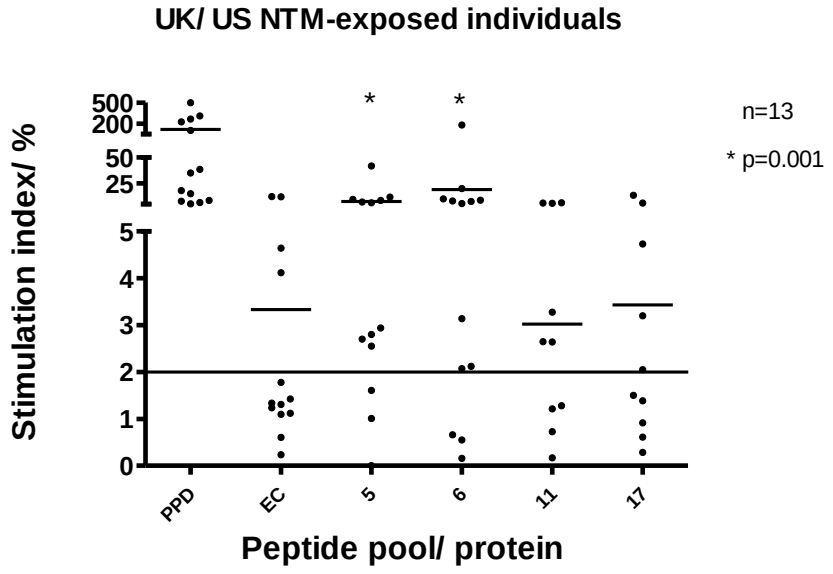
A 6 day proliferation assay was carried out on all remaining cryopreserved PBMC from the UK and US NTM-exposed individuals (n=13) and all remaining cord blood cells (n=9) (Section 2.8.3.7, Chapter 2, Methods). The gating strategy is outlined in Figure 5.8, and in Section 5.2.3. The proliferation assay used four NTM-specific peptide pools: 5, 6, 11 and 17, which were chosen on the basis of responses (pools 5 and 6) or borderline responses (pools 11 and 17) detected by *ex vivo* ELISpot (Figure 4.9). There were significant, high level proliferative responses to pools 5 and 6 in the UK based NTM exposed individuals, with 7 individuals responding to pool 5 and 5 to pool 6 (Figure 5.9). CD8 positive lymphocyte proliferation levels were low (data not shown), and cord blood responses were again low.

Figure 5-8: Gating strategy for the proliferation assay

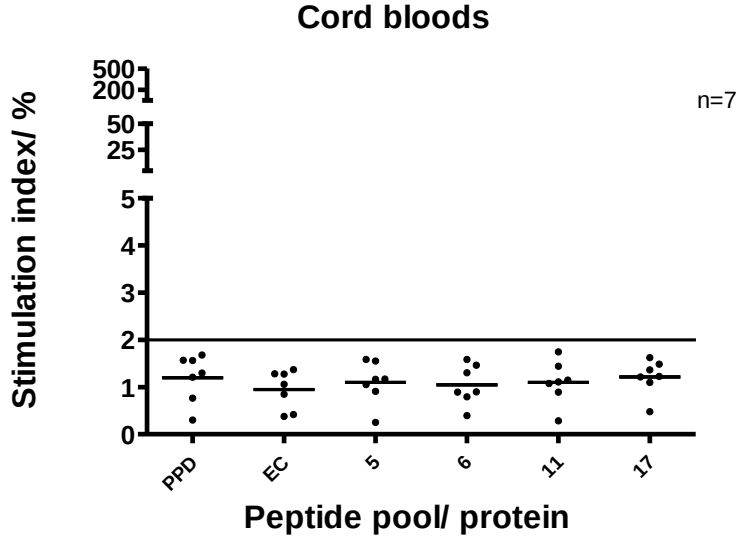


Gating strategy used in the 6 day proliferation assay. Time gates (excluding fluctuations in fluorescence), antibody aggregate exclusion gates and forward scatter/ side scatter gates (selecting singlets) were followed by gating on live CD3 positive cells, then CD4 or CD8 positive cells, then Ki67 positive or negative and Oregon Green positive or negative cells. Note that proliferating cells are Ki67 high but Oregan Green low (as a result of dye dilution following proliferation).

Figure 5-9: 6 day proliferation assay responses to pools of NTM-specific peptides



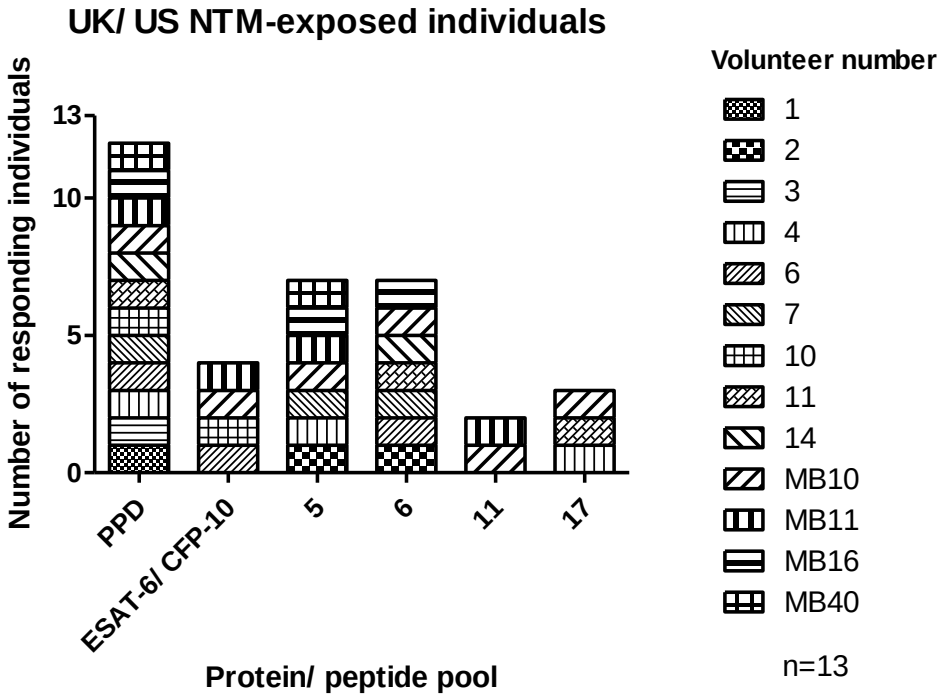
A



B

Figure continued over page

Figure 5 9 ctd



C

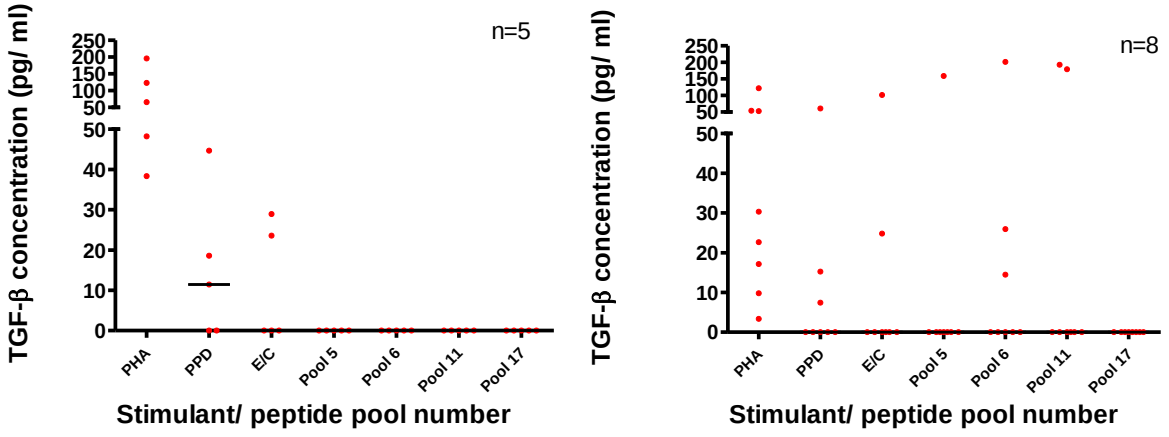
Percentage Ki67 positive proliferating CD4 + lymphocytes in UK and US NTM-exposed individuals (A), UK cord blood samples (B) and number of responding UK and US individuals to each peptide pool (C). Assay cut off = 3 median absolute deviations (MADs) above the median. Responses to pools 5 and 6 were significant, P=0.001 in both cases, Wilcoxon matched pairs test.

5.2.6 Multiplex data

A multiplex cytokine bead-based assay was used to determine concentrations of 11 different Th1 and Th2 cytokines (IFN-γ, IL2, IL4, IL5, IL6, IL 10, IL12, IL1β, TNF-α and TNF-β, in supernatants from PBMC stimulated with antigen for three days (Section 2.8.3.9, Methods). Cryopreserved PBMC samples from all NTM-exposed UK individuals, cord blood samples and

healthy South Africans with known negative responses to QuantiFERON®-TB Gold In-Tube were analysed, where remaining cell numbers permitted. Levels of cytokines detected were low, with cytokines detected to PPD in most cases, but to individual peptide pools in very few cases (data not shown). This included IFN-γ, demonstrating the increased sensitivity of ELISpot and proliferation assays over this methodology. A cytokine bead-based assay was also used to determine concentrations of transforming growth factor beta (TGF-β), which cannot be measured together with the above cytokines as TGF-β must first be acid-activated. The results are shown in Figure 5.10. No TGF-β responses were detected to any of the peptide pools in the UK NTM-exposed individuals, and low level responses were only detected in 25% healthy South African volunteers. Group numbers are reduced as samples were excluded where no response to PHA was detected.

Figure 5-10: TGF-β cytokine bead array results



A B

Results from TGF-β cytokine bead array, showing UK NTM-exposed individuals (A) and healthy South Africans (B). Samples were excluded where there was no response to either PHA or PPD. E/C: ESAT-6/ CFP-10 peptide pool. Differences were not statistically significant (P>0.05, Wilcoxon matched pairs test).

The assay failed in the thawed cord blood samples, as PHA responses in cord blood samples were all low, so these results are not included.

5.3 Discussion

The proliferation assay proved a robust and reproducible method for quantitating PBMC proliferation to a subset of NTM-specific peptide pools. Responses had already been shown to NTM-specific peptide pools 5 and 6 using *ex vivo* ELISpot in healthy South Africans, but not in NTM-exposed individuals from the UK and US. Using a more sensitive proliferation assay, significant responses were demonstrated to these pools in the UK and US cohorts. Non-significant responses were demonstrated to the two pools to which there were borderline responses in the South African cohort (pools 11 and 17). Pool 5 contains proteins from the Mce family of proteins which are known to be antigens in other species of mycobacteria, and pool 6 consisted of hypothetical proteins [146, 167, 168] (see Section 4.3, Chapter 4, Basic Immunology). It is interesting that responses were seen to the same groups of proteins, and this supports the *ex vivo* ELISpot results.

It would have been preferable to have tested all the NTM-specific pools of peptides by both methods, to allow an unbiased comparison of *ex vivo* ELISpot and the proliferation assay, rather than just testing four pools of peptides with the proliferation assay. However, the numbers of remaining PBMC limited the numbers of peptide pools that could be tested by this second method. It would also have been advantageous to have repeated the proliferation assay on samples from the South African cohort in order to compare responses to the two assays, but the method is also expensive for large scale screening, as well as being time consuming.

In the comparison of different proliferation assay protocols, the Oregon Green method was associated with high levels of cell death. This is consistent with published results which suggest that the related dye CFSE is toxic to cells. Given that CFSE also affects proliferation levels and the expression of cytokines and cell surface markers [191], it is possible that the same phenomenon may occur with the use of Oregon Green, and this is another reason why it might not be the optimal method to use.

A multiplex cytokine bead array detected very low levels of cytokine production in response to PPD and to individual NTM-specific peptide pools, looking at 11 different Th1 and Th2 cytokines (data not shown). A TGF-β bead array assay demonstrated low level responses to all of the pools in 25% of healthy South Africans, and no responses to NTM-specific peptide pools in the UK NTM-exposed individuals. Compared to studies of TB-infected patients, the levels of TGF-β detected in response to PPD were also very low [194]. The results from the levels of IFN-γ detection using the 11-plex multiplex assay suggest that IFN-γ ELISpot is more sensitive in the detection of IFN-γ. Secondly, the absence of high levels of the Th2 or regulatory cytokines IL4, IL5, IL10 and TGF-β suggest that the low levels of IFN-γ detected in the ELISpot assay are not explained by high levels of regulatory or Th2 cytokine production.

It should be noted that multiplex bead array data should be interpreted with care, particularly if comparing results between different laboratories or assays. In a large comparison of multiplex kits between different laboratories, using different lots and at different time points during initial HIV viraemia, levels of cytokines detected in serum varied between different kits, lots and within and between laboratories. Levels of agreement between laboratories and between kits

were greater when measuring some cytokines than others. Overall, IL-6, IL-8, IL-10 and TNF-α had good levels of detectability, but IL-6 levels were highly variable between multiplex kit used, and did not correlate with ELISA- based measurements [195]. All of the measurements in this experiment were performed on the same day using the same kit lot, but it is important to bear in mind that comparison of results between laboratories can be problematic.

5.4 Conclusion

In conclusion, a 6 day proliferation assay was used to investigate responses to the NTM-specific peptide pools 5, 6, 11 and 17, to which responses or borderline responses were detected by *ex vivo* ELISpot in the healthy South African cohort. Significant responses were detected to pools 5 and 6 using thawed cells from the UK and US NTM-exposed cohorts, suggesting that the proliferation assay may be more sensitive in this context than *ex vivo* ELISpot. There was no evidence of TGF-β or Th2 cytokine production in response to PBMC stimulation with NTM-specific peptide pools. These results are compatible with the description of NTM-specific peptides, to which statistically significant but low level responses have been demonstrated in UK and US NTM-exposed and South African probably exposed individuals.

6 Discussion

6.1 Summary of results

NTM- specific peptides were defined using bioinformatics, and splenocyte INF-γ responses to these peptides were measured in C57BL6 mice infected with *M. avium* and BCG. INF-γ responses were not detected to any pool of peptides, in either group of mice. INF-γ levels to NTM- specific peptides were also low when measured by ELISpot in PBMC from a cohort of UK individuals with evidence of isolation of NTM from sputum samples. A group of healthy individuals was recruited in the Western Cape, South Africa, and significant INF-γ responses were detected to two pools of peptides, pools 5 and 6. PBMC from the UK individuals and also from a group of US individuals from whom *M. avium* had been isolated from sputum samples were then evaluated using a more sensitive proliferation assay, and responses were again detected to pools 5 and 6, by *ex vivo* ELISpot. These two pools correspond to the mycobacterial cell entry (Mce) proteins and a family of hypothetical proteins. Low responses to a pool of ESAT-6/ CFP-10 peptides were detected in the healthy South African cohort, but there was no correlation between response to NTM-specific peptides and to ESAT-6/ CFP-10 peptide pool when responses from South Africans with known positive responses to QuantiFERON®-TB Gold In-Tube were evaluated.

It is encouraging that responses that appear to be specific to NTM exposure have been described to two pools of NTM-specific peptides. If verified, pools such as these could be used to compare levels of NTM exposure between individuals receiving TB vaccines at different sites,

and also between individuals at the same site, in whom differing levels of immune response to the vaccine are detected. However, the levels of responses detected are low, and these results require further confirmation before peptide pools 5 and 6 could be used as a reliable indicator of NTM exposure.

In this discussion, I will summarise the background of the project and its positive results, followed by the problems and challenges encountered. Finally, the implications of these results on the field are discussed, together with future directions.

6.2 Project overview

BCG has a low efficacy against pulmonary TB in the tropics, possibly as a result of high levels of exposure to NTM occurring in these areas. This is of relevance both to BCG, a vaccine which is likely to remain in widespread use in the tropics, and maybe also to the many new TB vaccines in development. Importantly, the tropical regions where BCG does not work are also those regions where TB disease has the greatest impact. It is also possible that some types of novel TB vaccine may be affected more by this phenomenon than others (for example, for example live vaccines based on BCG or attenuated *M. tuberculosis* may be more affected than sub-unit vaccines which do not replicate) [111].

Human evidence of the effect of NTM exposure on BCG efficacy is generally based on the analysis of immune responses to PPD. Studies have looked at responses to PPD-T in individuals prior to and following BCG vaccination, and have concluded that individuals may be exposed to higher levels of NTM in tropical compared to temperate countries, and that this exposure

appears to correlate with reduced efficacy of BCG vaccination [122]. Other studies have compared the ratio of response to PPD-T to response to PPD from different species of NTM (e.g. PPD from *M. avium* (PPD-A)) as a proxy of *M. tuberculosis* or NTM exposure, and there is some evidence that the ratio of responses to the different PPDs may give an indirect indication of NTM exposure from *M. tuberculosis* or other species of NTM [125, 126]. Defining antigens which are present in common NTM but absent from all members of the *M. tuberculosis* complex would allow further investigation of the relationship between immune responses to NTM, suggestive of NTM exposure, with TB vaccine efficacy.

In this study, NTM common to South Africa and the UK (where the two study sites were located) were defined, and a stringent methodology was used to define protein families which were present in NTM but absent from the *M. tuberculosis* complex. After the sequencing of a previously un-sequenced NTM genome and the prediction of protein sequences from un-annotated genome sequences, BLASTP was used to define a subset of proteins thought to be present in at least four common NTM species and absent from the *M. tuberculosis* complex. The subsequent use of TBLASTN ensured that the final selection of proteins absent from the *M. tuberculosis* complex was independent of genome annotation, the quality of which has been shown to be highly variable [152, 154]. The final BLASTP comparing short 20-mer sequences against the whole bacterial reference protein database provided an additional specificity filter, allowing the exclusion of peptides with homology to common bacterial species, and providing an additional check that there were no hits with reference proteins from members of the *M. tuberculosis* complex.

Proteins with a LocateP prediction to be secreted were selected on the basis that the most immunogenic of known mycobacterial proteins belong to this group of proteins. Protein location predictions were only available for reference proteins, and not for protein sequences which had been predicted using the programme Prodigal. However, the use of BlastClust allowed assumptions to be made about the locations of predicted protein sequences as a consequence of the reference proteins they clustered with. This is the first time a bioinformatics pipeline has incorporated this range of steps. The aim was to reduce the chances of unpredicted cross-reactivity, as has been noted to occur in other antigen discovery pipelines [151, 159, 160].

The proteins were first tested, in the form of overlapping 20-mer peptides, in a mouse model of *M. avium* infection, using BCG infection as a negative control. Significant splenocyte IFN-γ responses were detected to the control antigens PPD-T, PPD-A and a pool of overlapping Antigen 85A peptides in both groups of mice. However, no responses were detected to the NTM-specific peptides. The robust responses seen to the control antigens argues against a significant problem with the mouse model itself, but it is possible that responses to the NTM-specific proteins could be sub-dominant, whereas those to Antigen 85A and PPD are likely to be dominant. Therefore, the lack of responses detected in the mouse model could relate to the timing of mycobacterial infection and culling (sub-dominant responses peak early and over time become dominated by dominant responses [176]).

In humans, significant IFN-γ T cell responses to pools 5 and 6 of the NTM-specific peptides were detected from healthy volunteers in South Africa using *ex vivo* ELISpot. Similar responses were

not detected to any NTM-specific peptides pools using cryopreserved cells from UK and US individuals who had documented retrieval of NTM from sputum samples. However, the use of a 6 day proliferation assay demonstrated significant proliferative responses to the same two pools of peptides in this group of volunteers. No responses were detected by either method in cord blood samples, which are thought to have a low likelihood of either NTM or TB exposure. The magnitude of responses detected in all groups of responding volunteers was low, and possible explanations for this finding are discussed below.

To investigate the possibility of cross-reactivity between the NTM-specific peptide pools and *M. tuberculosis,* healthy South Africans with known positive responses to QuantiFERON®-TB Gold In-Tube test were recruited. There was no correlation between response to pool 5 or 6 (or any other pool tested) of the NTM specific peptides and to the ESAT-6/ CFP-10 peptide pool, supporting the suggestion that these peptides may be specific to NTM, and not present in the *M. tuberculosis* complex.

PBMC from healthy South Africans and from individuals from the UK with definite NTM exposure were evaluated using a multiplex cytokine bead-based assay, which demonstrated very low levels of 11 Th1 and Th2 cytokines in supernatants from PBMC stimulated for four days with NTM-specific peptides. A TGF-β bead-based assay also demonstrated low levels of the regulatory cytokine TGF-β in the same supernatants. These results suggest that the low levels of the Th1 cytokine INF-γ are not explained by the inhibitory effects of high levels of Th2 or regulatory cytokines.

6.3 Problems and challenges

The main possible reasons for the low levels of immune responses detected in response to stimulation with the NTM-specific peptides pools can be summarised in three main categories.

- Difficulty in defining NTM exposure in healthy individuals, and problems with low immune responses in an unhealthy 'definitely exposed' group.
- The immunogenicity of NTM.
- The immunogenicity of NTM-specific peptides as contrasted with peptides or proteins which are shared between NTM and the *M. tuberculosis* family.

6.3.1 Defining NTM exposure

The main challenge in the design of this study was in the definition of control groups on whom to test the peptides. The ideal groups would have included a group of healthy individuals with high level exposure to NTM and no possibility of TB exposure, a group of individuals with exposure to other infections, in particular *M. tuberculosis*, but no exposure to NTM, and a group with no possibility of exposure to either NTM or *M. tuberculosis.*

In practice it is very difficult to define such groups, in particular because most areas of high NTM exposure are also areas of high TB transmission. I therefore recruited a group of individuals with definite NTM exposure, as demonstrated by the isolation of NTM from sputum samples on at least two occasions, in the UK, where TB transmission levels are low. Unfortunately, since NTM only cause infections in individuals with significant levels of either

immunosuppression or lung pathology, this group was not healthy, which probably explains the low levels of immune responses detected to NTM-specific peptides and to control antigens.

A second cohort of patients was therefore recruited from a region where high levels of NTM exposure had previously been demonstrated [28, 121, 130]. This cohort consisted of healthy individuals, and testing was performed on fresh PBMC, in which immune responses detected are greater [181]. The major disadvantage with this cohort was that neither the timing nor the dose of NTM exposure could be determined, nor whether any individual was actually exposed to NTM. In addition, it is possible that some of the individuals recruited had been exposed to *M. tuberculosis,* so a pool of ESAT-6/ CFP-10 peptides was included as a negative control.

6.3.2 The immunogenicity of NTM

Although robust immune responses which have been attributed to NTM exposure have been described in studies looking at immune responses to PPD, it is possible that these responses actually result from previous exposure to a member of the *M. tuberculosis* complex such as *M. tuberculosis* or BCG, and that NTM themselves are not highly immunogenic. It is difficult to differentiate between exposure to *M. tuberculosis,* BCG and NTM using PPD.

6.3.3 The immunogenicity of NTM-specific peptides which are absent from the *M. tuberculosis* complex

It is also possible that NTM themselves are immunogenic, but only as a result of proteins shared between NTM and the TB complex of organisms, and that proteins which are present in NTM but not in the *M. tuberculosis* complex are simply not immunogenic. There are some plausible

reasons to explain this possibility. There is no doubt that robust immune responses can be detected to PPD from *M. avium* and from other species of NTM. It may be that the responses detected in assays using PPD-A are predominantly to antigens which are shared between NTM and the *M. tuberculosis* complex, and that those antigens specific to NTM and not present in the *M. tuberculosis* complex are not immunogenic. This possibility is credible; there is an evolutionary advantage in developing a robust immune response to *M. tuberculosis* antigens (and therefore to the numerous antigens which are shared between *M. tuberculosis* and NTM), but there is a less clear advantage to having a robust immune response specifically to non-pathogenic NTM. It is possible that the NTM-specific proteins are predominantly concerned with functions which *M. tuberculosis* does not possess, such as free-living in the environment, and also that this subset of proteins may only be expressed under certain conditions and not *in vivo* in humans.

6.4 Future directions

There are several ways in which the positive findings from this study could be validated, and this research taken further. Firstly, responses to NTM-specific peptides could be investigated in a large group of healthy individuals (for example, using cohorts from one of the Gates diagnostic antigens projects), in regions where NTM exposure is high. Since the prevalence of TB exposure is high in tropical regions, where NTM exposure is also thought to be high, TB-exposed individuals could be excluded using responses to the highly sensitive and (relatively) *M. tuberculosis*-specific proteins ESAT-6 and CFP-10 [14]. Ideally, a group with low risk of exposure to *M. tuberculosis* would be defined by epidemiological rather than by immunogical means,

but at present it is hard to see how this group could better be defined. Negative control groups could include individuals with ESAT-6 and CFP-10 responses suggestive of previous *M. tuberculosis* infection, and also with acute bacterial and viral infections and autoimmune diseases.

Secondly, modifications to the NTM-specific peptide pools could potentially result in the detection of larger IFN-γ responses in NTM-exposed individuals. It is likely that the immunogenic pools 5 and 6 contain a minority of peptides to which a IFN-γ response can be measured in NTM-exposed individuals, and a majority which elicit no response. Removing 'non-responding' peptides could improve the magnitude of responses detected, in particular because some peptides present in the pools may result in the inhibition of responses [178]. Testing the constituent peptides either individually or in a matrix may allow the identification of these responding peptides. A smaller pool size would have the additional advantage of containing a lower concentration of DMSO, which may be associated with a reduction in immune response detected [179].

Following optimisation of a pool of NTM-specific peptides and validation in an alternative population, the variability in immune response which is seen to the vaccine MVA85A could be investigated. Immune responses to NTM could be characterised in stored PBMC samples from individuals receiving the vaccine in South Africa, where responses were seen to be smaller, and compared with responses to NTM in stored PBMC from UK subjects. Likewise, since levels of response are variable within each trial site, responses to NTM could be compared between 'high' and 'low' responders to the vaccine within each site. Baseline PBMC samples would be

used in all cases, corresponding to a time point prior to the administration of the vaccine MVA85A, and compared with the peak response to the vaccine, which is seen 7 days after vaccination.

In conclusion, it is possible that testing an improved selection of peptides from the promising pools identified in this study on larger cohorts of individuals in a high NTM-prevalence area may result in the definition of NTM-specific peptides. Such peptides could then be used to study the influence of NTM exposure on the efficacy of the BCG vaccine, and also on the immunogenicity of new TB vaccines. Until peptides have been defined which are specifically associated with NTM exposure, it will remain difficult to further develop the theory that exposure to NTM negatively influences the efficacy of the BCG vaccine, and to investigate the possibility that this mechanism may also affect some or all of the new TB vaccines currently in development.

References

1. World Health Organisation, *Global Tuberculosis Control 2010*. 2010.
2. World Health Organisation, *The Stop TB Strategy*. 2010.
3. World Health Organisation, *Multidrug and extensively drug-resistant TB (M/XDR-TB) 2010 Global report on surveillance and response*. 2010.
4. Velayati, A.A., et al., *Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran.* Chest, 2009. **136**(2): p. 420-5.
5. Lawn, S.D. and A.I. Zumla, *Tuberculosis.* Lancet, 2011.
6. Walzl, G., et al., *Immunological biomarkers of tuberculosis.* Nat Rev Immunol, 2011. **11**(5): p. 343-54.
7. Barry, C.E., 3rd, et al., *The spectrum of latent tuberculosis: rethinking the biology and intervention strategies.* Nat Rev Microbiol, 2009. **7**(12): p. 845-55.
8. World Health Organisation, *Global tuberculosis control.* 2009.
9. Houben, R.M., et al., *Human immunodeficiency virus associated tuberculosis more often due to recent infection than reactivation of latent infection.* Int J Tuberc Lung Dis, 2011. **15**(1): p. 24-31.
10. Churchyard, G.J., et al., *Mycobacterial disease in South African gold miners in the era of HIV infection.* Int J Tuberc Lung Dis, 1999. **3**(9): p. 791-8.
11. Marais, B.J., et al., *A refined symptom-based approach to diagnose pulmonary tuberculosis in children.* Pediatrics, 2006. **118**(5): p. e1350-9.
12. Schluger, N.W. and J. Burzynski, *Recent advances in testing for latent TB.* Chest, 2010. **138**(6): p. 1456-63.
13. Elzi, L., et al., *Reducing tuberculosis incidence by tuberculin skin testing, preventive treatment, and antiretroviral therapy in an area of low tuberculosis transmission.* Clin Infect Dis, 2007. **44**(1): p. 94-102.
14. Ewer, K., et al., *Comparison of T-cell-based assay with tuberculin skin test for diagnosis of Mycobacterium tuberculosis infection in a school tuberculosis outbreak.* Lancet, 2003. **361**(9364): p. 1168-73.
15. Ling, D.I., et al., *Are interferon-{gamma} release assays useful for active tuberculosis in a high-burden setting?* Eur Respir J, 2011.
16. Mahomed, H., et al., *Predictive factors for latent tuberculosis infection among adolescents in a high-burden area in South Africa.* Int J Tuberc Lung Dis, 2011. **15**(3): p. 331-6.
17. Kim, B.J., et al., *Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (rpoB).* J Clin Microbiol, 1999. **37**(6): p. 1714-20.
18. Flynn, J.L., J. Chan, and P.L. Lin, *Macrophages and control of granulomatous inflammation in tuberculosis.* Mucosal Immunol, 2011. **4**(3): p. 271-8.
19. Kusner, D.J., *Mechanisms of mycobacterial persistence in tuberculosis.* Clin Immunol, 2005. **114**(3): p. 239-47.

20. Cooper, A.M., et al., *Disseminated tuberculosis in interferon gamma gene-disrupted mice.* J Exp Med, 1993. **178**(6): p. 2243-7.
21. Ottenhoff, T.H., et al., *Control of human host immunity to mycobacteria.* Tuberculosis (Edinb), 2005. **85**(1-2): p. 53-64.
22. Alcais, A., et al., *Tuberculosis in children and adults: two distinct genetic diseases.* J Exp Med, 2005. **202**(12): p. 1617-21.
23. de Jong, R., et al., *Severe mycobacterial and Salmonella infections in interleukin-12 receptor-deficient patients.* Science, 1998. **280**(5368): p. 1435-8.
24. Flynn, J.L., *Immunology of tuberculosis and implications in vaccine development.* Tuberculosis (Edinb), 2004. **84**(1-2): p. 93-101.
25. Cooper, A.M., et al., *Interleukin 12 (IL-12) is crucial to the development of protective immunity in mice intravenously infected with mycobacterium tuberculosis.* J Exp Med, 1997. **186**(1): p. 39-45.
26. Flynn, J.L., et al., *IL-12 increases resistance of BALB/c mice to Mycobacterium tuberculosis infection.* J Immunol, 1995. **155**(5): p. 2515-24.
27. Solovic, I., et al., *The risk of tuberculosis related to tumour necrosis factor antagonist therapies: a TBNET consensus statement.* Eur Respir J, 2010. **36**(5): p. 1185-206.
28. Corbett, E.L., et al., *The growing burden of tuberculosis: global trends and interactions with the HIV epidemic.* Arch Intern Med, 2003. **163**(9): p. 1009-21.
29. Lawn, S.D., et al., *Short-term and long-term risk of tuberculosis associated with CD4 cell recovery during antiretroviral therapy in South Africa.* AIDS, 2009. **23**(13): p. 1717-25.
30. Madhan Kumar, M. and A. Raja, *Cytotoxicity responses to selected ESAT-6 and CFP-10 peptides in tuberculosis.* Cell Immunol, 2010. **265**(2): p. 146-55.
31. Ho, P., et al., *Mycobacterium chelonae sensitisation induces CD4(+)-mediated cytotoxicity against BCG.* Eur J Immunol, 2009. **39**(7): p. 1841-9.
32. Lazarevic, V., D. Nolt, and J.L. Flynn, *Long-term control of Mycobacterium tuberculosis infection is mediated by dynamic immune responses.* J Immunol, 2005. **175**(2): p. 1107-17.
33. Sad, S., D. Kagi, and T.R. Mosmann, *Perforin and Fas killing by CD8+ T cells limits their cytokine synthesis and proliferation.* J Exp Med, 1996. **184**(4): p. 1543-7.
34. Flynn, J.L., et al., *Major histocompatibility complex class I-restricted T cells are required for resistance to Mycobacterium tuberculosis infection.* Proc Natl Acad Sci U S A, 1992. **89**(24): p. 12013-7.
35. Tully, G., et al., *Highly focused T cell responses in latent human pulmonary Mycobacterium tuberculosis infection.* J Immunol, 2005. **174**(4): p. 2174-84.
36. Khader, S.A., et al., *IL-23 and IL-17 in the establishment of protective pulmonary CD4+ T cell responses after vaccination and during Mycobacterium tuberculosis challenge.* Nat Immunol, 2007. **8**(4): p. 369-77.
37. Okamoto Yoshida, Y., et al., *Essential role of IL-17A in the formation of a mycobacterial infection-induced granuloma in the lung.* J Immunol, 2010. **184**(8): p. 4414-22.
38. Khader, S.A. and A.M. Cooper, *IL-23 and IL-17 in tuberculosis.* Cytokine, 2008. **41**(2): p. 79-83.
39. Vordermeier, H.M., et al., *Viral booster vaccines improve Mycobacterium bovis BCG-induced protection against bovine tuberculosis.* Infect Immun, 2009. **77**(8): p. 3364-73.

40. Scriba, T.J., et al., *Distinct, specific IL-17- and IL-22-producing CD4+ T cell subsets contribute to the human anti-mycobacterial immune response.* J Immunol, 2008. **180**(3): p. 1962-70.
41. de Cassan, S.C., et al., *Investigating the induction of vaccine-induced Th17 and regulatory T cells in healthy, Mycobacterium bovis BCG-immunized adults vaccinated with a new tuberculosis vaccine, MVA85A.* Clin Vaccine Immunol, 2010. **17**(7): p. 1066-73.
42. Sakaguchi, S., et al., *FOXP3+ regulatory T cells in the human immune system.* Nat Rev Immunol, 2010. **10**(7): p. 490-500.
43. Green, A.M., et al., *CD4(+) regulatory T cells in a cynomolgus macaque model of Mycobacterium tuberculosis infection.* J Infect Dis, 2010. **202**(4): p. 533-41.
44. Hougardy, J.M., et al., *Regulatory T cells depress immune responses to protective antigens in active tuberculosis.* Am J Respir Crit Care Med, 2007. **176**(4): p. 409-16.
45. Darrah, P.A., et al., *Multifunctional TH1 cells define a correlate of vaccine-mediated protection against Leishmania major.* Nat Med, 2007. **13**(7): p. 843-50.
46. Wilkinson, K.A. and R.J. Wilkinson, *Polyfunctional T cells in human tuberculosis.* Eur J Immunol, 2010. **40**(8): p. 2139-42.
47. Betts, M.R., et al., *HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells.* Blood, 2006. **107**(12): p. 4781-9.
48. Owen, R.E., et al., *HIV+ elite controllers have low HIV-specific T-cell activation yet maintain strong, polyfunctional T-cell responses.* AIDS, 2010. **24**(8): p. 1095-105.
49. Beveridge, N.E., et al., *Immunisation with BCG and recombinant MVA85A induces long-lasting, polyfunctional Mycobacterium tuberculosis-specific CD4+ memory T lymphocyte populations.* Eur J Immunol, 2007. **37**(11): p. 3089-100.
50. Scriba, T.J., et al., *Modified vaccinia Ankara-expressing Ag85A, a novel tuberculosis vaccine, is safe in adolescents and children, and induces polyfunctional CD4+ T cells.* Eur J Immunol, 2010. **40**(1): p. 279-90.
51. Abel, B., et al., *The novel tuberculosis vaccine, AERAS-402, induces robust and polyfunctional CD4+ and CD8+ T cells in adults.* Am J Respir Crit Care Med, 2010. **181**(12): p. 1407-17.
52. Caccamo, N., et al., *Multifunctional CD4(+) T cells correlate with active Mycobacterium tuberculosis infection.* Eur J Immunol, 2010. **40**(8): p. 2211-20.
53. Plattner, B.L., R.T. Doyle, and J.M. Hostetter, *Gamma-delta T cell subsets are differentially associated with granuloma development and organization in a bovine model of mycobacterial disease.* Int J Exp Pathol, 2009. **90**(6): p. 587-97.
54. Meraviglia, S., et al., *gammadelta T cells cross-link innate and adaptive immunity in Mycobacterium tuberculosis infection.* Clin Dev Immunol, 2011. **2011**: p. 587315.
55. Sutton, C.E., et al., *Interleukin-1 and IL-23 induce innate IL-17 production from gammadelta T cells, amplifying Th17 responses and autoimmunity.* Immunity, 2009. **31**(2): p. 331-41.
56. Berry, M.P., et al., *An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis.* Nature, 2010. **466**(7309): p. 973-7.

57. Maglione, P.J., J. Xu, and J. Chan, *B cells moderate inflammatory progression and enhance bacterial containment upon pulmonary challenge with Mycobacterium tuberculosis.* J Immunol, 2007. **178**(11): p. 7222-34.
58. Kagina, B.M., et al., *Delaying BCG vaccination from birth to 10 weeks of age may result in an enhanced memory CD4 T cell response.* Vaccine, 2009. **27**(40): p. 5488-95.
59. Kagina, B.M., et al., *Specific T cell frequency and cytokine expression profile do not correlate with protection against tuberculosis after bacillus Calmette-Guerin vaccination of newborns.* Am J Respir Crit Care Med, 2010. **182**(8): p. 1073-9.
60. Cheon, S.H., et al., *Bactericidal activity in whole blood as a potential surrogate marker of immunity after vaccination against tuberculosis.* Clin Diagn Lab Immunol, 2002. **9**(4): p. 901-7.
61. Kampmann, B., et al., *Novel human in vitro system for evaluating antimycobacterial vaccines.* Infect Immun, 2004. **72**(11): p. 6401-7.
62. World Health Organisation. *BCG Vaccine, WHO position paper*. 2004; Available from: http://www.who.int/immunization/wer7904BCG_Jan04_position_paper.pdf.
63. Hart, P.D. and I. Sutherland, *BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life.* Br Med J, 1977. **2**(6082): p. 293-5.
64. Ponnighaus, J.M., et al., *Efficacy of BCG vaccine against leprosy and tuberculosis in northern Malawi.* Lancet, 1992. **339**(8794): p. 636-9.
65. Colditz, G.A., et al., *Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature.* JAMA, 1994. **271**(9): p. 698-702.
66. Palmer, C.E., L.W. Shaw, and G.W. Comstock, *Community trials of BCG vaccination.* Am Rev Tuberc, 1958. **77**(6): p. 877-907.
67. Sutherland, I. and V.H. Springett, *Effectiveness of BCG vaccination in England and Wales in 1983.* Tubercle, 1987. **68**(2): p. 81-92.
68. Aronson, N.E., et al., *Long-term efficacy of BCG vaccine in American Indians and Alaska Natives: A 60-year follow-up study.* JAMA, 2004. **291**(17): p. 2086-91.
69. Comstock, G.W., *Identification of an effective vaccine against tuberculosis.* Am Rev Respir Dis, 1988. **138**(2): p. 479-80.
70. Baily, G.V., *Tuberculosis prevention Trial, Madras.* Indian J Med Res, 1980. **72 Suppl**: p. 1-74.
71. Trunz, B.B., P. Fine, and C. Dye, *Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness.* Lancet, 2006. **367**(9517): p. 1173-80.
72. Rodrigues, L.C., V.K. Diwan, and J.G. Wheeler, *Protective effect of BCG against tuberculous meningitis and miliary tuberculosis: a meta-analysis.* Int J Epidemiol, 1993. **22**(6): p. 1154-8.
73. al-Kassimi, F.A., et al., *Does the protective effect of neonatal BCG correlate with vaccine-induced tuberculin reaction?* Am J Respir Crit Care Med, 1995. **152**(5 Pt 1): p. 1575-8.
74. Hesseling, A.C., et al., *BCG vaccination in South African HIV-exposed infants--risks and benefits.* S Afr Med J, 2009. **99**(2): p. 88-91.
75. Hesseling, A.C., et al., *Disseminated bacille Calmette-Guerin disease in HIV-infected South African infants.* Bull World Health Organ, 2009. **87**(7): p. 505-11.

76. *Revised BCG vaccination guidelines for infants at risk for HIV infection.* Wkly Epidemiol Rec, 2007. **82**(21): p. 193-6.
77. Barreto, M.L., et al., *Evidence of an effect of BCG revaccination on incidence of tuberculosis in school-aged children in Brazil: Second report of the BCG-REVAC cluster-randomised trial.* Vaccine, 2011. **29**(31): p. 4875-7.
78. Roth, A.E., et al., *Effect of revaccination with BCG in early childhood on mortality: randomised trial in Guinea-Bissau.* BMJ, 2010. **340**: p. c671.
79. Cunha, S.S., et al., *BCG revaccination does not protect against leprosy in the Brazilian Amazon: a cluster randomised trial.* PLoS Negl Trop Dis, 2008. **2**(2): p. e167.
80. Zwerling, A., et al., *The BCG World Atlas: a database of global BCG vaccination policies and practices.* PLoS Med, 2011. **8**(3): p. e1001012.
81. Wilson, M.E., H.V. Fineberg, and G.A. Colditz, *Geographic latitude and the efficacy of bacillus Calmette-Guerin vaccine.* Clin Infect Dis, 1995. **20**(4): p. 982-91.
82. Abebe, F. and G. Bjune, *The emergence of Beijing family genotypes of Mycobacterium tuberculosis and low-level protection by bacille Calmette-Guerin (BCG) vaccines: is there a link?* Clin Exp Immunol, 2006. **145**(3): p. 389-97.
83. World Health Organisation, *Issues relating to the use of BCG in immunization programmes*. 1999.
84. Brosch, R., et al., *Genome plasticity of BCG and impact on vaccine efficacy.* Proc Natl Acad Sci U S A, 2007. **104**(13): p. 5596-601.
85. Kozak, R. and M.A. Behr, *Divergence of immunologic and protective responses of different BCG strains in a murine model.* Vaccine, 2011. **29**(7): p. 1519-26.
86. *Randomised controlled trial of single BCG, repeated BCG, or combined BCG and killed Mycobacterium leprae vaccine for prevention of leprosy and tuberculosis in Malawi. Karonga Prevention Trial Group.* Lancet, 1996. **348**(9019): p. 17-24.
87. Mahomed, H., et al., *The impact of a change in bacille Calmette-Guerin vaccine policy on tuberculosis incidence in children in Cape Town, South Africa.* Pediatr Infect Dis J, 2006. **25**(12): p. 1167-72.
88. Hawkridge, A., et al., *Efficacy of percutaneous versus intradermal BCG in the prevention of tuberculosis in South African infants: randomised trial.* BMJ, 2008. **337**: p. a2052.
89. Gorak-Stolinska, P., et al., *Immunogenicity of Danish-SSI 1331 BCG vaccine in the UK: comparison with Glaxo-Evans 1077 BCG vaccine.* Vaccine, 2006. **24**(29-30): p. 5726-33.
90. Brahmajothi, V., et al., *Association of pulmonary tuberculosis and HLA in south India.* Tubercle, 1991. **72**(2): p. 123-32.
91. Bellamy, R., et al., *Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans.* N Engl J Med, 1998. **338**(10): p. 640-4.
92. Goldfeld, A.E., et al., *Association of an HLA-DQ allele with clinical tuberculosis.* JAMA, 1998. **279**(3): p. 226-8.
93. Gagneux, S., et al., *Variable host-pathogen compatibility in Mycobacterium tuberculosis.* Proc Natl Acad Sci U S A, 2006. **103**(8): p. 2869-73.
94. Tsenova, L., et al., *Virulence of selected Mycobacterium tuberculosis clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the bacilli.* J Infect Dis, 2005. **192**(1): p. 98-106.

95. Manca, C., et al., *Virulence of a Mycobacterium tuberculosis clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha /beta.* Proc Natl Acad Sci U S A, 2001. **98**(10): p. 5752-7.
96. Cowley, D., et al., *Recent and rapid emergence of W-Beijing strains of Mycobacterium tuberculosis in Cape Town, South Africa.* Clin Infect Dis, 2008. **47**(10): p. 1252-9.
97. Hesseling, A.C., et al., *Mycobacterial genotype is associated with disease phenotype in children.* Int J Tuberc Lung Dis, 2010. **14**(10): p. 1252-8.
98. Kong, Y., et al., *Association between Mycobacterium tuberculosis Beijing/W lineage strain infection and extrathoracic tuberculosis: Insights from epidemiologic and clinical characterization of the three principal genetic groups of M. tuberculosis clinical isolates.* J Clin Microbiol, 2007. **45**(2): p. 409-14.
99. Lan, N.T., et al., *Mycobacterium tuberculosis Beijing genotype and risk for treatment failure and relapse, Vietnam.* Emerg Infect Dis, 2003. **9**(12): p. 1633-5.
100. Sun, Y.J., et al., *Association of Mycobacterium tuberculosis Beijing genotype with tuberculosis relapse in Singapore.* Epidemiol Infect, 2006. **134**(2): p. 329-32.
101. Thwaites, G., et al., *Relationship between Mycobacterium tuberculosis genotype and the clinical phenotype of pulmonary and meningeal tuberculosis.* J Clin Microbiol, 2008. **46**(4): p. 1363-8.
102. Grode, L., et al., *Increased vaccine efficacy against tuberculosis of recombinant Mycobacterium bovis bacille Calmette-Guerin mutants that secrete listeriolysin.* J Clin Invest, 2005. **115**(9): p. 2472-9.
103. Tsenova, L., et al., *BCG vaccination confers poor protection against M. tuberculosis HN878-induced central nervous system disease.* Vaccine, 2007. **25**(28): p. 5126-32.
104. Martineau, A.R., et al., *High-dose vitamin D(3) during intensive-phase antimicrobial treatment of pulmonary tuberculosis: a double-blind randomised controlled trial.* Lancet, 2011. **377**(9761): p. 242-50.
105. Lalor, M.K., et al., *BCG vaccination: a role for vitamin D?* PLoS One, 2011. **6**(1): p. e16709.
106. Martineau, A.R., et al., *IFN-gamma- and TNF-independent vitamin D-inducible human suppression of mycobacteria: the role of cathelicidin LL-37.* J Immunol, 2007. **178**(11): p. 7190-8.
107. Martineau, A.R., et al., *A single dose of vitamin D enhances immunity to mycobacteria.* Am J Respir Crit Care Med, 2007. **176**(2): p. 208-13.
108. Chilima, B.Z., et al., *Distribution of environmental mycobacteria in Karonga District, northern Malawi.* Appl Environ Microbiol, 2006. **72**(4): p. 2343-50.
109. Cooper, A.M., *T cells in mycobacterial infection and disease.* Curr Opin Immunol, 2009. **21**(4): p. 378-84.
110. Rhoades, E.R., A.A. Frank, and I.M. Orme, *Progression of chronic pulmonary tuberculosis in mice aerogenically infected with virulent Mycobacterium tuberculosis.* Tuber Lung Dis, 1997. **78**(1): p. 57-66.
111. Brandt, L., et al., *Failure of the Mycobacterium bovis BCG vaccine: some species of environmental mycobacteria block multiplication of BCG and induction of protective immunity to tuberculosis.* Infect Immun, 2002. **70**(2): p. 672-8.

112. Flaherty, D.K., et al., *Exposure to Mycobacterium avium can modulate established immunity against Mycobacterium tuberculosis infection generated by Mycobacterium bovis BCG vaccination.* J Leukoc Biol, 2006. **80**(6): p. 1262-71.
113. Orme, I.M., A.R. Roberts, and F.M. Collins, *Lack of evidence for a reduction in the efficacy of subcutaneous BCG vaccination in mice infected with nontuberculous mycobacteria.* Tubercle, 1986. **67**(1): p. 41-6.
114. de Lisle, G.W., et al., *The efficacy of live tuberculosis vaccines after presensitization with Mycobacterium avium.* Tuberculosis (Edinb), 2005. **85**(1-2): p. 73-9.
115. *Trial of BCG vaccines in south India for tuberculosis prevention.* Indian J Med Res, 1979. **70**: p. 349-63.
116. Edwards, M.L., et al., *Infection with Mycobacterium avium-intracellulare and the protective effects of Bacille Calmette-Guerin.* J Infect Dis, 1982. **145**(5): p. 733-41.
117. Hernandez-Pando, R., et al., *Pathogenesis of tuberculosis in mice exposed to low and high doses of an environmental mycobacterial saprophyte before infection.* Infect Immun, 1997. **65**(8): p. 3317-27.
118. Demangel, C., et al., *Differential effects of prior exposure to environmental mycobacteria on vaccination with Mycobacterium bovis BCG or a recombinant BCG strain expressing RD1 antigens.* Infect Immun, 2005. **73**(4): p. 2190-6.
119. Young, S.L., et al., *Environmental strains of Mycobacterium avium interfere with immune responses associated with Mycobacterium bovis BCG vaccination.* Infect Immun, 2007. **75**(6): p. 2833-40.
120. Orme, I.M. and F.M. Collins, *Efficacy of Mycobacterium bovis BCG vaccination in mice undergoing prior pulmonary infection with atypical mycobacteria.* Infect Immun, 1984. **44**(1): p. 28-32.
121. Hatherill, M., et al., *Isolation of non-tuberculous mycobacteria in children investigated for pulmonary tuberculosis.* PLoS One, 2006. **1**: p. e21.
122. Black, G.F., et al., *BCG-induced increase in interferon-gamma response to mycobacterial antigens and efficacy of BCG vaccination in Malawi and the UK: two randomised controlled studies.* Lancet, 2002. **359**(9315): p. 1393-401.
123. Burl, S., et al., *The tuberculin skin test (TST) is affected by recent BCG vaccination but not by exposure to non-tuberculosis mycobacteria (NTM) during early life.* PLoS One, 2010. **5**(8): p. e12287.
124. Mahaba, H., N. Ismail, and M. El Teheawy, *Evaluation of BCG vaccination: an intervention study in Rail Region, Saudi Arabia.* J Egypt Public Health Assoc, 1995. **70**(5-6): p. 485-95.
125. Farhat, M., et al., *False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria?* Int J Tuberc Lung Dis, 2006. **10**(11): p. 1192-204.
126. Weir, R.E., et al., *Interferon-gamma and skin test responses of schoolchildren in southeast England to purified protein derivatives from Mycobacterium tuberculosis and other species of mycobacteria.* Clin Exp Immunol, 2003. **134**(2): p. 285-94.
127. Buddle, B.M., et al., *Update on vaccination of cattle and wildlife populations against tuberculosis.* Vet Microbiol, 2011.
128. Corbett, E.L., et al., *Nontuberculous mycobacteria: defining disease in a prospective cohort of South African miners.* Am J Respir Crit Care Med, 1999. **160**(1): p. 15-21.

129. Corbett, E.L., et al., *Mycobacterium kansasii and M. scrofulaceum isolates from HIV-negative South African gold miners: incidence, clinical significance and radiology.* Int J Tuberc Lung Dis, 1999. **3**(6): p. 501-7.
130. September, S.M., V.S. Brozel, and S.N. Venter, *Diversity of nontuberculoid Mycobacterium species in biofilms of urban and semiurban drinking water distribution systems.* Appl Environ Microbiol, 2004. **70**(12): p. 7571-3.
131. Donoghue, H.D., E. Overend, and J.L. Stanford, *A longitudinal study of environmental mycobacteria on a farm in south-west England.* J Appl Microbiol, 1997. **82**(1): p. 57-67.
132. Pickup, R.W., et al., *Mycobacterium avium subsp. paratuberculosis in lake catchments, in river water abstracted for domestic use, and in effluent from domestic sewage treatment works: diverse opportunities for environmental cycling and human exposure.* Appl Environ Microbiol, 2006. **72**(6): p. 4067-77.
133. Yates, M.D., A. Pozniak, and J.M. Grange, *Isolation of mycobacteria from patients seropositive for the human immunodeficiency virus (HIV) in south east England: 1984-92.* Thorax, 1993. **48**(10): p. 990-5.
134. Martin-Casabona, N., et al., *Non-tuberculous mycobacteria: patterns of isolation. A multi-country retrospective survey.* Int J Tuberc Lung Dis, 2004. **8**(10): p. 1186-93.
135. Hyatt, D., et al., *Prodigal: prokaryotic gene recognition and translation initiation site identification.* BMC Bioinformatics, 2010. **11**: p. 119.
136. Villegas, A. and A.M. Kropinski, *An analysis of initiation codon utilization in the Domain Bacteria - concerns about the quality of bacterial genome annotation.* Microbiology, 2008. **154**(Pt 9): p. 2559-661.
137. Madden, T., *The BLAST sequence analysis tool*, in *The NCBI Handbook* NCBI, Editor. 2002, National Library of Medicine: Bethseda.
138. Zhou, M., et al., *LocateP: genome-scale subcellular-location predictor for bacterial proteins.* BMC Bioinformatics, 2008. **9**: p. 173.
139. McShane, H., et al., *Enhanced immunogenicity of CD4(+) t-cell responses and protective efficacy of a DNA-modified vaccinia virus Ankara prime-boost vaccination regimen for murine tuberculosis.* Infect Immun, 2001. **69**(2): p. 681-6.
140. Soares, A., et al., *Novel application of Ki67 to quantify antigen-specific in vitro lymphoproliferation.* J Immunol Methods, 2010. **362**(1-2): p. 43-50.
141. Day, C.L., et al., *Functional Capacity of Mycobacterium tuberculosis-Specific T Cell Responses in Humans Is Associated with Mycobacterial Load.* J Immunol, 2011.
142. Koh, W.J., et al., *Pulmonary TB and NTM lung disease: comparison of characteristics in patients with AFB smear-positive sputum.* Int J Tuberc Lung Dis, 2006. **10**(9): p. 1001-7.
143. NCBI. *ftp://ftp.ncbi.nih.gov/genomes/Bacteria/*. 2011.
144. Sanger, *http://www.sanger.ac.uk/*. 2011.
145. Covaris. 2011; Available from: http://www.covarisinc.com/dna-shearing.html.
146. Xu, G., et al., *Effect of recombinant Mce4A protein of Mycobacterium bovis on expression of TNF-alpha, iNOS, IL-6, and IL-12 in bovine alveolar macrophages.* Mol Cell Biochem, 2007. **302**(1-2): p. 1-7.
147. Meyer, C.G., et al., *Pulmonary tuberculosis: virulence of Mycobacterium africanum and relevance in HIV co-infection.* Tuberculosis (Edinb), 2008. **88**(5): p. 482-9.

148. de Jong, B.C., M. Antonio, and S. Gagneux, *Mycobacterium africanum--review of an important cause of human tuberculosis in West Africa.* PLoS Negl Trop Dis, 2010. **4**(9): p. e744.
149. Demers, A.M., et al., *Mycobacterium africanum is not a major cause of human tuberculosis in Cape Town, South Africa.* Tuberculosis (Edinb), 2010. **90**(2): p. 143-4.
150. Geluk, A., et al., *From genome-based in silico predictions to ex vivo verification of leprosy diagnosis.* Clin Vaccine Immunol, 2009. **16**(3): p. 352-9.
151. Cockle, P.J., et al., *Identification of novel Mycobacterium tuberculosis antigens with potential as diagnostic reagents or subunit vaccine candidates by comparative genomics.* Infect Immun, 2002. **70**(12): p. 6996-7003.
152. Fleischmann, R.D., et al., *Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.* J Bacteriol, 2002. **184**(19): p. 5479-90.
153. Gattiker, A., et al., *The Microbe browser for comparative genomics.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W296-9.
154. Jungblut, P.R., et al., *Proteomics reveals open reading frames in Mycobacterium tuberculosis H37Rv not predicted by genomics.* Infect Immun, 2001. **69**(9): p. 5905-7.
155. Jones, G.J., et al., *Screening of predicted secreted antigens from Mycobacterium bovis reveals the immunodominance of the ESAT-6 protein family.* Infect Immun, 2010. **78**(3): p. 1326-32.
156. Jones, G.J., R.G. Hewinson, and H.M. Vordermeier, *Screening of predicted secreted antigens from Mycobacterium bovis identifies potential novel differential diagnostic reagents.* Clin Vaccine Immunol, 2010. **17**(9): p. 1344-8.
157. Cho, D. and M.T. Collins, *Comparison of the proteosomes and antigenicities of secreted and cellular proteins produced by Mycobacterium paratuberculosis.* Clin Vaccine Immunol, 2006. **13**(10): p. 1155-61.
158. Haslov, K., et al., *Guinea pig cellular immune responses to proteins secreted by Mycobacterium tuberculosis.* Infect Immun, 1995. **63**(3): p. 804-10.
159. Araoz, R., et al., *Towards an immunodiagnostic test for leprosy.* Microbes Infect, 2006. **8**(8): p. 2270-6.
160. Pidot, S.J., et al., *Serological evaluation of Mycobacterium ulcerans antigens identified by comparative genomics.* PLoS Negl Trop Dis, 2010. **4**(11): p. e872.
161. Illumina. 2011; Available from: http://www.ucl.ac.uk/wibr/services/solexa/DeNovoAssembly.pdf.
162. Havlak, P., et al., *The Atlas genome assembly system.* Genome Res, 2004. **14**(4): p. 721-32.
163. Zerbino, D.R. and E. Birney, *Velvet: algorithms for de novo short read assembly using de Bruijn graphs.* Genome Res, 2008. **18**(5): p. 821-9.
164. Pop, M., *Genome assembly reborn: recent computational challenges.* Brief Bioinform, 2009. **10**(4): p. 354-66.
165. Nagarajan, H., et al., *De Novo assembly of the complete genome of an enhanced electricity-producing variant of Geobacter sulfurreducens using only short reads.* PLoS One, 2010. **5**(6): p. e10922.
166. Hewinson, R.G., et al., *Recent advances in our knowledge of Mycobacterium bovis: a feeling for the organism.* Vet Microbiol, 2006. **112**(2-4): p. 127-39.

167. Zhang, F. and J.P. Xie, *Mammalian cell entry gene family of Mycobacterium tuberculosis.* Molecular and cellular biochemistry, 2011. **352**(1-2): p. 1-10.
168. Flesselles, B., et al., *Disruption of the mycobacterial cell entry gene of Mycobacterium bovis BCG results in a mutant that exhibits a reduced invasiveness for epithelial cells.* FEMS Microbiol Lett, 1999. **177**(2): p. 237-42.
169. Gioffre, A., et al., *Mutation in mce operons attenuates Mycobacterium tuberculosis virulence.* Microbes Infect, 2005. **7**(3): p. 325-34.
170. Ahmad, S., et al., *Mammalian cell-entry proteins encoded by the mce3 operon of Mycobacterium tuberculosis are expressed during natural infection in humans.* Scand J Immunol, 2004. **60**(4): p. 382-91.
171. Ohara, N., et al., *Analysis of the genes encoding the antigen 85 complex and MPT51 from Mycobacterium avium.* Infect Immun, 1997. **65**(9): p. 3680-5.
172. Bethunaickan, R., et al., *Antibody response in pulmonary tuberculosis against recombinant 27kDa (MPT51, Rv3803c) protein of Mycobacterium tuberculosis.* Scand J Infect Dis, 2007. **39**(10): p. 867-74.
173. Sander, C., *Investigation of pre-existing immunity on the immunogenicity and efficacy of new TB vaccines*, in *Nuffield Department of Clinical Medicine*. 2007, Oxford University.
174. Smith, S.G., et al., *Identification of major factors influencing ELISpot-based monitoring of cellular responses to antigens from Mycobacterium tuberculosis.* PLoS One, 2009. **4**(11): p. e7972.
175. Suneetha, P.V., et al., *Effect of peptide pools on effector functions of antigen-specific CD8+ T cells.* J Immunol Methods, 2009. **342**(1-2): p. 33-48.
176. Gray, P.M., G.D. Parks, and M.A. Alexander-Miller, *A novel CD8-independent high-avidity cytotoxic T-lymphocyte response directed against an epitope in the phosphoprotein of the paramyxovirus simian virus 5.* J Virol, 2001. **75**(21): p. 10065-72.
177. Kloverpris, H., et al., *Dimethyl sulfoxide (DMSO) exposure to human peripheral blood mononuclear cells (PBMCs) abolish T cell responses only in high concentrations and following coincubation for more than two hours.* J Immunol Methods, 2010. **356**(1-2): p. 70-8.
178. Plebanski, M., et al., *Altered peptide ligands narrow the repertoire of cellular immune responses by interfering with T-cell priming.* Nat Med, 1999. **5**(5): p. 565-71.
179. Mustafa, A.S., et al., *Efficient testing of large pools of Mycobacterium tuberculosis RD1 peptides and identification of major antigens and immunodominant peptides recognized by human Th1 cells.* Clin Vaccine Immunol, 2008. **15**(6): p. 916-24.
180. Conese, M., *Cystic fibrosis and the innate immune system: therapeutic implications.* Endocr Metab Immune Disord Drug Targets, 2011. **11**(1): p. 8-22.
181. Kvarnstrom, M., M.C. Jenmalm, and C. Ekerfelt, *Effect of cryopreservation on expression of Th1 and Th2 cytokines in blood mononuclear cells from patients with different cytokine profiles, analysed with three common assays: an overall decrease of interleukin-4.* Cryobiology, 2004. **49**(2): p. 157-68.
182. Campion, S., et al., *Improved detection of latent Mycobacterium tuberculosis infection in HIV-1 seropositive individuals using cultured cellular assays.* Eur J Immunol, 2011. **41**(1): p. 255-7.

183. Godkin, A.J., H.C. Thomas, and P.J. Openshaw, *Evolution of epitope-specific memory CD4(+) T cells after clearance of hepatitis C virus.* J Immunol, 2002. **169**(4): p. 2210-4.
184. Todryk, S.M., et al., *Correlation of memory T cell responses against TRAP with protection from clinical malaria, and CD4 CD25 high T cells with susceptibility in Kenyans.* PLoS One, 2008. **3**(4): p. e2027.
185. Keating, S.M., et al., *Durable human memory T cells quantifiable by cultured enzyme-linked immunospot assays are induced by heterologous prime boost immunization and correlate with protection against malaria.* J Immunol, 2005. **175**(9): p. 5675-80.
186. Belz, G.T. and F. Masson, *Interleukin-2 tickles T cell memory.* Immunity, 2010. **32**(1): p. 7-9.
187. McKinstry, K.K., T.M. Strutt, and S.L. Swain, *Regulation of CD4+ T-cell contraction during pathogen challenge.* Immunol Rev, 2010. **236**: p. 110-24.
188. Bradley, L.M., L. Haynes, and S.L. Swain, *IL-7: maintaining T-cell memory and achieving homeostasis.* Trends in immunology, 2005. **26**(3): p. 172-6.
189. Last'ovicka, J., et al., *Assessment of lymphocyte proliferation: CFSE kills dividing cells and modulates expression of activation markers.* Cell Immunol, 2009. **256**(1-2): p. 79-85.
190. Parish, C.R., *Fluorescent dyes for lymphocyte migration and proliferation studies.* Immunol Cell Biol, 1999. **77**(6): p. 499-508.
191. Wallace, P.K., et al., *Tracking antigen-driven responses by flow cytometry: monitoring proliferation by dye dilution.* Cytometry A, 2008. **73**(11): p. 1019-34.
192. Miller, J.D., et al., *Human effector and memory CD8+ T cell responses to smallpox and yellow fever vaccines.* Immunity, 2008. **28**(5): p. 710-22.
193. Gerdes, J., et al., *Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67.* J Immunol, 1984. **133**(4): p. 1710-5.
194. Vidyarani, M., et al., *1, 25 Dihydroxyvitamin D3 modulated cytokine response in pulmonary tuberculosis.* Cytokine, 2007. **40**(2): p. 128-34.
195. Breen, E.C., et al., *Multisite comparison of high-sensitivity multiplex cytokine assays.* Clin Vaccine Immunol, 2011. **18**(8): p. 1229-42.

8 Appendices

8.1 Appendix one: trial documentation

Pto

UNIVERSITY OF OXFORD

STUDY PROTOCOL

Study Reference: Lab-T002

Title: Investigation of immune responses following exposure to environmental mycobacteria

Sponsor: University of Oxford
Chief Investigator: Dr Anna Checkley
Supervisor: Dr Helen McShane

Version Number: 2
15th June 2009

Investigator Agreement
"I have read this protocol and agree to abide by all provisions set forth therein.
I agree to comply with the International Conference on Harmonisation Tripartite Guideline on Good Clinical Practice."

Investigator signatures **dates**

Confidentiality Statement
This document contains confidential information that must not be disclosed to anyone other than the Sponsor, the Investigator Team, and members of the Independent Ethics Committee. This information cannot be used for any purpose other than the evaluation or conduct of the clinical investigation without the prior written consent of the sponsor (University of Oxford).

Contents

1. Background information and scientific rationale

1.1. Tuberculosis and BCG

Tuberculosis remains a major health problem worldwide, and BCG works variably in preventing pulmonary TB. A more effective vaccine is urgently needed. BCG consistently works better in temperate than tropical regions, with efficacies of 70-80% in UK trials, compared with 0% in Malawi. A large meta-analysis of clinical trials suggested that geographical latitude explains up to 41% of this difference. Understanding why this occurs is important in designing new TB vaccines, as it may also occur in vaccines currently under development.

1.2. Environmental mycobacteria

A leading hypothesis is that immunity induced by exposure to related environmental mycobacteria, common in the tropics, affects the immune response to BCG vaccination. It has been shown that *in vitro* T cell IFN-γ responses to PPD increase dramatically in UK individuals following BCG vaccination, but are high before vaccination in Malawians with very little increment after BCG vaccination. PPD contains many antigens also present in environmental mycobacteria, and exposure to environmental mycobacteria is thought to account for these results.

1.3. MVA85A

MVA85A is a leading TB vaccine candidate developed by Dr McShane in Oxford. It consists of a viral vector expressing antigen 85A (Ag85A), an immunodominant protein present in all known species of mycobacteria. Phase I and II trials have shown that the vaccine is very immunogenic, but that it induces higher T cell IFN-γ responses in individuals in Oxford compared to those in South Africa. It is important to understand why this difference exists, in order to improve vaccine efficacy. We propose that high levels of environmental mycobacteria exposure in South Africa result in reduced immune responses to this vaccine.

1.4. Current research

One of the research projects currently under way in the group is focussed on looking at the influence of exposure to environmental mycobacteria on the immunogenicity of the vaccine. This work is based on stored samples from trials of the vaccine MVA85A. At the time of the trials, all volunteers provided full informed consent for participation, including the collection of blood samples at various times. Leftover cells from these samples (peripheral blood mononuclear cells, PBMCs) and serum have been frozen and stored for future immunological analysis in relation to tuberculosis. Consent for storage and future analysis of the sample was provided in the individual trials at the time, and a separate ethics application to cover this work has also been accepted (UK NHS ethics committee reference 06/Q1602/146). All of these samples are anonymised, identifiable only by initials and/or trial specific identifiers.

The work described above is already fully covered by ethical approval. This application relates to the following two specific aims. We would like to recruit a small group of individuals, and take a single blood sample from them, to be used as a positive control in the

assays under development. We would also like to use anonymised cord blood samples, provided by the National Blood Service, as a negative control in our assays.

1.5. Positive controls

We propose to invite subjects who are patients at the chest clinic, and who have had positive sputum cultures for environmental mycobacteria, to participate. We propose to identify these patients via the microbiology laboratory of the John Radcliffe Hospital, Oxford. In developing an assay to compare levels of exposure to environmental mycobacteria between vaccine recipients in Africa and Oxford, we need to have a control group who we know are exposed to environmental mycobacteria, and in whom we would expect our assay to be positive.

1.6. Negative controls

Cord blood samples are routinely collected by the National Blood Service for research purposes. They are collected with full informed consent of the mother, and are totally anonymised. For our purposes, they will provide an effective negative control for the assay, since levels of exposure to environmental mycobacteria will be expected to be low.

2. Objectives

To provide positive and negative control samples in assays which are currently being developed to compare immune responses to environmental mycobacteria in individuals receiving the vaccine MVA85A in different trial sites.

3. Study design

3.1. Overview

Individuals currently attending the chest clinic, Churchill Hospital, who have had environmental mycobacteria isolated from sputum samples, will be identified by the microbiology department, John Radcliffe Hospital. They will be invited to participate in the study by means of a letter from the chest clinic. This will be followed up a week later by a phone call from a member of staff at the chest clinic, to ensure that they received the letter, and to check whether they have any questions. Alternatively, they may be approached by a member of staff at the chest clinic when they attend a routine appointment, and informed of the study.

If they are interested in participating, or have further queries, an appointment with a study investigator will be arranged. The investigator will go through the participant information sheet, ensuring that they understand it, and answering any questions they may have.

At that point they will have the option of declining the invitation, taking time to think about it, arranging an appointment at another date, or participating in the study.

If they decide to take part, they will be asked to provide basic demographic information including age, ethnic origin and country of birth, and sufficient information to allow the investigators to determine that they meet the inclusion criteria and exclusion criteria. They will be asked about history of BCG vaccination, travel history, previous TB infection, history of conditions causing immunosuppression, anti-tuberculous medication, other medications, and examined for the presence of a BCG scar. They will then be asked to provide a single 50ml blood sample. This will be the only procedure in the study.

They will be allocated a unique study identifier, so all samples and data stored will be anonymous.

3.2. Study visit

At the visit, the investigator or clinical research nurse will fully inform the subject of all aspects of the study. The aims of the study and all tests to be carried out will be explained. The subject will be given the opportunity to ask about details of the study, and will then have time to consider whether or not to participate. Subjects will have the choice of continuing to the consent and phlebotomy stage on the same day (for convenience) or at a later date, depending on their preference. If they do decide to participate, they will sign and personally date two copies of the consent form, one for them to take away and keep, and one for the investigator. These forms will also be signed and dated by the investigator or clinical research nurse. All subjects will personally sign and date the informed consent form before any study specific procedures are performed.

3.2. Duration of study

3.2.1. Cord blood samples

These will be collected and stored on an ongoing basis, to be used as a negative control in our assays.

3.2.2. Blood samples from individuals with positive sputum culture for environmental mycobacteria

Each participant will be asked to provide one blood sample. These samples will be stored, if specific consent is given, for 15 years. If at any point consent is withdrawn for the storage of these samples, they will be destroyed.

3.3. Definition of end of study

The last collection of blood from the last recruited volunteer in the study.

3.4. Study site

Study administration, management and co-ordination will be conducted at the Jenner Vaccine Institute, ORCRB, Roosevelt Drive, Oxford, OX3 7DQ and the Centre for Clinical Vaccinology and Tropical Medicine (CCVTM), Churchill Hospital, Oxford, OX3 7LJ.

3.4.1. Lone worker policy

When an individual visits a patient at their home, they will routinely inform a second member of the clinical research team (who already has access to this information) of details of the planned visit, including time (and expected time of return) and location of visit. They will inform the same individual when they have safely returned.

3.5. Sample storage

All samples will be stored at the Jenner Vaccine Institute, ORCRB, Roosevelt Drive, Oxford, OX3 7DQ and the CCVTM, Churchill Hospital, Oxford, OX3 7LJ. Both of these buildings are part of the Churchill Hospital site in Oxford. The work detailed in this study will be conducted mainly at these sites. It may also be conducted at other collaborating institutions at the discretion of the Chief Investigator.

4. Selection and withdrawal of study participants

4.1. Study population

This study will enrol adult patients at the Oxford Radcliffe Hospitals Trust chest clinics who have, as a part of clinical investigation, had environmental mycobacteria isolated from sputum samples submitted to the John Radcliffe Hospital microbiology laboratory.

4.2. Recruitment

Participants will be recruited by two means. Firstly, patients currently attending the chest clinic and with environmental mycobacteria isolated from sputum samples will be contacted by a letter containing a patient information leaflet and information on how to get in touch with the study team. This letter will be followed a week later by a phone call from one of the chest clinic nurses, to check that the letter has been received and to check if the individual has any queries. Secondly, an individual may be approached by one of the chest clinic staff at a routine clinic appointment, and given a patient information leaflet. If an individual expresses interest in the study, they will have the options of either a home visit, or meeting the investigator or study nurse at the chest clinic or the CCVTM.

4.3. Inclusion criteria

- Isolation of environmental mycobacteria from sputum
- Competent to consent as judged by the investigator
- Over the age of 18 years

4.4. Exclusion criteria

- Currently taking systemic steroids or other immunosuppressive medication
- Taken steroids or other immunosuppressive medication in the last 3 months
- Other known cause of immunosuppression

4.5. Informed consent

4.5.1. Cord blood samples

Informed consent is obtained prior to cord blood donation. The use of the blood for medical research is specifically mentioned in the consent procedure (consent form enclosed). The cord blood sample is then withdrawn by the National Blood Service, which subsequently ensures that the sample is fully anonymised.

4.5.2. Blood samples from individuals with positive sputum culture for environmental mycobacteria

All subjects will personally sign and date the informed consent form before any study specific procedures are performed. The information sheet will be made available to the volunteer before they are asked to participate in the study. A copy of the informed consent document will be given to the subjects for their records. A member of the study team will be present, to answer any questions that may occur.
The following general principles will be emphasised:
• Participation in the study is entirely voluntary
• Refusal to participate involves no penalty or loss of medical benefits
• The subject may withdraw from the study at any time
• The subject is free to ask questions to allow him or her to understand the purpose of the study and the procedures involved
• There is no direct benefit for participating
• The volunteer's chest physician will be contacted to inform them of their participation

• The chest clinic notes will be consulted to corroborate their history
In any subjects without fluent English language skills, an interpreter will be obtained.

4.6. Subject confidentiality

All records will be kept in a locked filing cabinet, which is accessed only by the investigators and the study nurse(s). All computer entry and networking programs will be done with coded numbers and initials only. Only the investigators and clinical trials staff at the study site will have access to the records. Subject data and information generated will be held in strict confidence. No information concerning the study or the data will be released to any unauthorized third party, without prior written approval of the sponsor.

5. Laboratory techniques

The most important form of immune response in effective immunity against TB infection is thought to be cell-mediated immunity.

5.1. Cell mediated immunity

The standard assay of cell-mediated immunity performed in the laboratory is the interferon gamma (IFN-γ) ELISPOT assay, using peripheral blood mononuclear cells (PBMCs) obtained from volunteers. For the ex vivo IFN-γ ELISPOT assay PBMC are stimulated with environmental mycobacterial peptides, control peptides (such as TB peptides), recombinant antigens (rAg85A) and purified protein derivative. Other exploratory immunological assays aimed at investigating the immune response to the vaccine may be performed at the discretion of the investigator. These may include, for example: ELISPOT assays for other cytokines, and *in vitro* assays of cellular proliferation and intracellular cytokine staining and flow cytometry.

5.2. Antibody-mediated immunity

Antibody responses to environmental mycobacteria may also be measured, at the discretion of the investigators.

5.3. HLA testing

Genetic variations, such as HLA type, influence the response to infection. The HLA type may also be determined.

6. Data handling

The Chief Investigator will be the data manager with responsibility for receiving, entering, cleaning, querying and analysing all data that accrues from the study. The data will be entered into the subjects' CRFs and may be transferred to a database. This includes clinical and laboratory data. Dr Helen McShane will be the custodian of the data, with ongoing responsibility for data storage.

All samples are identifiable only by a study specific subject number and code. Data relating to this work will be stored for a period of 15 years.

7. Financing and insurance

7.1. Insurance

Oxford University investigators participating in this study will receive insurance cover from the University clinical trials insurance policy. The volunteer information sheet includes the following text:
"Compensation for harm arising from an accidental injury and occurring as a consequence of your participation in the study may be covered by the University of Oxford. If you are harmed and this is due to someone's negligence then you may have grounds for legal action for compensation against the University of Oxford (in respect of any harm arising out of the participation in the Clinical Study)"

7.2. Compensation and expenses

Participants will be reimbursed £10 for their time and inconvenience. Transport costs will also be paid, if required.

THE JENNER INSTITUTE
Old Road Campus Research Building
Roosevelt Drive, Oxford OX3 7DQ

REC 09/H0605/75

Information about the research

Measurement of immune responses to environmental mycobacteria

Study protocol number: Lab-T002

We would like to invite you to take part in our research study. Before you decide, we would like you to understand why the research is being done, and what it would involve for you.

Why is this research being done?

We are trying to find out if some common bacteria in the environment reduce the effectiveness of vaccines used against the disease tuberculosis.

Tuberculosis (also known as TB) is already responsible for more deaths worldwide than any other infectious disease, and is becoming more common in this country. At present, BCG is the only vaccine we have to prevent TB, but it works least well in tropical countries- where it is needed the most. We need to understand why BCG is not very effective in these countries, as part of work to develop a new and more effective vaccine against TB.

We think that BCG may not work well in tropical countries because of high levels of 'environmental mycobacteria'. These are bacteria that are distantly related to TB, which are commonly found in the environment.

We are working to improve BCG by boosting its effect with a new TB vaccine called MVA85A. This new vaccine is currently being tested in Oxford and also in Africa. At the same time we need to find out if environmental mycobacteria affect the success of this new vaccine. So we are planning to look at the effects of these bacteria on people who have received this new vaccine, in trials in the UK and Africa.

Before we can do this, we need to try out our test on people who we know have definitely been exposed to environmental mycobacteria. These bacteria are very common in the UK, and many people are exposed to them.

Why have I been asked to take part in this research?

You have had environmental mycobacteria grown from a sputum sample that you gave to your doctor or nurse. We are interested in people who have had this bacteria found in their sputum, as we know these people have definitely been exposed to it. That is why we are asking you to help.

Will it harm me?

It is very common to find this sort of bacterium in sputum samples, particularly in people with other lung problems, and it usually causes no harm. Your doctor is aware of the result, and will have discussed whether it advisable to consider treatment. In the majority of cases, no treatment is required.

Can I take part?

To take part you need to be at least 18 years old, and to have had a positive sputum sample for environmental mycobacteria. You also must *not* be taking medication such as steroids, which may suppress the immune system.

What will this research involve for me, if I decide to take part?

If you decide you would like to take part, you will only need to attend for a single visit, which could take place either at your home or at the Churchill Hospital (the Centre for Clinical Vaccinology and Tropical Medicine or the Chest Clinic). One of the researchers will go through this information sheet with you, and answer any questions. Once you are happy that you understand what the study involves, you will be asked to sign two copies of a consent form.

Following this, the researcher will go through a few questions for administrative purposes and some related to your health. Next, the researcher will draw a 50ml blood sample (about 4 tablespoons). This will be the only procedure performed, and after this you will be free to go.

Are there any risks from taking part in the study?

Drawing blood may cause slight pain and occasionally bruising at the site where the needle enters. Rarely, people feel light-headed or even faint. If you were to have a disorder of your immune system or blood cells that has previously not been diagnosed (these are very rare), it is possible that our tests may bring it to light. If this were the case, we would offer to refer you to the appropriate specialist for further management. The measurements we would perform on the blood sample may also occasionally indicate that you might be infected by TB. If this was the case, we would discuss these results with you and with your clinic doctor.

What are the advantages of taking part?

This study will not be of direct benefit to you. However, information gained from this study may aid in the fight against TB around the world.

Will my taking part in this study be kept confidential?

Yes. All information that is collected about you during the course of the study will be kept strictly confidential. It is available to the study team and authorised staff from the organiser or 'sponsor' of the research (University of Oxford). Any information about you that leaves the hospital/clinic will have your name and address removed so that you cannot be recognised from it. Your clinic doctor will be informed about your participation.

Every effort will be taken to maintain confidentiality. Information about you may be stored electronically on a secure computer server, and paper notes will be kept in a locked filing cabinet at the study centre.

What tests will be done on my blood?

A series of immunology tests will be done to the blood sample you have supplied, looking at your body's response to the environmental mycobacteria found in your sputum. Tests to assess your body's response to tuberculosis will also be performed. Some blood will be used to look at the pattern of your genes that can affect the immune system (the HLA [Human Leukocyte Antigen] genes). If you would like to know precise details of what tests are being done, a list can be provided.

With your consent some of your leftover blood samples will be stored and may be used for further immunology studies of the body's immune response to environmental mycobacteria or TB. No studies concerning diseases other than TB will be done. Samples will be stored for up to fifteen years unless you would prefer us not to. Upon your request at any time, your remaining blood samples will be destroyed. Your participation in this study will not be affected by your decision to allow or not allow storage and future use of your leftover blood samples.

What if something goes wrong?

The investigators will make every effort to ensure your safety and well-being. If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions. If you remain unhappy and wish to complain formally, we will provide you with details of the University's complaints procedure. In addition, the Patient Advice and Liaison Services (PALS) is available at the Churchill Hospital, Monday to Thursday 9.00am - 5.00pm, tel: 01865 235855, email: PALSCH@orh.nhs.uk.

Compensation for harm arising from an accidental injury and occurring as a consequence of your participation in the study may be covered by the University of Oxford. If you are harmed and this is due to someone's negligence then you may have grounds for legal action for compensation against the University of Oxford (in respect of any harm arising out of the participation in the Clinical Study).

What if I change my mind?

If you decide you would rather not have a blood sample taken, you will be entirely free to change your mind about taking part. If you change your mind after we have taken the blood sample, we will ensure that it is destroyed, and not used for further tests. You are free to withdraw at any time, without giving a reason. This will not affect your subsequent medical care in any way.

Will I be paid for taking part?

You will be compensated £10 for your time, in addition to travel expenses for any additional visits as part of the study.

What will happen to the results of the research study?

At the end of the study we aim to publish the results in scientific journals and may talk about them at scientific or medical conferences but you will not be identified in any report or publication. If you would like a summary of the study findings, please ask.

Who is organising and funding the research?

This study is funded by the Medical Research Council, a government funding body. The study is designed and organised by the investigators who are staff of the University of Oxford. Neither your doctor nor the researchers are paid for recruiting you into this study.

Who has reviewed the study?

Ethical review has been carried out by Oxford Research Ethics Committee B (ref 09/H0605/75).

I hope this information sheet has helped answer any questions you might have. If you would like to take part in this study, or find out more about it, please contact:

Dr. Anna Checkley
Clinical Research Fellow

Address Jenner Vaccine Institute, ORCRB, Roosevelt Drive, Oxford OX3 7DQ
Tel (01865) 617615
Email anna.checkley@ndm.ox.ac.uk

αβχ

OxREC Ref: 09/H0605/75

Oxford Centre for Respiratory Medicine
Churchill Hospital
Old Road
Headington
Oxford
OX3 7LJ
Tel: 01865 225252
Fax: 01865 225221

Date:

To

Re: A study you may be interested in and eligible for

I am writing to you because we are collaborating with a research group in Oxford, who are working to design new improved vaccines against tuberculosis (TB). In order to help do this, they are carrying out some studies looking at the effects of bacteria which are distantly related to TB on the body.

In some individuals, bacteria such as these are found in a sputum sample. This usually causes no harm to the health of the person concerned, but blood samples from these people can be a useful way of studying the effects of these bacteria on the immune system.

I am writing because you are possibly eligible for this study. An information leaflet from the study team is enclosed with this letter. This gives much more detail about what the study involves. Once you have read this, if you are interested, please do not hesitate to contact the study investigator with any questions or to arrange a visit.

Yours sincerely

Dr Henry Bettinson
Locum Respiratory and Intensive Care Physician

THE JENNER INSTITUTE
Old Road Campus Research Building
Roosevelt Drive, Oxford OX3 7DQ

Consent form

OxREC No. 09/H0605/75

Evaluating immune responses to environmental mycobacteria.

Principle Investigator: Anna Checkley, MBChB MRCP

Please initial box

1. I confirm that I have read and understand the Participant Information Leaflet (version 2, dated 15/06/09) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I have received enough information about the study. ☐
3. I understand that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. ☐
4. I understand that relevant sections of my medical notes and data collected during the study may be looked at by study individuals from the University of Oxford, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. ☐
5. I agree for the research team to inform my physician in the Chest Clinic that I am taking part in this study. ☐
6. I agree that some of my leftover blood samples will be stored and may be used for further studies of the body's immune response to environmental mycobacteria. I agree to gift blood samples taken for the purpose of the research study to the University of Oxford.Samples will be stored for up to fifteen years and I have the right to ask for these samples to be disposed of at any time. ☐
7. I agree to take part in this study. ☐

Signed: Date:

Name: (in block letters)

Signature of Investigator: Date:

Name of Investigator: (in block letters)

Thank you for taking the time to read this leaflet, if you would like more information or to donate your placenta or cord please contact :

The Cord Blood Collection Team
National Blood Service
John Radcliffe Hospital
Oxford OX3 3BQ

Telephone: 01865 447944

Professor C. W. G. Redman
Nuffield department of Obstetrics
and Gynaecology
John Radcliffe Hospital
Oxford OX3 9DU

Telephone: 01865 221009

Oxfordshire Research Ethics Committee Study nos. C02.313 and C02.090

NHS Oxford Radcliffe Hospitals NHS Trust

NHS Blood and Transplant

Your placenta or cord could be used for research into:

Prevention of pregnancy complications such as premature labour or high blood pressure

New treatments for diseases such as cancer and heart disease

Understanding the way blood cells develop

Your opportunity to take part in scientific research . . .

Taking part in research using the afterbirth (placenta) or the cord or cord blood

Introduction

There are several research teams associated with the John Radcliffe Hospital that are investigating serious conditions — some that can occur during pregnancy and others that may affect all women and men. The link between these teams is that the research involves the afterbirth (placenta) or the baby's cord, which is left behind, or the blood that remains in the cord. All of these are usually discarded after delivery.

What is the purpose of this research?

The research being undertaken includes work using the placenta or cord on problems such as premature labour and high-blood pressure in pregnancy. The research on cord blood studies the way different types of cells are made in the body and this work could lead to better prevention or treatment of diseases such as cancer and heart disease.

What would I have to do?

You can assist this work by allowing us to use your placenta, cord blood cells or the cord itself for our research after you have had your baby. This would in no way interfere with the birth itself and all the studies involved have been given Oxfordshire Research Ethics Committee approval.

Will my donation be used for profit?

Most of the research is funded by the government or charities but some grants originate with commercial companies. However no tissue would be sold or used for commercial purposes. Likewise there would be no benefit financially to you or your child if you decide to help us with this research and would have no future rights to the samples.

Would my taking part in this study be kept confidential?

Some information relating to you and the donation may be stored on a database but the samples for research will not be identifiable to the researchers by your name or address. For a few of the research projects it is necessary for your cord blood to be tested for Hepatitis B and C, HTLV and HIV, If you do not want this testing done your donation can still be used for the other research projects and this can be indicated on the consent form.

Will my donation be stored?

If research cannot be carried out immediately then samples may be stored to avoid wasting this valuable resource. However any future research would only be done with Ethics Committee approval. Samples would be disposed of according to ethical guidelines issued by the government and would not be used for any treatment purposes at any time.

What is the next step?

You may be approached at the hospital to ask for your consent to donate. There is no obligation to agree, and if you do not want to donate your treatment will not be affected in any way. If you would like to donate, the cord blood collection team member will ask you to sign a consent form. On the day of your delivery your placenta and cord would be collected after the birth of your child for an ethically approved research project. We aim to collect as many donations as possible, but depending on when your baby is born we may not be able to collect your donation.

. . . would you like to help?

αβχ

TAKING PART IN RESEARCH USING THE AFTERBIRTH (PLACENTA) OR THE CORD OR THE CORD BLOOD

Oxfordshire Research Ethics Committee Study nos. C02.313, C03.090 & 05/Q1606/117

Introduction.
There are several research teams associated with the John Radcliffe Hospital that are investigating serious conditions - some that can occur during pregnancy and others that may affect all women and men. The link between these teams is that the research involves the afterbirth (placenta) or the baby's cord, which is left behind, or the blood that remains in the cord. All of these are usually discarded after delivery.

What is the purpose of this research?
The research being done includes work using the placenta or cord on problems such as premature labour and high blood pressure in pregnancy. The research on cord blood studies the way different types of cells are made in the body and this work can lead to better prevention or treatment of diseases such as cancer and heart disease.

What would I have to do?
You can assist this work by allowing us to use your placenta, cord blood cells or the cord itself for our research after you have had your baby. This would in no way interfere with the birth itself and all the studies involved have been given Oxfordshire Research Ethics Committee approval.

Will my donation be used for profit?
Most of the research is funded by the government or charities but some grants originate with commercial companies. However no tissue would be sold or used for commercial purposes. Likewise there would be no benefit financially to you or your child if you decide to help us with this research and you would have no future rights to the samples.

Would my taking part in this study be kept confidential?
Some information relating to you and the donation may be stored on a database but the samples for research will not be identifiable to the researchers by your name or address. For a few of the research projects it is necessary for your cord blood to be tested for Hepatitis B and C, HTLV and HIV. If you do not want this testing done your donation can still be used for the other research projects and so this form has a separate section for this consent.

Will my donation be stored?
If research cannot be carried out immediately then samples may be stored to avoid wasting this valuable resource. However any future research would only be done with Ethics Committee approval. Samples would be disposed of according to ethical guidelines issued by the government and would not be used for any treatment purposes at any time.

Your treatment will not be affected in any way if you prefer not to give permission.
You can obtain more information by contacting one of the co-ordinators of the research listed below the consent form.

αβχ

TAKING PART IN RESEARCH WORK LOOKING FOR NEW TREATMENTS.

Consent form

I have read the attached information dated 06/06/06 and have been given a copy to keep. I have been able to ask questions and know why the research is being done.

Please Initial Box ☐

I agree that my placenta and the cord blood and the cut end of the cord itself can be used for research work that has been approved by the Ethics Committee and understand that I or my baby have no further rights to the samples. I understand there will be no benefit financially to me or my child from this research.

Please Initial Box ☐

I agree that my donation may be stored for use in future projects that have been given Ethics Committee approval and I understand that any information stored on a database will be kept confidential.

Please Initial Box ☐

I agree that my cord blood may be tested for Hepatitis B and C, HTLV and HIV.

(In the unlikely event of a positive test a trained health advisor and your GP would offer you appropriate advice.)

Please Initial Box ☐

Signed ..... Date .....

Name in Capitals .....

Consent obtained by ..... Date .....

Name in Capitals .....

List of research personnel who may be contacted for further information

For work using cord blood cells: Cord Blood Collection Team 01865 447944
National Blood Service
Level 2, John Radcliffe Hospital, Oxford OX3 9BQ

For work using placenta and cord tissue: Professor C.W.G. Redman 01865 221009
Nuffield Department of Obstetrics and Gynaecology
John Radcliffe Hospital, Oxford OX3 9DU

8.2 Appendix two: Annotated genome sequences

1: *Mycobacterium avium* 104 [TIGR]

2: *Mycobacterium avium subsp. paratuberculosis* K-10 [University of Minnesota]

3: *Mycobacterium abscessus* [Genoscope]

4: *Mycobacterium bovis* AF2122/97 [Sanger Institute]

5: *Mycobacterium bovis* BCG [Fiocruz - FAP]

6*: Mycobacterium bovis* BCG str. Pasteur 1173P2 [Mycobacterium bovis sequencing teams]

7: *Mycobacterium gilvum* PYR-GCK [DOE Joint Genome Institute]

8*: Mycobacterium leprae* Br4923 [Institut Pasteur/Institut Pasteur PF1]

9: *Mycobacterium leprae* TN [Sanger Institute]

10: *Mycobacterium marinum* M [Welcome Trust Sanger Institute]

11: *Mycobacterium smegmatis* str. MC2 155 [TIGR]

12: *Mycobacterium sp. JLS* [DOE Joint Genome Institute]

13*: Mycobacterium sp. KMS* [DOE Joint Genome Institute]

14: *Mycobacterium sp. MCS* [DOE Joint Genome Institute]

15: *Mycobacterium tuberculosis* CDC1551 [TIGR]

16: *Mycobacterium tuberculosis* F11 [Broad Institute]

17: *Mycobacterium tuberculosis* H37Ra [Chinese National HGC, Shanghai/Fudan University, P.R. China, Shanghai/Johns Hopkins University, Department of Molecular Microbiology & Immunology, Bloomberg School of Public Health, USA, Baltimore]

18: *Mycobacterium tuberculosis* H37Rv [Sanger Institute]

19: *Mycobacterium ulcerans* Agy99 [Institut Pasteur]

20: *Mycobacterium vanbaalenii* PYR-1 [DOE Joint Genome Institute]

8.3 Appendix three: Un-annotated genome sequences

1: *Mycobacterium intracellulare* ATCC 13950 [McGill University, Canada]

2: *Mycobacterium kansasii* ATCC 12478 [McGill University, Canada]

3. *Mycobacterium africanum* (downloaded from the Sanger centre FTP site)