Studies on *in vitro* human T cell reactivity to antigens of *Mycobacterium tuberculosis*

Hannah Priyadarshini Gideon

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Medical Microbiology

Faculty of Health Sciences

UNIVERSITY OF CAPE TOWN

June 2010

For

Appa and Amma

with Love

DECLARATION

I, Hannah Priyadarshini Gideon, hereby declare that the work on which this dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: …….

Acknowledgements

This work has been performed under the supervision of Professor Robert J Wilkinson and Dr Katalin A Wilkinson, at the Institute of Infectious Disease and Molecular Medicine, Faculty of health science, University of Cape Town; and supported by a joined grant from Bill and Melinda gates foundation through Grand challenges for global health and Welcome trust.

I am very grateful to Professor Robert J Wilkinson and Dr Katalin A Wilkinson, my supervisors for giving me this wonderful opportunity to do my PhD, for providing the funds for the research, teaching me techniques, guidance, discussion, encouragement, support, for beliving in me and helping me at all times. Thank You.

I thank,

Dr David Sherman and Dr Tige R Rustad, Seattle Biomedical Research Institute, Washington, USA, for providing the transcriptional profiles of MTB, which formed the basis of this work.

Professor Douglas B Young, MRC National Institute for Medical Research, and Imperial College London, UK; for his valuable comments and encouragement.

Dr Marcel Behr, McGill University Health Centre, Montreal, for providing the RD2 knock out mutant MTB strains.

Dr Martin Vordermeier, Veterinary Laboratories Agency, Weybridge, UK, for providing synthetic peptides.

Dr Szilvia Bosze and Dr Kata Horvati, Estovs University, Budapest, Hungary, for the synthesis of Rv2654c and Rv1733c overlapping peptides, and for the collaboration.

Dr Shea Hamilton, Imperial College, London, for the SELDI analysis.

Dr Tolu Oni, Ubuntu Clinic, Khayelitsha, South Africa, for recruiting HIV uninfected active and latent tuberculosis persons; Dr Molebogeng Rangaka and Dr Graeme Meintjes' for recruiting HIV-1 infected persons, and Dr Dominique Pepper for recruiting healthy volunteers, and also the nurses and staff at the clinic for collecting the samples.

Ronnette Seldon, Kathy Wood and Heinner Guio for helping me in the lab, to process the samples, and everyone at the laboratory and department for the wonderful time.

I am grateful to Lutheran World Federation, University of Cape Town, and Professor Robert Wilkinson for the financial assistance for my PhD.

My Parents have nurtured me with love and support all the way through, been my strength, and encouraged me at all times. I also thank my family and friends for standing by me throughout.

Last but not the least, I thank God, for his His grace and mercy, for making all this possible.

Abstract

Hannah Priyadarshini Gideon **June 2010**

Studies on in vitro human T cell reactivity to antigens of Mycobacterium tuberculosis

Studies on *Mycobacterium tuberculosis* (MTB) antigens are important to improve immunodiagnostics and vaccine efficacy. A novel genome based strategy for antigen discovery is to relate what is highly expressed by bacilli *in vivo* or *in vitro*, to what is recognized by human T cells as antigens.

As hypoxia is a relevant stimulus that MTB encounters *in vivo*, whole genome based transcriptional profiles of *M. tuberculosis* subject to prolonged hypoxia (described as the Enduring hypoxic response [EHR]) were analyzed, to guide the discovery of novel potential antigens, by a combined bioinformatic and empirical approach and to determine evidence of infection stage specific recognition.

29 candidate genes were selected by bioinformatic prediction and evaluated empirically using IFN-γ and IL-2 ELISpot, and Acr-1, CFP-10, ESAT-6 as reference antigens. 26 of 29 proteins induced an IFN-γ response in PBMC of persons with active or latent tuberculosis. IL-2 responses were of lower magnitude than that of IFN-γ and the antigenic repertoire far more focused.

Ten proteins induced a dominant IFN-γ response in ≥ 50% of the donors. Five novel immunodominant proteins: Rv1957, Rv1954c, Rv1955, Rv2022c and Rv1471, showed responses similar to CFP-10 and ESAT-6 in both magnitude and frequency. With the exception of Rv1471, these novel genes are shared only within MTB complex. With the exception of modest IFN-γ response to the peptide pools of Rv0849-2, Rv2659c-2, Rv2693c-1 and very low levels for Rv2021c, there was no strong evidence of infection stage specific recognition.

RD2 encoded, Rv1986 was found to be a major target of central memory CD4+ T cells, inducing IL-2 response, several times greater than ESAT-6 or CFP-10, and that this response is directed narrowly to two epitope regions. This immunodominant target is absent in most commonly used BCG vaccine strains.

In addition, T cell determinants of species specific Rv2654c and dosR regulated Rv1733c were mapped. T cell epitope mapping of Rv2654c revealed, that the dominant epitope recognized in this population (p51-65) is different from what is previously described.

These findings revealed that a number of genes induced by hypoxia are in fact potent T-cell targets and therefore offers general support to the important role of hypoxia in the natural course of tuberculosis infection. Only moderate evidence of infection stage specific recognition of antigens was observed. The approach described for antigen discovery could be potentially employed when predicting antigens from pathogens other than *M. tuberculosis*.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

µ	Micro
°C	Degree Celsius
ΔRD2	RD2 mutant
ΔRD2::Rv1986	RD2 mutant complemented with Rv1986
7H9	Middlebrook 7H9 liquid media
αα	Amino acid
Ab	Antibody
Acr	Alpha crystallin (Acr1, Rv2031c)
af	Allelic frequency
Ag	Antigen
APC	Antigen presenting cell
ATB	Active tuberculosis
BCG	Bacille Calmette-Guérin
BLAST	Basic local alignment search tool
BLASTN	BLAST-nucleotide sequence with nucleotide database
BLASTP	BLAST-protein sequence with protein database
BSA	Bovine serum albumin
cART	Combined antiretroviral therapy
CD27	CD27 receptor
CD3	Cluster of differentiation 3, T cell co-receptor
CD4	Cluster of differentiation 4, T cell co-receptor, T helper cells

CD45RA	CD45 receptor
CD8	Cluster of differentiation 4, T cell co-receptor, cytotoxic T cells
CFP-10	Culture filtrate protein-10
CO_2	Carbon dioxide
DTH	Delayed type hypersensitivity
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dosR	Dormancy survival regulon
DR	Major HLA (MHC) class II cell surface receptor
EAST-6	6kDa early secretory antigenic target
EHR	Enduring hypoxic response
ELISA	Enzyme linked sorbent assay
ELISPOT	Enzyme linked immunosorbent spot
F	Female
FACS	Fluorescence-activated cell sorting
FCS	Fetal calf serum
FI	Fold induction
GO	Gene ontology
g	Centrifugal force
H37Rv	Standard laboratory strain of *Mycobacterium tuberculosis*
HI FCS	Heat inactivated Fetal calf serum
HIV	Human immunodeficiency virus

HLA	Human leukocyte antigen
ICS	Intracellular staining
IEDB	Immune epitope database
IFN-γ	Interferon gamma
IGRA	Interferon gamma release assay
IL	Interleukin
INH	Isoniazid; isonicotinylhydrazine
IQR	Inter quartile range
kDa	kiloDalton
LTBI	Latent tuberculosis
M	Male
mAb	Monoclonal antibody
MDR	Multi drug resistant
mg	milligram
MHC	Major histocompatibility complex
MHCII	MHC class II
MIP1a	Macrophage inflammatory protein 1 alpha
MIP1b	Macrophage inflammatory protein1 beta
mls	milli liter
mm	milli meter
MPT64	Immunogenic protein MPT64
MTB	*Mycobacterium tuberculosis*

NA	Not applicable
ng	nanogram
NI	Normalized intensity
O_2	Oxygen
OD	Optical density
ORF	Open reading frame
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PE/PPE	Group of gene family with the presence of N-terminal ProGlu/ProProGlu motifs
pg	picogram
PHA	phytoheamagglutinin
phiRv	Phage insert
PPD	Purified protein derivative
QFT GIT	QuantFERON TB Gold In Tube
Rantes	chemokine
RD	Region of difference
REC	Research ethics committee
RNA	Ribonucleic acid
rpm	Rotations per minute
SEB	Staphylococcal enterotoxin B

SELDI	Surface-enhancer laser desorption/ionisation
SFC	Spot forming cells
SOP	Standard operating procedure
SSC	Side scatter
T-cell	T lymphocytes
TA	Toxin-antitoxin module
TB	Tuberculosis
TCM	Central memory T cell
TCR	T cell receptor
Tdiff	Terminally differentiated T cell
TEM	Effector memory t cell
Th	Helper T cells
Th1	Type 1 helper T cell
Th2	Type 2 helper T cell
TMHMM	Transmembrane prediction using hidden markov Models
TNF	Tumor necrosis factor
TSPOT	ELISpot based commercial IGRA
TST	Tuberculin skin test
XDR	Extensively drug resistance

"Studies on *in vitro* human T cell reactivity to antigens of *Mycobacterium tuberculosis"*

- **Chapter 1** is an Introduction.
- **Chapter 2** is the materials and methods.
- **Chapter 3** describes the comparison of various techniques and evaluation of performance of synthetic peptides and recombinant proteins with the discovery that there is poor correlation between techniques and that the peptides are degraded, while the recombinant proteins are cleaved in long-term whole blood cultures.
- **Chapter 4** describes antigenic mapping infection stage specific antigens: species specific RD-11 encoded Rv2654c (TB7.7) and dosR encoded Rv1733, with the discovery that the commercially used p38-55 peptide of Rv2654c is not immuno-dominant in this population.
- **Chapter 5** describes informatics and empirical analysis of novel hypoxia inducible targets of human T cell response in subjects with active and latent TB with the discovery of immuno-dominant antigens and that these antigens do not induce a differential recognition between the 2 clinical groups.
- **Chapter 6** describes the analysis of hypoxia-inducible species-specific determinants of the human T cell response with the discovery that RD-2 encoded Rv1986, which is deleted from the commonly used BCG strains, constitutes a major target of memory T cells.
- **Chapter 7** describes the overall conclusions of this thesis.

Chapter 1: Introduction

The chapters of this thesis describe the *in vitro* reactivity of human T cell response to novel T cell antigenic targets of *Mycobacterium tuberculosis* (MTB) using bioinformatic prediction algorithm and empirical testing. The antigens tested are either hypoxia induced: under *in vitro* culture conditions (relating to the physiology of the bacilli *in vivo*) or are encoded by various regions of difference. Therefore, this introductory chapter reviews a brief history of tuberculosis, its pathology, T cell antigen discovery in *Mycobacterium tuberculosis* and the techniques used in antigen discovery.

Historical background

Tuberculosis caused by *Mycobacterium tuberculosis* has claimed its victims throughout much of known human history and it is considered the greatest present bacterial cause of death. Evidence of tuberculosis, in the pre-historic era, dates back to 8000 BC. Johann Lukas Schönlein of Würzburg, listed scrofula, tubercles and phthisis as separate diseases under the heading of 'haematoses' and in 1834, he coined the term 'tuberculosis' to describe affliction with tubercles (Herzog, 1998).

However, understanding the pathogenesis of tuberculosis began only in the 19th century with the identification of tubercle bacilli as the etiological agent by Robert Koch (in 1882). Within few years after the discovery of the bacilli, Koch also produced a glycerin extract of dead tubercle bacilli, which he initially named 'lymph' and later 'tuberculin' and promoted it as an effective remedy. Clemens von Pirquet, developed the tuberculin skin test (TST) in 1907 using Koch's tuberculin. He introduced the term "latent tuberculosis"- describing children who did not manifest any symptoms of tuberculosis but had a positive skin response to tuberculin. In 1908, Charles Mantoux introduced the use of cannulated needle and syringe to inject tuberculin intracutaneously in the current format and Florence

Seibert further developed purified protein derivative (PPD), in the form, which is currently in use (Daniel, 2006; Herzog, 1998).

A major breakthrough in the history of tuberculosis was the successful attenuation of *M. bovis* for use as a vaccine, by Albert Calmette and his associate Camille Guérin, from where the name Bacille Calmette-Guérin (BCG) derives and was successfully developed in 1921. Over the years this vaccine was accepted and remains in use all over the world today (Daniel, 2006).

Another equally important contribution to the diagnosis of TB is the revolutionary advance in the field of physics in the year 1895: the discovery of X-rays by Wilhelm Conrad von Röntgen. With the aid of the rapidly evolving radiological techniques, the development, course and severity of tuberculosis could now be accurately monitored and studied.

Rapidly progressing discoveries in tuberculosis saw another milestone with the development of antibiotic Streptomycin in the 1940s for the treatment of tuberculosis. Following this several other agents were successfully developed: p-aminosalicylic acid (1949), isoniazid (1952), pyrazinamide (1954), cycloserine (1955), ethambutol (1962) and rifampin (rifampicin; 1963).

Current Global burden of tuberculosis

Even after such rapid progress in discoveries ranging from diagnostics to treatment and prevention in the field of tuberculosis, the latest global burden of tuberculosis estimates that almost one third of the world's population is infected with tuberculosis, with 8.9-9.9 million incident cases, 9.6-13.0 million prevalent cases, 1.1-1.7 million deaths amongst HIV-1 uninfected and an additional 0.45-0.62 million deaths in HIV-1 infected persons (WHO, 2010). The estimated incidence worldwide is depicted in figure 1.

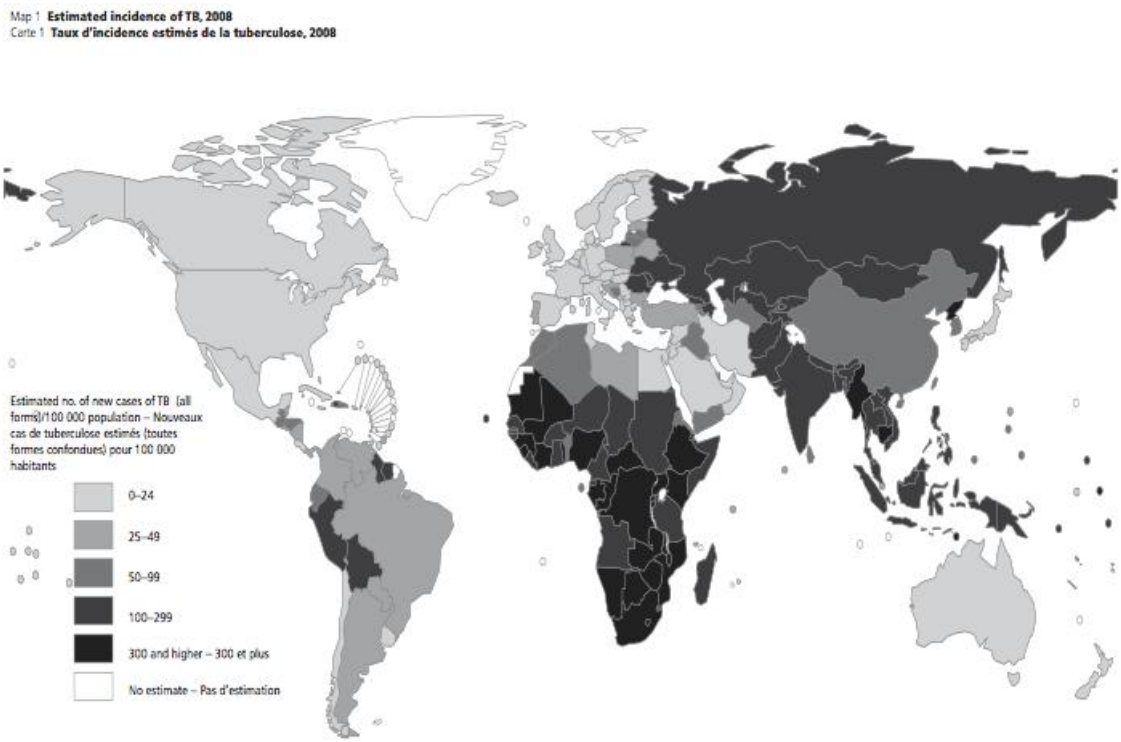


Figure 1 Estimated incidence of Global tuberculosis (As of 2008) (WHO, 2010)

Control of Tuberculosis

Control strategies against tuberculosis still remain largely dependent on the partially effective BCG vaccine and the early diagnosis and treatment of active tuberculosis. Isoniazid preventive therapy and antiretroviral therapy for HIV-1 infected people are further measures. The enormous reservoir of latent tuberculosis also adds potential threat to the control strategy, increasing the lifetime chances of developing active form of tuberculosis, especially in the HIV-1 infected persons. In addition, administration of a live vaccine to HIV-1 co-infected children carries the potential risk of developing disseminated BCG disease (Hesseling et al., 2007; Hesseling et al., 2003). In addition, the HIV-1 pandemic has made the diagnosis and treatment of tuberculosis difficult, due to atypical clinical presentation and dissemination (Maartens and Wilkinson, 2007).

Diagnosis of Tuberculosis

The diagnosis of tuberculosis plays a key role in the control of tuberculosis. In general, diagnosis largely relies on clues from the clinical presentation and symptoms of the illness such as cough, night sweats, weight loss, fever and radiological changes.

Laboratory diagnosis is based on either (i) detecting the bacilli or (ii) detecting an immune response to the bacilli.

(i) Detecting the bacilli

Presence of MTB bacilli in the stained (Ziehl Neelsen) sputum smear, viewed under light microscopy is the most commonly used technique. The sputum smear is highly specific for the diagnosis of pulmonary tuberculosis, cost effective and does not need specialized laboratory, but it is labour intensive and insensitive. The use of fluorescent dye and fluorescent microscopy increases the sensitivity of the sputum smear detection (Mitchison, 1972; Steingart et al., 2006). However, these facilities are not always available in resource-limited settings. Sputum microscopy however, is not sensitive for detection of MTB in extrapulmonary and paucibacillary tuberculosis. In HIV-1 infected persons and in children, the sputum smear is often negative.

Culture of sputum, is the current gold standard in the laboratory diagnosis of tuberculosis. The major draw back however is the time delay, as it can take up to 3-4 weeks to grow in solid media, and 1-10 days in the liquid media (Lagrange and Herrmann, 2008; Lalvani, 2007; Lalvani and Pareek, 2009a; Lange and Mori, 2010; Lange et al., 2009).

While microscopy and culture are still the major backbone for laboratory diagnosis of tuberculosis (TB), new methods including molecular diagnostic tests have evolved over the last two decades (Palomino, 2005). The majority of molecular tests have been focused on: (i) detection of nucleic acids both DNA and RNA, which are specific to *Mycobacterium*

tuberculosis, by amplification techniques such as polymerase chain reaction (PCR); and (ii) detection of mutations in the genes which are associated with resistance to anti-tuberculosis drugs by sequencing or nucleic acid hybridization. These are extensively reviewed (Balasingham et al., 2009; Lange and Mori, 2010).

(ii) Detection of immune response

Humoral immune response

For several decades, B cell targets of MTB were considered with the primary purpose of developing serodiagnostic assays. Evidence in the literature suggests that circulating antibodies to mycobacterial antigens have been described as early as 1965 (Daniel et al., 1979) which employed the use of sterile culture filtrates of mycobacterial culture, which lacked specificity (Daniel and Debanne, 1987; Grange, 1984). Progress, in terms of specificity and sensitivity, were achieved with the aid of monoclonal antibodies directed against immunodominant species-specific antigen determinants (Hewitt et al., 1982). Further use of specific whole antigen molecules like 38-kDa 19- kDa and 14-kDa molecules increased the sensitivity of diagnosis in both smear positive and smear negative tuberculosis (Bothamley et al., 1992; Jackett et al., 1988).

Recently, Steingart and colleagues conducted a systematic review and meta-analysis on the performance of serodiagnosis and found 254 studies evaluating 51 distinct single antigens and 30 multiple- antigen combinations for diagnosing pulmonary tuberculosis. The authors conclude that (i) none of the antigens achieve sufficient sensitivity to replace sputum microscopy, (ii) no single antigen provides a sensitivity that is sufficient for a single antigen to be used to device a serodiagnostic test for TB and (iii) that it is unlikely that a single antigen based serodiagnostic test could be devised; however, (iv) combination of selected antigens provide higher sensitivity than a single antigen (Steingart et al., 2009).

Cellular immune response

The TST (Tuberculin skin test: intradermal injection of a standard dose (2 Tuberculin units) of PPD) detects a delayed type hypersensitivity (DTH) reaction against MTB infection. It is based on the cellular reaction to PPD after 48-72 hours after intradermal injection and defined by a local skin induration and is still considered and used as an effective tool for the diagnosis of tuberculosis, for over a century. The major drawbacks include, operational disadvantages requiring two patient visits, and variability in measuring induration (Huebner et al., 1993). In addition, the response is not MTB specific, as it contains a crude mixture of proteins, which are shared widely by other mycobacteria, including BCG, and therefore cannot readily distinguish between BCG vaccination, MTB infection or exposure to other environmental mycobacteria. The sensitivity however, is highly dependent on the host immune status, and dramatically decreases with acquired immunodeficiency like HIV-1 infection, in children, treatment of corticosteroids (Menzies et al., 2007). In addition, on an average 10-20% of persons with active tuberculosis do not react, and the reaction is variable (Huebner et al., 1993).

Advances in the genomic research of MTB lead to the landmark discovery of the genomic region of difference (RD1) which is absent in non virulent *M. bovis* vaccine strains (BCG) (discussed in detail later), led to the development of *in vitro* T cell based immunodiagnosis measuring Interferon- gamma (IFN-γ) release in response to the RD1-encoded antigens ESAT-6 and CFP-10. The assays show promising results for diagnosis of latent tuberculosis with operational advantage requiring only one patient visit, higher sensitivity compared to the TST, and unaffected by prior BCG vaccination (Pai et al., 2009) and considered the major advance in the last decade (Lalvani and Pareek, 2009a; Lange and Mori, 2010; Pai et al., 2008). These assays are also less affected by the immune status of the person tested (Rangaka et al., 2007b). However, these assays do not differentiate between MTB infection and active disease (Lalvani and Pareek, 2009b).

Treatment of tuberculosis

Antibiotic regimes used as anti-TB treatment, have been shown to be associated with cure rates of up to 95%, and require at least six months treatment. The necessity for long-term treatment has contributed to poor treatment adherence, which may be confounded by other factors such as re-treatment, poor medical supervision, irregular supply of drugs and inappropriate drug dosage, which eventually lead to the development of drug resistance. Alternative approaches like DOTS were introduced to handle the treatment services (Weis et al., 1994).

It is estimated that 0.5 million cases of multidrug resistant TB (MDR-TB) (defined as TB that are resistant to at least isoniazid and rifampicin) occurs each year. In addition, 58 countries have reported at least one case of extensively drug resistant tuberculosis XDR-TB (defined as MDR-TB plus resistance to a fluoroquinolone and at least one second-line injectable agent: amikacin, kanamycin and/or capreomycin) as of March 2010 and 5.4% of MDR-TB cases are XDR-TB (WHO, 2010).

Mycobacterium tuberculosis infection

Infection with MTB follows a pattern of events that are established through observation in humans and in animal models. Tuberculosis is spread by airborne droplet nuclei, which contain *Mycobacterium tuberculosis*. It remains viable for minutes to hours, after expectoration by people with pulmonary or laryngeal tuberculosis during coughing, sneezing, or talking. The infectious droplet nuclei are inhaled and lodge in the alveoli. In the alveolar sacs, *M. tuberculosis* is engulfed by alveolar macrophages, key cells of the innate immune system that act as a front-line barrier against infection. The bacilli can resist the bactericidal mechanisms induced by the macrophages (i.e., phagolysosome formation), and multiply inside the phagosome (thus avoiding the macrophage apoptosis), in a protected intracellular niche. These macrophages invade the sublying epithelial layer and

leads to a local inflammatory response that results in recruitment of mononuclear cells from the neighboring blood vessels, providing fresh host cells for the expanding bacterial population. These cells constitute the granuloma, the hallmark of tuberculosis disease pathology. With the development of acquired immune response and the arrival of lymphocytes, the granuloma acquires a more organized, stratified structure (Russell et al., 2010).

The development of immune response (DTH) about 4-6 weeks after the primary infection is indicated by a positive TST reaction. The balance between the host immunity and the bacillary multiplication determines the outcome of the tuberculosis.

An encounter with MTB is classically regarded to give rise to three outcomes: (i) a minority of the population (5%) rapidly develop primary active TB disease – with clinical symptoms; (ii) majority of the infected persons show no disease symptoms but develop an effective acquired immune response and referred to have latent tuberculosis infection; (iii) a proportion of those latently infected persons might reactivate and develop the disease and known as post primary active disease.

Active TB

5–10% of infected individuals (mainly infants) will develop rapidly progressive clinical disease and referred to as primary active TB. Primary TB usually occurs within 1–2 years after the initial infection. This results from local bacillary multiplication and spread in the lung or spread in the blood, through out the body, with seeding of bacilli in various tissues and organs.

Post-primary, or secondary, TB can occur many years after infection owing to loss of immune control or the reactivation of latent bacilli. The immune response of the patient results in a pathological lesion that is characterized by localized, often extensive tissue damage, and cavitations. The characteristic feature of the active post primary tuberculosis

includes extensive lung destruction with cavitation, positive sputum smear, and upper lobe involvement. Patients with these lesions are the main transmitters of infection.

Latent TB infection

Majority of MTB infected persons will not develop any clinical symptoms, and are referred as to have "latent" TB infection. Latent TB is defined on solely on the evidence of sensitization by mycobacterial proteins: a positive result in either TST or an *in vitro* interferon gamma (IFN-γ) release assay (Pai et al., 2008). However, TST and IFN-γ release assays do not distinguish latent TB from active disease. In contrast to the risk of MTB infection, which is largely determined by an exogenous factor, the risk of TB following MTB infection is largely endogenous, determined by the individual's immune status. A significant improvement in the understanding of the MTB latency is necessary to be able to answer question on the control of TB.

The bacilli are then thought to enter a state of latency based resistant bacillary forms capable of surviving under stressful conditions generated by the host (Cardona, 2007). The actual physical location of bacilli during latent infection remains a topic of considerable debate over a decade and whether latency is synonymous to chronic persistent infection with viable MTB in all instances is not known. The knowledge of physical nature of MTB during latency is important to determine treatment, post exposure vaccine and diagnosis. The two common viewpoints, (although not mutually exclusive), that exists on latent TB are: (1) MTB persists mostly in a "lazy" state within granulomatous lesions, periodically recrudescent, a dynamic immunological interplay between MTB and the host, and a continuous recruitment of cells into the granuloma. Reactivation occurs when this dynamic cellular exchange becomes dysregulated; (2) dormant MTB reside within alveolar epithelial cells in the lung, with reactivation being associated with the upregulation of

resuscitation promoting factors within MTB and the escape of newly dividing microorganisms into alveoli and bronchi (Ehlers, 2009).

Spectrum of tuberculosis infection

As a complication of purely clinical definition of active and latent tuberculosis, TB is commonly thought of as having a simple binary distribution. Although latent TB is generally related to bacterial containment in some inactive form, the current definition of latent TB includes a diverse range of individuals from those who have completely cleared the infection to those who are incubating actively replicating bacteria in the absence of clinical symptoms (Young et al., 2009), as also evident from animal models (Lin et al., 2009b).

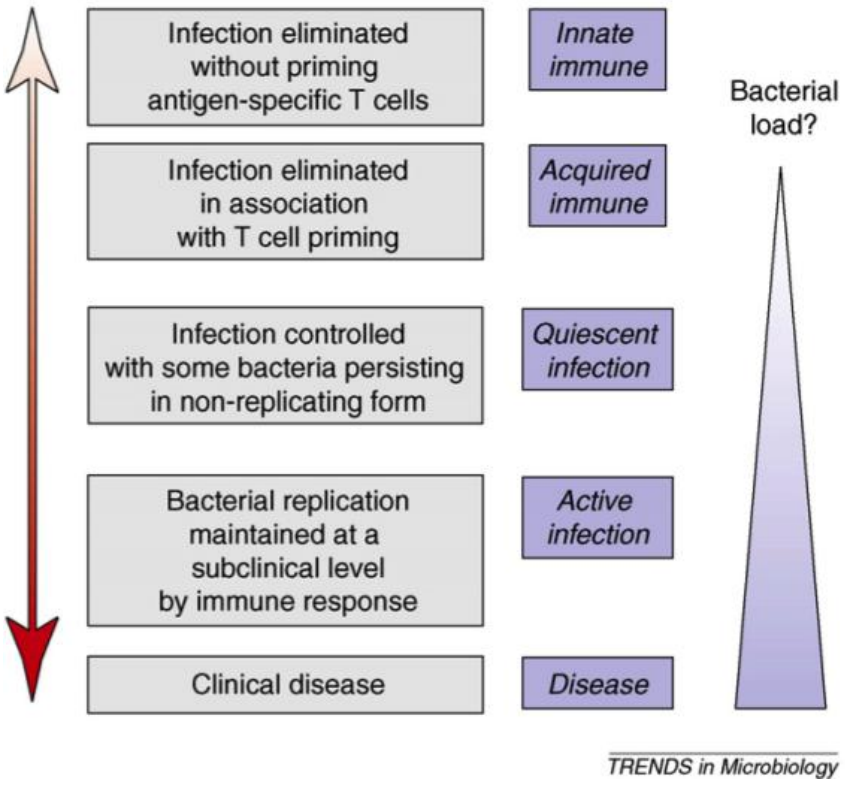


Figure 2 Tuberculosis infection as a spectrum (Young et al., 2009)

Likewise, a range of diverse pathological presentation characterizes active TB in humans and nonhuman primates, ranging from sterile tissue to caseous hypoxic lesion containing variable number of bacteria, to liquefied cavities with a massive load of replicating

bacteria. Therefore it is proposed that the MTB infection may be better viewed as a continuous spectrum extending from sterilizing immunity, to subclinical active disease, to fulminant active disease corresponding to partially overlapping regions of biological heterogenicity (Barry et al., 2009; Young et al., 2009).

Current Vaccine: BCG

Bacille Calmette-Guérin (BCG) was one of the first live-attenuated vaccines to be used in humans. It remains the most commonly used and only available vaccine for tuberculosis worldwide, with over 120 million doses administered each year. BCG immunization is considered effective in children, providing 80% protection against severe and disseminated tuberculosis, such as tuberculous meningitis and miliary disease (Trunz et al., 2006; Walker et al., 2006). BCG immunization also reduces the risk of tuberculosis in adults, by an average of 50% as shown in a recent meta-analysis, however various reports suggests wide range of efficacy in different populations (Brewer, 2000; Colditz et al., 1994).

Comparative genomics studies investigated the genetic diversity of BCG vaccine strains. In addition to the regions of difference, discovered in avirulent BCG in relation to *M. tuberculosis* and virulent *M. bovis* (Mahairas et al., 1996), genetic divergence with various BCG vaccine strains were also found. These can be characterized by the presence or absence of RD2, RD14, RD15 and RD16 (Behr and Small, 1997; Behr et al., 1999; Brosch et al., 2002; Gordon et al., 2001; Mahairas et al., 1996). BCG vaccines were thus divided in to early strains, represented by BCG Japan, BCG Birkhaug, BCG Sweden, and BCG Russia; and late strains including BCG Pasteur, BCG Danish, BCG Glaxo and BCG Prague (shown in Figure 3) (Brosch et al., 2007). Potential influence of these genetic differences, through antigenic variation, on the protective immunity and efficacy of BCG immunization with various vaccine strains has generated considerable concerns internationally (Knezevic and Corbel, 2006).

Behr and Small, first suggested that attenuation of BCG vaccine strains induced progressive loss of its efficacy and therefore it fails to mimic the natural infection of tuberculosis (Behr and Small, 1997). The genetic diversity of different BCG vaccine strains has also been shown to have variable effect on immunogenicity (Aguirre-Blanco et al., 2007; Behr, 2002; Hussey et al., 2002; Mostowy and Behr, 2002; Narayanan, 2006; Vijayalakshmi et al., 1995; Wu et al., 2007). Most of the meta analysis of BCG vaccine efficacy, however used data set obtained from the use of late BCG strains (Colditz et al., 1994).

In South Africa, Davids et al., showed that BCG Japan induced higher frequencies of IFN-γ producing CD4 and CD8 T cells, greater secretion of IFN-γ, TNF and IL-2 and greater T cell proliferation, than BCG Danish (Davids et al., 2006). This led to the speculation that early BCG vaccines strains might confer better protection against tuberculosis (Brosch et al., 2007).

Recently, immune response to various BCG vaccine strains used world wide in both animal models and in humans has been reviewed. In animal models, BCG-Pasteur, BCG-Denmark and BCG-Glaxo were found to associate more with better protection in mice and BCG-Danish more effective in guinea pigs. In humans, only small number of studies compared the immunological response between the use of different strains, however the results were either contradicting or inconclusive, and insufficient data available, amidst a good evidence to support the perception that the induced immune response and protection afforded against TB differs between BCG vaccine strains. The authors concluded that there is an urgent need for well designed, adequately powered studies to compare the immune response induced by different, genetically distinct BCG vaccine strains in humans, to better understand the in vitro correlates of protective immunity against tuberculosis (Ritz et al., 2008).

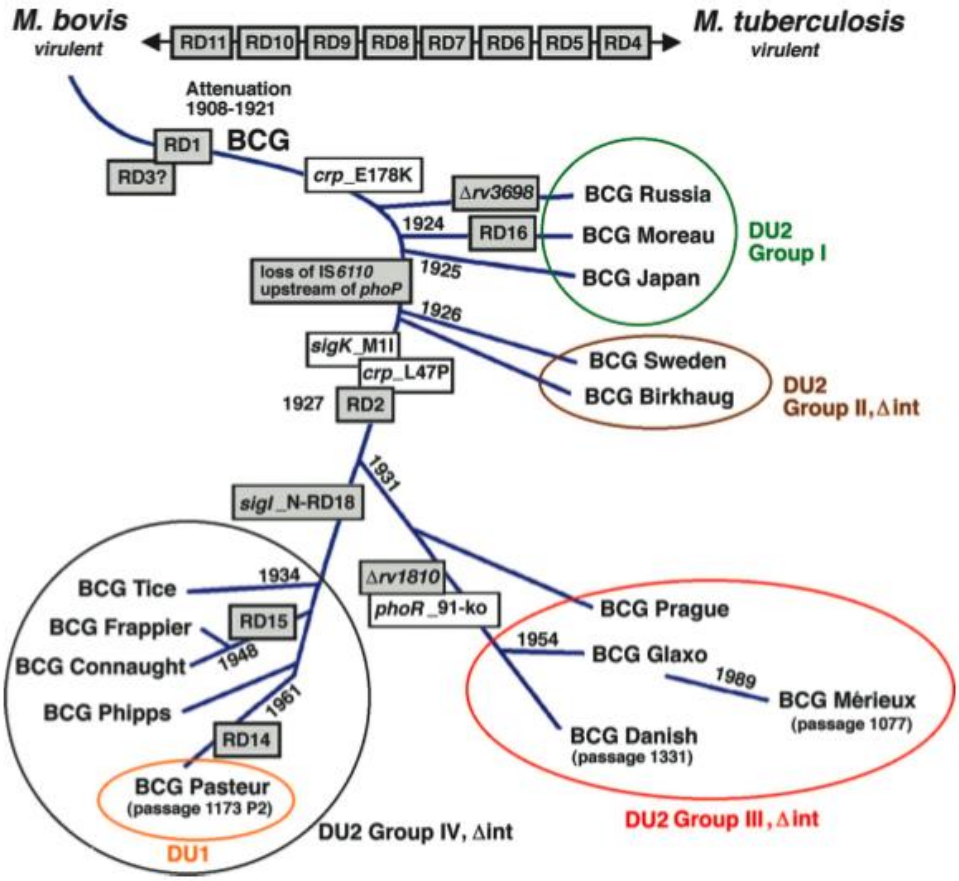


Figure 3 Phylogeny of BCG vaccine strains (Brosch et al., 2007)

With potential draw backs in all control measures, there is an urgent need for new immunodominant antigenic candidates to be developed as new vaccine and / or diagnostic tests: which could be potentially (i) useful as a post exposure prophylaxis in the latently infected persons, (ii) differentiate active and latent tuberculosis, (iii) used as diagnostic marker to identify those at high risk of developing active tuberculosis, and (iv) used to monitor treatment response.

Correlates of protection

In humans, correlates of protection against TB are not completely known. Studies conducted in relevant animal models indicate that the cellular immune response, both in

the CD4+ and CD8+ T cell subsets (Lazarevic et al., 2005), are essential components of an effective immune response, but very little is known about the specific phenotypes of these cells and about the M. tuberculosis antigen targets (Delogu and Fadda, 2009). Moreover, it is thought that the humoral response against *M. tuberculosis* may not be relevant for protection (Kaufmann, 2001).

The majority of T cells at the disease site in humans are CD4+ T cells, as their primary role is to produce cytokines that assist and orchestrate other immune cells in the environment. TB-specific CD4+ T cells produce primarily type 1 cytokines, or Th1 cytokines, which include interferon-gamma (IFN-γ), IL-2 and tumour necrosis factor (TNF) (Hanekom et al., 2007).

In murine models, roles of CD4 T cells, IFN-γ and tumor necrosis factor (TNF) in controlling acute and latent MTB have been demonstrated (Botha and Ryffel, 2003; Flynn et al., 1993; Flynn et al., 1995; Mogues et al., 2001). In humans studies on those infected with HIV-1 provides evidence that increase in CD4+ T cell numbers with antiretroviral therapy correlate with the decrease in susceptibility to develop active tuberculosis, suggesting that these T cells play an important role in the protection against TB (Maartens and Wilkinson, 2007). The protective role of IFN-γ was best demonstrated by the fact that people with defective type 1 cytokine pathway poorly control non pathogenic mycobacteria (Newport et al., 1996). Administration of anti-TNF monoclonal antibody infliximab treatment was shown to reactivate TB in some patients (Keane et al., 2001).

Although IFN-γ is essential to human defense against mycobacteria, it is increasingly recognized that assay of PBMC secretion of IFN-γ is a poor correlate of protection in field studies of tuberculosis (Hanekom et al., 2007). Greater attention to markers such as IL-2, which might better reflect immunological memory, is now being paid. The emerging picture is that distinct IFN-γ /IL-2 functional profiles correlate with different models of

infection. In a study assessing the capacity for IFN-γ and IL-2 secretion by MTB specific T cells in HIV uninfected persons with active tuberculosis indicated a co-dominance of IFN-γ single positive and IFN-γ / IL-2 double positive T cell, followed by a shift to a dominance of IFN-γ / IL-2 double positive and IL-2 single positive T cells after treatment (Millington et al., 2007). In HIV-1 infected persons, disease progression and increase in viral load is found to correlate with the loss of IL-2 secretory function by CD4 T cells. This novel finding implicates that IL-2 or IL-2 / IFN-γ as correlate of protection (Day et al., 2008).

Antigen discovery in tuberculosis

Most humans have the inherent potential to control tuberculosis either partially or completely, following an exposure or infection. Evidence suggests that in those infected with *Mycobacterium tuberculosis*, 90% control the pathogen and prevent development of disease *via* the immune response. Therefore as suggested by Behr and Small (Behr and Small, 1997), a vaccine that could elicit an immune response similar to the natural human infection with MTB should be effective, and thereby contain bacterial multiplication, dissemination, limit tissue damage, and block the development of the disease. Identifying antigenic candidates capable of inducing TB specific cellular immune responses are thus crucial for vaccine development.

Both T and B cell are thought to play roles in conjunction with each other in the immune regulation. The discovery of immunodominant antigens in *M. tuberculosis* has hitherto largely been based on dominance in antibody responses that are not the basis of protection against tuberculosis.

Antigen discovery in tuberculosis is one of the most promising areas of current research priorities. Advances in the technologies such as genomics, proteomics and informatics provide substantial support to study various aspects of host-bacilli interaction (de Jonge et

al., 2005; Lin and Ottenhoff, 2008a; Vordermeier et al., 2009; Zvi et al., 2008). Empirical evaluation is still considered ultimately the most reliable methodology for such discoveries. However, technologies complementing and reducing the length and breath of experimental evaluation are widely desired.

T cell recognition

T cells are lymphocytes that recognize antigens, *via* their T cell receptors (TcR), in the form of processed peptide fragments presented with MHC molecules. The T cell subset responsible for regulating the immune response to *M. tuberculosis* is considered to be the CD4 T cells. CD4 T cell binds with MHC II molecules (described below) on the antigen presenting cells. T cells expressing CD4 molecules recognize exogeneous antigens. The complex formed by the MHC II-peptide-TcR triggers an immune response and the release of various cytokines via type 1 consisting mainly of proinflamatory cytokines (IFN-γ, TNF and IL-2) or type 2 regulatory / inflammatory cytokine (IL-4, IL-10, IL5, IL6, IL13) (van Crevel et al., 2002).

T cell recognition is a fundamental mechanism of the adaptive immune system by which the host identifies and responds to foreign antigens. Identifying T cell epitopes are important in understanding the disease pathogenesis and is crucial for the development of vaccine candidates. It has been theorized that effective vaccines can be developed using the minimum essential subsets of T cell epitopes (Lafuente and Reche, 2009).

Components of T cell recognition

Major histocompatibility complex (MHC) proteins, also known as human leukocyte antigens (HLA), play a vital role in the adaptive immunity mediated by the T cells (Dimitrov et al., 2010). MHC molecules are glycoprotein which bind short peptides, derived from host and/or pathogen proteins and presents them to T cells. There are two types of MHC molecules: MHC class I generally expressed in all the nucleated cells. They

bind to peptides that are processed from endogenously synthesized antigen and are recognized by CD8 T cells. In contrast MHC class II molecules are expressed on most professional antigen presenting cells (APC), and they bind to peptides that are processed from extracellular antigens, and are recognized by the CD4 T cells. MHC (HLA) gene expression is codominant, and an individual typically can express 6 different MHC class I molecules and 12 or more MHC class II molecules. In the entire population, there are over a hundred of different MHC alleles rendering the HLA locus the most polymorphic gene system known (Terasaki et al., 2007).

In the MHC class II molecules, the peptide-binding grove is open, allowing the N- and C terminal ends of the peptide to extend beyond the binding grove. Therefore the MHC II-bound peptides vary widely in length (9-22 residue), although only a core of 9 residues (peptide binding core) fits in to the MHC class II binding groove. The MHC class II binding grove has pockets that are outlined by polymorphic residues in which peptide side chains are nestled. Therefore, positions 1 and 9 of the peptide-binding core of MHCII-bound peptides are anchor positions (Wang et al., 2001). A major contribution however is due to a set of conserved hydrogen bonds that form between the amino acid components of the peptide core and the MHCII molecule (Lafuente and Reche, 2009). In addition, the peptide binding core flanking residues (PFR) also contribute to the overall binding and the resulting T cell response (O'Brien et al., 2008; Wilkinson et al., 1997a)

The most important factor that contributes to the immune response to a protein is that the antigen-presenting cells (APC) must process and present T cell epitopes, which in turn triggers a T cell response. T cells recognize linear epitopes, which are derived from proteins that are processed by the APC. During the process, large numbers of peptide fragments are formed. However, only 2 % of all the fragments might have the right amino acid side chain that would allow them to bind to the MHC binding grove and be presented on the surface of APC. One of the critical determinants of the immunogenicity is the

strength of T cell epitope binding to MHC molecule. Only peptides with the higher binding affinity to the MHC molecules are more likely to be recognized by the T cell receptor (Weber et al., 2009).

T cell Epitope prediction

The ability of MHC class II molecules to bind peptides has been traditionally assayed *in vitro,* using purified glycoprotein. Competition assays between the labeled ligand-peptide and the competitor peptides, using enzymatic detection system, were the most reliable and commonly used methods of comparing peptide-affinities. Assays utilize the direct binding of labeled peptides to live antigen presenting cells have the advantage that the binding occurs in a native environment (Buus, 1999). The peptide binding motifs have been studied by comparing peptide analogues with multiple single amino acid substitution or C – or N-terminal truncation.

Identification of T cell epitopes requires the synthesis of overlapping peptides that span the entire length of the protein, followed by experimental assay for each peptide, such as proliferation assays by thimidine incorporation (Jurcevic et al., 1996), measurement of IFN-γ and IL-2 secretion by ELISA or ELISpot (Pathan et al., 2001), Intracellular cytokine staining (Lalvani et al., 1998) or Tetramers (Altman et al., 1996), allowing the identification of a small number of T cell epitopes. Although these methods allow accurate determination of the breadth of T cell responses *in vitro*, the systematic T cell epitope mapping of larger proteins is costly; cumbersome and restricts the ability to assay multiple protein candidates.

Computer aided epitope prediction

Alternative computational approaches have been developed for the prediction of T cell epitopes, which aim to reduce the burden that is associated with the empirical epitope

identification. These bioinformatic tools promise rapid movement from genome sequence to antigen discovery and vaccine design (Tsurui and Takahashi, 2007).

There are multiple methods and models available for predicting MHC-peptide binding, which differ in the techniques and algorithms that are applied. In general, Peptide binding algorithms can be classified into 2 larger groups: (i) those that are generated from sequences of peptide binders (sequence based models or peptide-data driven models) and (ii) those that do not require any peptide binding data and are based solely on the sequence and structure of MHC molecules.

The sequence based models in turn can be classified in to two types. (a) Those based on qualitative data that describe the ability of peptides to bind to MHC molecules (binders and non-binders); referred as binding pattern recognition models. These models predict, with some certainty, whether a test peptide binds to a given MHC molecule. (b) Those trained on quantitative data consisting of peptides whose binding affinity for MHC molecules has been determined; referred as quantitative binding affinity models. These models target the prediction of actual binding affinity for the peptides to MHC. The MHC binding peptide prediction algorithms are described in Figure 3 and the methods are reviewed in (Lafuente and Reche, 2009).

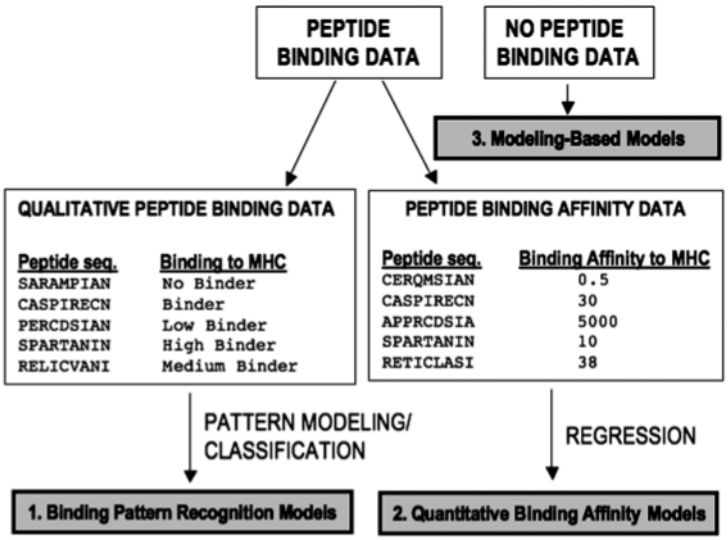


Figure 4 Methods of MHC binding peptide prediction algorithms (Lafuente and Reche, 2009).

Evaluation of MHC class II peptide algorithms

MHC prediction algorithms are comprehensively and critically reviewed and suggest, the potentially advantages and drawbacks of each prediction algorithm, and significant progress has been reported with regard to *in silico* predictive accuracy and the range of MHC molecules that it can be targeted for peptide binding predictions. Peptide-data driven models were thought to overwhelmingly outperform modeling based methods (Lafuente and Reche, 2009; Nielsen et al., 2010; Stern and Calvo-Calle, 2009; Tsurui and Takahashi, 2007). Several studies were carried out using experimental evaluation and compared the performance of these prediction algorithms with *in vitro* screening methods. Although these studies show some splendid examples of the utility of these methodologies, they also describe high levels of discrepancy in the prediction efficiency especially in the MHCII prediction system. Therefore, it has been widely suggested that until an efficient stand-alone algorithm becomes available, these techniques need to be used with caution (Gowthaman and Agrewala, 2008; Gowthaman and Agrewala, 2009; Lin et al., 2008).

Utility of in silico methods in the context of Mycobacterium tuberculosis antigen discovery

With the advent of the whole genome sequence of MTB and the description of various genomic regions of interest, MHC-peptide binding prediction algorithms are widely used in the antigen discovery in human and bovine tuberculosis, in addition to other infectious diseases (Mustafa, 2009a; Mustafa, 2009b; Vordermeier et al., 2009; Wang et al., 2009a). Computational methodologies were used in the selection of 189 vaccine candidates from whole genome (3989 ORF) analysis of MTB (Zvi et al., 2008), and to identify novel diagnostics (Bertholet et al., 2008; Ewer et al., 2006; Hewinson et al., 2006).

Chapter 5 in this thesis describes the use of *in silico* methodology to predict MHCII-binding peptides using the ProPred algorithm. ProPred (Singh and Raghava, 2001) (http://www.imtech.res.in/raghava/propred/index.html) web interface allows to predict MHC Class-II binding regions in an antigen sequence, using virtual quantitative matrices as described by Sturniolo et al., (Sturniolo et al., 1999) and assists in locating promiscuous binding regions. Quantitative matrices are the most widely used models in predicting MHC-peptide binding. This has been developed by the use of peptide libraries to obtain a quantitative representation of the interaction of all natural amino acid residues with structurally identified binding pockets of HLA-DR molecule (pocket binding profile). ProPred allows prediction of all 51 HLA alleles, performs analysis for each HLA-DR allele independently and predicts the binding strength. Besides the server also has the option to plot the threshold profile (threshold versus binding peptides), which can assist the users in selecting an appropriate threshold for locating promiscuous binders. In addition to analyzing the best scoring peptide, the server computes the binding strength of all peptides in the selected subsequence and presents a calculated binding score.

ProPred has been critically evaluated for its performance with empirical testing in comparison with other *in silico* prediction algorithms. RANPKPEP, SVMHC and ProPred were found to perform better than other programs for prediction of peptides binding to HLA-class II molecules (Gowthaman and Agrewala, 2008), while NETMHCIIPAN, ProPred, IEDB and MULTIPRED were found to be best predictors in vaccine research (Lin et al., 2008).

Empirical techniques used in antigen discovery

Assay of the T cell response to antigens of *Mycobacterium tuberculosis* is critical markers in the evaluation of new antigenic candidates. The immunogenicity of such candidate

antigens will be critical to decide which should move on to further studies, and the ability to compare results would be an important aspect. However, these potential comparisons are confounded by variation in individual laboratory approaches that use different readout methodologies to define the immunogenicity of such candidates. These include:

1. Nature of cells to be stimulated: whole blood or PBMC
2. Nature of candidate antigen: overlapping synthetic peptides or recombinant proteins
3. Time of culture: short term (< 24 hours), intermediate (2-3 days) or prolonged (>5 days)
4. Methods of read out: ELISA or ELISpot or FACS.

These read out methodologies that are most widely in use; and the factors that influence the results of immunogenicity evaluation of candidate antigens are reviewed below.

Use of Whole blood or PBMC

Time to Incubation of whole blood and isolation of PBMC

Time to incubation of whole blood and the isolation of PBMC are considered critical for the T cell-specific evaluation of antigens. The maximum time delay from the time of collection to the setting up of assay is 4 hours in our laboratory. However, delay in the incubation > 2 hours were reported to reduce T cell IFN-γ production in short term cultures (Hanekom et al., 2004). However, It has been shown that prolonged incubation time assays were not affected (Lin et al., 2007).

Whole blood (Agger et al., 2003; Black et al., 2009; Black et al., 2002; Cockle et al., 2002; Ewer et al., 2006; Hanekom et al., 2008; Sidders et al., 2008; Weir et al., 1994)	**PBMC** (Bertholet et al., 2008; Brock et al., 2004; Hanekom et al., 2008; Jurcevic et al., 1996; Leyten et al., 2006; Wilkinson et al., 1998a; Wilkinson et al., 1998b)
Advantages • Smaller blood volumes required • Assessment of a all types of peripheral blood leukocytes • The measured T cell response will reflect cellular and soluble components that affect antigen presentation and T cell activation • Requires relatively few resources • Assessment may often be completed in 2 phases: (i) immediate incubation with antigen, (ii) cryopreservation of plasma, fixed white cells (after red cell lysis) or RNA for later analysis.	**Advantages** • Cleaner system for the study of novel antigen responses. • PBMC-based assays are incubated on a per-cell basis and the results are corrected to number of cells used; resulting in increased sensitivity in T cell depletion, such as in HIV-1 infection. • Allow for a description of T cell responses without the influence of other whole blood components. • Cryopreserved PBMC also enables batched analysis, which is of particular importance for reducing variability in antigen screening and also in longitudinal vaccination studies • Cryopreserved PBMC are available at a later date for application of scientific advances and re evaluation of novel finding.
Disadvantages • Exact cell numbers are not known, T cell depletion therefore results in poor sensitivity • Measurement of response reflects not only T cells, but other peripheral blood components • Need for immediate processing of collected blood limits versatility in what may be measured	**Disadvantages** • PBMC isolation results in considerable cellular loss and therefore require larger blood volumes • Culture of PBMC requires additional nutrients in the forms of foetal calf or unrelated human serum; however, serum-free media may also be used • In short term assays, the cellular activation as a result of PBMC isolation may result in higher backgrounds, compared with similar incubation of whole blood. • Optimal procedures such as liquid nitrogen storage are critical for long-term storage of PBMC.

Comparison of whole blood and PBMC based culture systems

The comparability for the use of whole blood and PBMC based culture systems has been tested in tuberculosis and in other conditions. In a study comparing tuberculosis specific (ESAT-6 and CFP-10) IFN-γ response in overnight in house ELISpot using PBMC and 72 hours whole blood culture, showed good correlation ($r=0.689$, $p\leq0.0001$) and an agreement of 81% ($k=0.59$) (Scholvinck et al., 2004). However, when either in-house ELISpot (Adetifa et al., 2007) or commercial T.SPOT (PBMC) (Lee et al., 2006) was compared with the commercial QFT-GIT (whole blood), the PBMC based ELISpot / T.SPOT were found to have higher sensitivity than whole blood based QFT-GIT for detecting TB disease (Adetifa et al., 2007) and MTB infection (Lee et al., 2006) while the QFT-GIT was suggestive of higher specificity for detecting MTB infection (Lee et al., 2006).

Effect of cryopreservation on T cells

The frequencies and type 1 / type 2 cytokine signatures of recall antigen specific CD4+ and CD8+ T cell were shown to be unaffected after cryopreservation as examined by ELISpot (Kreher et al., 2003). The use of media additives and the temperature of the wash medium for thawing (37°C) were found to be important factors for maximizing the retention of antigen specific lymphocyte function after cryopreservation (Disis et al., 2006; Maecker et al., 2005).

Nature of stimulating antigen

Recombinant proteins (Chen et al., 2009; Leyten et al., 2006; Lin et al., 2009a; Schuck et al., 2009; Wilkinson et al., 1997b)	**Synthetic overlapping peptides** (Aagaard et al., 2004; Ewer et al., 2006; Jurcevic et al., 1996; Millington et al., 2007; Wilkinson et al., 1997b)
▪ Offers the advantage of studying T cell response as a whole protein. ▪ To study functional, structural studies of the Protein, protein-protein interaction and enzyme kinetic. ▪ The production of recombinant proteins, the purification required for use in T cell assays is laborious and costly.	▪ The advantage to use mixture of overlapping peptides include synthesise and purification in larger quantities, low cost as compared to the recombinant proteins. ▪ Allow evaluation of dominant epitope regions within the proteins. ▪ Allows design and selection of promiscuous HLA presentable regions of the protein and thereby reducing the cost of impracticability of larger number of proteins (Jurcevic et al., 1996).

Duration of incubation

The fundamental difference in the assays used for the evaluation of antigen candidates in various laboratories is the duration of culture. The assays may have:

Short incubation: assays measure the cytokine production by the short lived effector cells that are particularly well detected by ELISpot and Flow Cytometry.

- 6-12 hours (Whole blood/PBMC-FACS): For flow cytometry analysis, whole blood or PBMC are typically incubated with specific antigens for 6 to 12 hours. Brefeldin-A or Monensin is added for the last few hours (5 hrs) of incubation to capture cytokines intracellularly. The cells may be cryopreserved before further analysis. The advantage of this assay is that multiple parameters may be measured at the single cell level (Bertholet et al., 2008; Hanekom et al., 2004; Soares et al.,

2008; Tena-Coki et al., 2010). The cost and resources required particularly for multiparameter analysis remains a major drawback.

- 16-24 hours (Whole blood / PBMC): Most widely used incubation time for ELISpot assays, and also for ELISA and FACS.

Intermediate term incubation:

- 48-72 hours: Intermediate time assays measure both effector and memory T cell response. These are widely used to culture PBMC or whole blood and to measure the cytokine secretion in the supernatent (Scholvinck et al., 2004), but also been used for ELISpot based assays to identify response to antigens (Bertholet et al., 2008).

Prolonged incubation:

- >5 days (whole blood, PBMC culture / Lymphocyte stimulation test and proliferation assay): assays measure the response of central memory T cells that are required to differentiate before producing enough cytokine response. Whole blood (diluted) or PBMC is cultured with specific antigens and incubated for 5-7 days, supernatents harvested and secreted cytokines are measured by ELISA or multiple cytokine measurement technologies (Black et al., 2002; Hanekom et al., 2008; Weir et al., 1994).

  The advantages of these assays include relative ease, and the low volume of samples necessary allowing a degree of automation and processing of large sample numbers

- Cultured ELISpot assay: where PBMC is cultured with antigenic stimuli (+/- IL-2) for prolonged periods (6-9 days), cells were washed and used for ELISpot assay.

Cultured ELISpot were shown to reveal more durable memory T cell response (Todryk et al., 2009)

- Proliferation of PBMC is conventionally used with thymidine incorporation, by BrdU incorporation or by other non-isotopic stains (Roche). PBMC proliferation and soluble cytokine production has been extensively described for mycobacterial antigen studies (Jurcevic et al., 1996; Wilkinson et al., 1998a) and in novel vaccine trial (Beveridge et al., 2007). However, Thymidine incorporation, while sensitive, provides a single outcome, which cannot be combined with other markers, and leads to the generation of radioactive waste, with consequent disposal costs and hazards.

Correlation of assays with various incubation time points

Major discrepancy was reported between the results of two short-incubation assays (ELISA, ELISpot, FACS) and prolonged incubation time assay (6 day lymphocyte stimulation test or cultured FACS) using the same antigen stimulation. These differences have been attributed mainly to the variation in time duration of incubation (Leyten et al., 2007; Schuck et al., 2009). Discordant results was also reported between overnight PBMC based ELISpot and 72 hour whole blood assay (Scholvinck et al., 2004). However these differences were suggested to be a result of the measurement of various components of the T cell response.

Read out methods

The most widely used read out methods in antigen evaluation are (i) ELISpot, (ii) ELISA and (iii) FACS.

ELISpot (Enzyme-linked immunosorbent spot) uses nitrocellulose membranes and anti cytokine monoclonal antibodies for detection of T cell response. It is a commonly used technique for quantifying the occurrence of T lymphocytes cells secreting a cytokine, after

stimulation with an antigen or peptide. It measures the number of spot forming cells (SFC) [cytokine producing cells] at a specific concentration of effector cells. Cytokine secreting T cells are detected as coloured spots *via* staining with an enzyme linked anti-cytokine monoclonal antibody. The spots can be enumerated with the naked eye, however the development of ELISpot readers have automated the enumeration of spot forming cells and facilitated the ongoing development of standardized protocols. The assay was shown to have high sensitivity for detecting fewer antigen specific T cells per million PBMC. This techniques has been widely used in antigen evaluation and screening (Lalvani et al., 2001b; Bertholet et al., 2008; Chen et al., 2009).

ELISA (Enzyme-linked immunosorbent assay) measures the quantity of cytokine secreted in to the culture supernatants (PBMC or whole blood) in response to the antigen stimulation. This is measured as the intensity of optical density produced by the enzyme liked substrate solution, read at an optimal wavelength. ELISAs for a cytokine (such as IFN-γ) is most widely used assay, to measure the response to antigen in medium to long term culture assays (Leyten et al., 2006; Wilkinson et al., 2005a).

Fluorescence-assisted cell sorting (FACS): Flow cytometry and intracellular cytokine staining enables the simultaneous measurement of multiple cell surface antigens and cytokine producing capacity at a single cell level. It also enables the measurement of the proportion of specific subtypes within a population of cells and the absolute number of cells in a volume of whole blood. Flow cytometry in conjunction with fluorescent conjugated monoclonal antibodies (mAbs) are used to characterize distinct cell types and activation states. The use of anti-cytokine antibodies in a staining technique that allows the intracellular detection of cytokines and cell surface molecules, without the separation of cells, has greatly enhanced the use of flow cytometry to determine immunophenotype in human cells, particularly T cell response to antigenic stimuli.

Strategies in Antigen discovery

There have been many strategies for identification of mycobacterial immunodominant antigens using post genomic methods including, proteomic identification, T cell epitope prediction, genomic comparison to identify pathogen specific open reading frames, and biology driven (expression under certain conditions). For the purpose of this thesis, further descriptions are restricted only to the later two.

Genomic location: Role in T cell antigen discovery

The major landmark in the T cell based antigen discovery of *Mycobacterium tuberculosis* was the description of distinct genomic regions of difference. Mahairas et al., (Mahairas et al., 1996) described 3 genomic regions of difference (designated RD1 to RD3) found to be deleted in avirulent BCG, when studied by subtractive genomic hybridization to identify genetic differences with virulent *M. bovis* and *M. tuberculosis* and avirulent BCG. Of them, it was found that RD1 has been conserved in all virulent laboratory and clinical isolates of *M. bovis* and *M. tuberculosis*. The reintroduction of RD1 into BCG was found to initiate expression of at least 10 proteins and was found to resulted in a protein expression profile almost identical to that of virulent *M. bovis* and *M. tuberculosis* as determined by two-dimensional gel electrophoresis (Mahairas et al., 1996). Another significant finding was also that RD2-encoded segment was found to be deleted only from sub strains derived from the original BCG Pasteur strain after 1925. In addition they also identified that an ORF (ORF1C) in the RD1 encoded region contained previously described novel, immunogenic antigen esat-6 (Andersen et al., 1995; Boesen et al., 1995; Sorensen et al., 1995) and ORF2G in the RD2-encoded region contained Mpt-64 (Li et al., 1993) a secreted protein absent in BCG and reported to be absent from some BCG vaccine strains (Mahairas et al., 1996)

Since the completion of the whole genome annotation of *Mycobacterium tuberculosis* (Cole et al., 1998) several different approaches have been employed to compare the genomes of the members of MTB complex, extending from various DNA array technologies (Behr et al., 1999; Gordon et al., 1999a; Gordon et al., 1999b) to highly sensitive whole genome comparison (Brosch et al., 2007; Gordon et al., 2001) which detected the full range of polymorphisms and gene rearrangements. Many of these studies compared the virulent and avirulent strains in the hope of understanding the difference linked to changes in pathogenesis.

Several genes encoded by the regions of difference, attracted researchers as they encoded potential virulence factors similar to those characterized for other microbial pathogens. Pym *et al*., further experimentally demonstrated that reintroduction of RD-encoded genes in to BCG does result in a significant increase in virulence (Pym et al., 2002) and Lewis KN *et al.,* demonstrated that deletion of RD1 from MTB resembles attenuated BCG (Lewis et al., 2003).

Two RD1-encoded, co regulated secreted proteins ESAT-6 and CFP-10 (Berthet et al., 1998) have attracted considerable immunological interest as a result of potent antigenicity for T and B cells. Mice and Guinea pigs vaccinated with the recombinant strain of BCG complemented with ESAT-6 and CFP-10 were found to be better protected against challenge with *M. tuberculosis*, showing less severe pathology and reduced dissemination of the pathogen, as compared with control animals immunized with BCG alone (Pym et al., 2003).

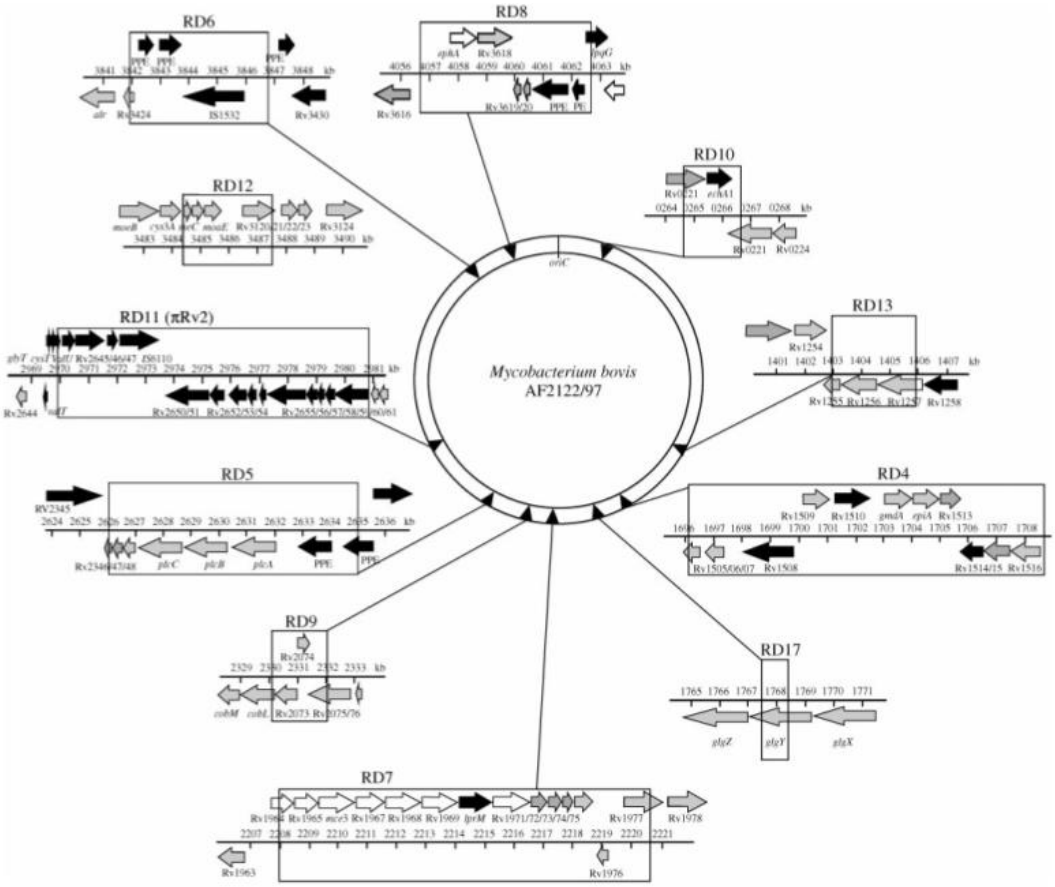


Figure 5 Deletions in the *M. bovis* genome in relation to MTB (Gordon et al., 2001)

ESAT-6 and CFP-10 have been extensively investigated as T cell targets for immunodiagnosis of active tuberculosis (Arend et al., 2000a; Lalvani et al., 2001b; Mustafa et al., 2000; Ravn et al., 1999; Vincenti et al., 2003). Both antigens were found to be frequently recognized *in vivo* and *in vitro* based on the induction of delayed-type hypersensitivity responses and the ability to induce interferon-gamma production by lymphocytes, respectively. The combination of ESAT-6 and CFP10 was found to be highly sensitive and specific for both *in vivo* and *in vitro* diagnosis (van Pinxteren et al., 2000). Studies were also carried out to differentiate MTB infection with that of *Mycobacterium avium* complex (Lein et al., 1999) and to differentiate between BCG vaccinated with the hope to replace the non-specific PPD based Skin test.

A few other RD encoded proteins (other than RD1) were evaluated to find potential candidates to be used in diagnostic tests to improve specific diagnosis of tuberculosis.

In bovines, 13 ORFs from RD1, RD2 and RD14 regions of MTB genome was evaluated using overlapping peptide pools spanning the ORF were studied in *M. bovis* infected, BCG vaccinated and control cattle's and 8 immunodominant antigens (Rv1983, Rv1986, Rv3873, Rv3878, Rv3879c, Rv1970c, and Rv1769) were identified that could either differentiate BCG vaccinated and *M. bovis* infected or improve the specificity than PPD (Cockle et al., 2002).

RD2 and RD11 encoded proteins such as MPT-64, CFP-21, nrdf1, PPE57, PPE59, Rv1978, were found to increases the sensitivity and specificity in detecting active TB when PPE57 or nrdf1 was used in addition to ESAT-6 or CFP-10 in IFN-γ based ELISpot (Chen et al., 2009). CFP21-MPT64 fusion protein was found to improve the diagnosis of latent TB with increased sensitivity and specificity than TST (Fu et al., 2009).

RD11 encoded Rv2654c was found to specific for the diagnosis of active TB (Aagaard et al., 2004) and one dominant peptide (p4) has been included in the commercialized IGRA QFT-IT to improve the specificity (Brock et al., 2004; Parkash et al., 2009).

Al-Attiyah and Mustaffa, evaluated cellular immune response induced by 11 RD-encoded antigens (RD1, RD4, RD5, RD6, RD7, RD9, RD10, RD11, RD12, RD13 and RD15) using synthetic overlapping peptides covering the entire region. The authors conclude that RD1 encoded antigens induced the best overall Th1 response in both active and latent tuberculosis; and that RD-encoded could be divided into two group, that preferentially induces the secretion of either IFN-γ (RD1, RD5, RD7, RD9 and RD10) or IL-10 (RD12, RD13 and RD15) suggesting a possible role in protection and pathogenesis of tuberculosis (Al-Attiyah and Mustafa, 2008).

Whilst genomic region based candidates offered MTB specific antigenic candidates and the potential of development into vaccine candidates, they however did not have the ability to differentiate various stages of *Mycobacterium tuberculosis* infection.

MTB biology driven antigen discovery

An alternative genome based strategy of antigen discovery is to relate what is highly expressed by the bacilli *in vivo* or *in vitro* and there by available as an antigen, and relate it to what is recognized *in vivo*.

It is believed that the bacilli themselves adapt to the stress condition upon the encounter under the adverse conditions created by the host immune response (latent TB). The ability to resume growth when conditions become favorable within the granuloma is however retained. These persistent bacteria are thought to have encounter depletion of nutrients, shifts in pH, production of growth limiting products, and depletion of oxygen and led to the various models of latency (Zhang, 2004), where *in vitro* bacterial cultures are subject to various stress conditions like hypoxia (Wayne and Sohaskey, 2001), nutritional starvation (Betts et al., 2002), acidic pH (Roxas and Li, 2009), nitric oxide, and high temperature (Stewart et al., 2002). MTB subject to these culture conditions (stress) were shown to induce distinct set of gene expression profiles (Betts et al., 2002; Roxas and Li, 2009; Rustad et al., 2008; Stewart et al., 2002; Zucker et al., 2001), different from normal optimum growth conditions. A detailed analysis of MTB genes that are switched on predominantly during latent stage infection was considered a research priority to lead to the identification of new antigenic targets for anti-TB strategies (Lin and Ottenhoff, 2008a; Lin and Ottenhoff, 2008b). As MTB is believed to encounter these stress condition *in vivo*, the genes that are highly induced under stress condition may also be highly expressed *in vivo*, and therefore these genes are considered to be available as potential antigens for T cell immune response and more likely antigenic. This approach has been shown to aid antigen discovery in bovines (Sidders et al., 2008).

Hypoxia

From a variety of human, animal and *in vitro* studies, it is apparent that the oxygen tension is intimately associated with the outcome of MTB infection (Rustad et al., 2009). These

observations have led to description of models of latency and reactivation that suggest a prominent role for Oxygen status (Boshoff and Barry, 2005; Gomez and McKinney, 2004; Rao et al., 2008; Via et al., 2008). These *in vitro* culture models represent proxies of conditions that tubercle bacilli are thought to encounter *in vivo* during persistence in immune competent host.

Initial hypoxic response and DosR (Dormancy Survival Regulon)

The advent of MTB genome sequence made it possible to characterize the whole transcriptomic response of MTB in response to defined *in vitro* conditions using microarrays. Sherman et al., first described the initial hypoxic response of MTB where the transcriptional profile of H37Rv was evaluated, with samples taken from aerated log phase cultures and cultures exposed to 2 hours of 0.2% O2, which consisted of about 48 genes. Many of these genes were found to play a role in the adaptation of hypoxic stress. These genes were controlled by a two-component transcriptional regulator system DosR/DosS (Sherman et al., 2001)

This was supported by the observation that the genes of DosR regulon, which are expressed as a part of adaptation to the initial hypoxic response of MTB, are upregulated *in vitro* in IFN-γ activated MTB infected murine macrophages (Schnappinger et al., 2003; Shi et al., 2003; Timm et al., 2003), and have increased level of RNA in human lung tissue in patients with active TB (Timm et al., 2003).

DosR is widely considered essential for the TB latency and therefore, led to many studies looking at the immunogenicity of these candidate genes, and were found to encode a number of MTB antigens with the potential to induce a strong T cell IFN-γ responses (Black et al., 2009; Demissie et al., 2006; Leyten et al., 2006; Roupie et al., 2007; Wilkinson et al., 1998a). Acr-1 is one of the well described dosR regulated antigen, shown to induce a dominant B and T cell response (Geluk et al., 2007; Wilkinson et al., 1998b).

In animal models, mice persistently infected with MTB were shown to induce an immune response against Rv1733c, Rv2031c, and Rv2626c (Roupie et al., 2007) Another study evaluated 51 dosR antigens in 3 African populations and found Rv1733c, Rv0081, Rv1735c, and Rv1737c to induce high positive response in MTB infected persons (Black et al., 2009).

DosR encoded immunodominant antigens were termed as "latency antigens" (Demissie et al., 2006; Leyten et al., 2006; Schuck et al., 2009), as some of these antigens were preferentially recognized, by a higher IFN-γ response, by those with latent infection, when compared to those with active tuberculosis.

Evidence of infection stage specific antigens are reported in some assay condition: Acr-1 (Demissie et al., 2006; Geluk et al., 2007; Wilkinson et al., 1998b); Rv1733c, Rv2029c, Rv2627c and Rv2628 (Leyten et al., 2006); Rv0569, Rv1733c, Rv1734, Rv2003, Rv2005c, Rv2006 (Schuck et al., 2009), were shown to induce strong IFN-γ response in those with latent tuberculosis and Rv2628 was found to be associated with cured tuberculosis and remote infection (Goletti et al., 2009).

Further studies to characterize DosR, also revealed the regulon appears expressed in conditions marked by significant bacterial replication, suggesting that the role of dosR may not be specific to latency and that other factors may also be involved. This led to further characterization of the hypoxic response in more depth revealing that the initial hypoxia response regulated by dosR contributes modestly to survival under hypoxic conditions *in vitro* and it is dispensable for virulence in mice (Rustad et al., 2008). These observations led to investigation of responses of MTB to extended hypoxia that are downstream of DosR.

Enduring Hypoxic Response (EHR) of MTB

The transcriptional profile of *Mycobacterium tuberculosis* H37Rv was evaluated through an extended time course, with samples taken from aerated log phase cultures and cultures exposed to four, eight, twelve, 24, 96, and 168 hours of hypoxia. Each hypoxic time point was analyzed with at least three biological replicates using high density oligonucleotide microarrays (Rustad et al., 2008).

This lead to the observation that the ~50 genes of the DosR regulon are maximally induced early in the time course, but gradually declined such that by 24 hours, about half the regulon was not significantly induced.

However, a second transcriptional response, much larger than the DosR regulon that was induced for much longer period, is now recognized. The number of induced genes increased until 96 hours into the experiment. The growth of these gene sets was largely found to be additive- each subsequent set contained most of the genes induced at the preceding time point. By four days of hypoxia, the number of induced genes seemed to stabilize. The enduring hypoxic response (EHR), was thus defined as the set of MTB genes not induced initially, but that are significantly up-regulated at four and seven days of hypoxia (Rustad et al., 2008).

In addition, several EHR genes were found to show overlap between genes induced by nutrition deprivation and high temperature models (Rustad et al., 2008). Therefore, the analysis of these newly described genes of the EHR were considered interesting to offer insights into the hypoxic response of MTB and the resulting bacteriostasis. In addition, EHR genes are also considered as potential drug targets (Murphy and Brown, 2008). EHR encoded proteins have not been evaluated for T cell immune response so far.

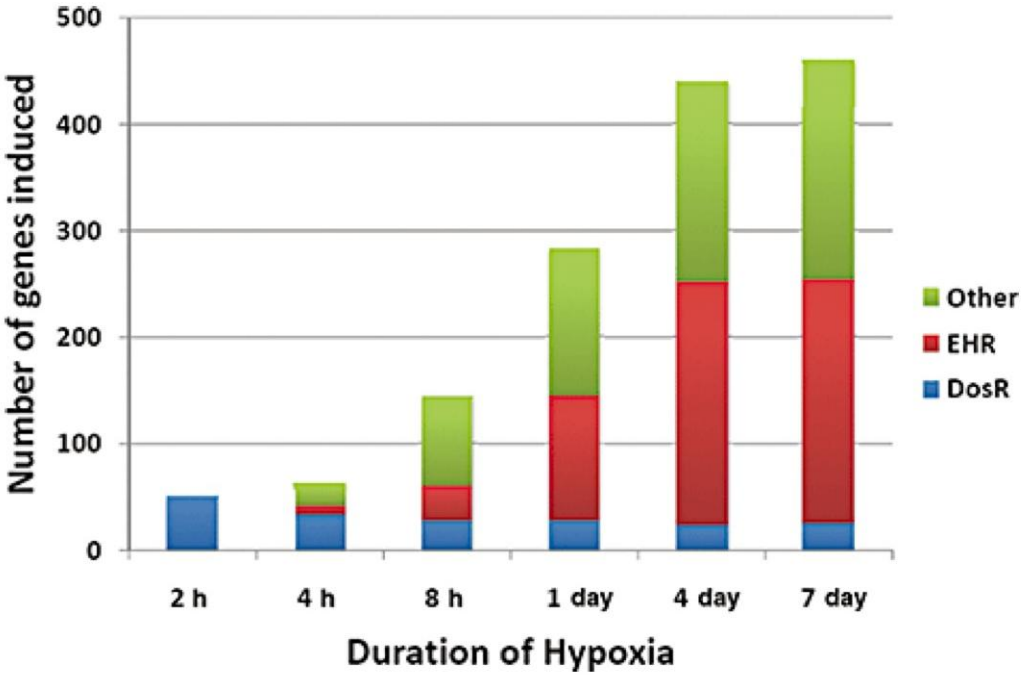


Figure 6 Enduring hypoxic response of MTB (Rustad et al., 2008; Rustad et al., 2009)

Overall Aim

This thesis presents the evaluation of *in vitro* human T cell reactivity of novel *Mycobacterium tuberculosis* antigens.

The main focus of this thesis is the analysis of whole genome based, transcriptional profiles of *Mycobacterium tuberculosis* subject to 7 days of hypoxia (enduring hypoxic response) to guide antigen discovery, using bioinformatic analysis and empirical testing.

Specific experiments directed to address the above aim are described in Chapters 5 and 6. The hypothesis include:

1. Molecules highly induced when *Mycobacterium tuberculosis* was subject to hypoxia *in vitro*, may be expressed in vivo and thereby could be targets of immune response
2. These molecules might provide further understanding in the biology of latent infection and may be preferentially recognized by the T cells during latent tuberculosis
3. Genomic information from a pathogen might aid in antigen prediction
4. Species specific genes of *Mycobacterium tuberculosis* are potential diagnostic targets.

However, a preliminary comparison of various techniques and the performance of synthetic overlapping peptides in T cell assays are presented in Chapter 3.

In addition, the T cell epitope mapping of previously described RD11-encoded Rv2654c and dosR encoded Rv1733c are described in Chapter 4.

Chapter 2: Materials and Methods

The studies presented in this thesis, examine the *in vitro* reactivity of human T cells to antigens of *Mycobacterium tuberculosis* and bioinformatic analysis of hypoxia inducible MTB antigens. A detailed description of patients enrolled, as well as the immunological, microbiological and bioinformatic methodologies employed in these studies is described below.

Ethical Approval

The University of Cape Town research ethics committee approved these study (REC336/2004; REC 296/2007; REC 245/2009).

Human Subjects

Persons with active or latent tuberculosis were recruited at the Ubuntu clinic at Khayelitsha site B, South Africa. All were of Xhosa ethnicity (Chapter 4-7). In addition, 44 healthy volunteers (laboratory and clinical staff) of mixed ethnicity, were enrolled in for the studies described in Chapters 3 and 4.

Active tuberculosis (ATB) was defined by smear positivity for and/or culture of *M. tuberculosis* from one or more sputum specimens. Latent tuberculosis (LTBI) was defined by transverse TST reactivity of > 15 mm in response to 2 TU PPD (RT23) at 48- 72 hours or an interferon-gamma (IFN-γ) Enzyme linked immunospot (ELISpot) response to ESAT-6 or CFP-10 of > 20 spot forming cells (SFC)/10^6 PBMC in the absence of clinical symptoms or radiographic abnormality and with a negative sputum smear and culture for *M. tuberculosis*.

ATB patients were sampled prior to commencing anti-tubercular chemotherapy. Known immunosuppression for other reasons, age < 18 years and pregnancy formed other exclusion criteria. All persons underwent voluntary counseling and testing for HIV-1 infection and positivity was an exclusion criterion, with the exception described below.

11 HIV-1 infected otherwise healthy persons, who were asymptomatic for any other infection, and naïve to anti retroviral therapy, were included in the study presented in Chapter 3. 19 HIV-1 infected persons who were starting combined anti-retroviral therapy (cART) and followed up for 36 weeks were included in the study presented in Chapter 6.

Persons with ATB and / or HIV-1 infection were treated according to South African national guidelines. The baseline characteristics of subjects enrolled are described in the respective Chapters.

Clinical Specimen

Peripheral whole blood samples were used for all experiments presented in this thesis as described below.

Heparinised peripheral whole blood

30 ml of venous blood was obtained in tubes containing sodium heparin, from all patients enrolled in the studies presented in Chapters 3-6. They were transported to the laboratory at room temperature and were processed in less than 4 hours after venepuncture.

QuantiFERON-TB Gold In Tube

Commercially available, QuantiFERON-TB Gold In Tube (QFT-GIT) tests were performed as per manufacturer's instructions, in studies described in Chapters 3 and 4. One ml of whole blood is collected by venepuncture directly into each of the QuantiFERON®-TB Gold IT blood collection tubes [Nil Control (Grey cap), TB Antigen (Red cap) and Mitogen Control (Purple cap)] and mixed thoroughly (as antigens have been dried onto the inner wall of the blood collection tubes).

In the peptide boosting experiments described in Chapter 4 the following modification was included: in addition to the three QFT-GIT, blood was also collected in an additional TB antigen and Nil tube, to which 5μg/ml of Rv2654c peptide covering region 51-65 was added. The tubes were transferred to a 37°C incubator within 4 hours of collection and incubated for 16-20 hours. After incubation of the tubes at 37°C, the plasma was harvested by centrifuging tubes for 15 minutes at 2000 to 3000 RCF (g). The gel plug separates the cells from the plasma. The plasma was collected from over the gel plug and stored at -80°C for batched measurement by QFT ELISA.

QuantiFERON TB Gold ELISA

The QFT Gold ELISA was performed for the study presented in Chapter 3 and 4, according to manufactures instruction. All plasma samples and reagents, except for conjugate 100X concentrate, were brought to room temperature (22°C ±5°C) before use (at least 60 minutes for equilibration). The freeze dried Kit Standard was reconstituted with deionised or distilled water to a concentration of 8.0 IU/ml and mixed gently to minimize frothing and to ensure complete solubilisation. Using the reconstituted kit standard, a 1 in 4 dilution series of IFN-γ in green diluent (GD) were produced [S1 (standard 1) contains 4 IU/ml, S2 (standard 2) contains 1 IU/ml, S3

(standard 3) contains 0.25 IU/ml, and S4 (standard 4) contains 0 IU/ml (GD alone)] and included in triplicates. The freeze-dried conjugate 100X was reconstituted with 0.3ml of deionised or distilled water. The working strength of the conjugate was prepared according to the manufactures instruction (60 µl of conjugate in 6 mls of GD to be used for one plate (12 strips)). 50 µl of working strength conjugate was added to all the required wells. 50µl of test plasma (Nil, TB antigen and mitogen stimulated) and standards (S1-4) were added to respective wells. The plate was mixed thoroughly for 1 minute in a microplate shaker and incubated at room temperature for 2 hours. After incubation, the plates were washed with 400 µl of working strength wash buffer, and 100 µl of enzyme substrate solution was added to each well. Following 30 minutes of incubation, 50 µl of enzyme stopping solution was added to stop the reaction and the optical density (OD) of each well was measured within 5 minutes using the Bio-Rad microplate reader, fitted with a 450nm and with a 655nm reference filters. The OD values are used to calculate results using QuantiFERON TB Gold analysis software v2.50. The cut-of for a positive result is defined by >0.35 IU/ml of IFN-γ measured in TB antigen minus nil and as negative when <0.35IU/ml. The results for mitogen minus Nil <0.5IU/ml or Nil >8IU/ml of IFN-γ were considered indeterminate.

Whole blood culture

A diluted whole blood culture of 7 day incubation (Beveridge et al., 2008; Weir et al., 1994) was performed for the study presented in Chapter 3. Whole blood was diluted 1:10 in RPMI 1640 medium containing 1% L-Glutamine. A 48 well (flat bottom) plate was set up with the antigens and controls with a final volume of 1000µl. The antigens were reconstituted in the desired concentration with RPMI 1640 Medium containing 1% L-Glutamine and added as a final volume of 100µl per well. PHA was

included as a mitogen / positive control and RPMI 1640 medium with 1% L-Glutamine as a negative control. 900 µl of diluted blood was added to the antigen plate and sealed with micro-pore tape, to avoid evaporation during incubation at 37°C with 5% CO_2. After incubation for 7 days, the supernatants were harvested and stored at -20°C until used for IFN- γ ELISA.

Peripheral Blood Mononuclear Cell (PBMC) Isolation

Whole blood was diluted 1:1 with 1 X Dulbecco's Phosphate Buffered Saline (PBS) (Sigma®, UK) and slowly pipetted over a 10 ml layer of Ficoll-Paque PLUS (GE Healthcare Bio-Sciences AB) in a 50ml Falcon universal tube (Sterilin). The tube was centrifuged at 1820 rpm for 20 minutes with the brakes off. The whole blood separated into layers and the buffy layer was removed using a plastic pasture pipette. The buffy layer consisting of the PBMC was washed in RPMI-1640 Medium (Sigma®) without additives and centrifuged at 600 RCF for 10 minutes with the brakes on. The RPMI was poured off leaving a pellet of PBMC behind. This pellet was re-suspended in 10ml RPMI with 10% heat inactivated Fetal-Calf Serum (HI-FCS) (RPMI/10%FCS) (Sigma), after which the number of PBMC was counted. A 1:1 dilution of PBMC in RPMI/10% FCS was added to a 0.4% trypan blue solution (Sigma®) and counted in a disposable, 10 chamber plastic counting slide. PBMC were frozen in a 1:1 dilution of RPMI / 10% FCS and cell freeze medium (20% DMSO in FCS) in a stepwise manor using MrFrosty at -80ºC freezer overnight, then transferred to liquid nitrogen, and stored in vapour phase until used for the experiments in batches.

In house IFN-γ release assay

In house ELISA were performed for the experiments described in chapter 3 as per the approved standard operating procedure of the Laboratory (Connell et al., 2007; Rangaka et al., 2007a). Plates (Costar 3590, 96 well EIA/RIA plate, high binding) were coated with captured antibody (Mouse anti-human IFN-γ monoclonal capture antibody (BD #551221) overnight at 4ºC. The plates were washed and blocked for 2 hours at room temperature, and washed. 100 µl of samples and standards (recombinant human IFN-γ (BD #554617) were added to the respective wells and incubated overnight at 4ºC. Following incubation, the plates were washed and 100µl of biotinylated secondary antibody (Biotinylated mouse anti-human IFN-γ detection antibody (BD#554550) added and incubated for 45 min at room temperature, washed and 100 µl of avidin peroxidase (sigma A3151) added and incubated at room temperature for 30 min. Following incubation, the plates were washed and 100µl of OPD solution (SigmaFast OPD tablet set) was added and incubated in dark for 5 min or until color develops. The reaction was stopped by 50µl of 2M H2SO4 (Sigma 32,0501). The plates are read at 490nm with BioRad plate reader. The detection range of the ELISA was 0-3000 pg/ml of IFN-γ.

Experimental procedures using PBMC

Thawing and resting of PBMC

Cryopreserved PBMC were thawed and rested overnight for all studies described in chapter 3-6. The cryovial containing the desired PBMC were transferred from liquid nitrogen to a 37ºC incubator (with the cap slightly loose to allow the nitrogen to escape during thawing). When the PBMC almost thawed (a small bit of ice remains), the cell suspension was transferred with a 1 ml micropipette to warm RPMI

containing 10% FCS (37ºC) in a drop wise fashion with constant gently mix (1 ml cell suspension to 9 ml RPMI in a 50 ml Falcon centrifuge tube) and centrifuged at 1680 rpm for ten minutes. The supernatants are discarded and the pellet are suspended in 10ml of RPMI containing 10% HI-FCS. Viable cell numbers are determined using the cell counting chamber with trypan blue stain, which differentiates dead and viable cells, (dead cell nucleus takes up the stain). The tubes containing the cell suspension are washed by centrifugation at 1680 rpm for 10 minutes and pellets are reconstituted at 10^6 per ml with RPMI containing 10% FCS. The cell suspension is rested overnight at 37ºC with 5% CO_2, with the lid slightly loose. After overnight resting (12-16 hrs), tubes are centrifuged at 1680 rpm for ten minutes and cells counted. The pellets were re-suspended at the required concentration for ELISpot, cell culture or FACS experiments.

ELISpot assay

The interferon-gamma ELISpot assay was performed in studies described in chapter 3-6 (Wilkinson et al., 2009). Ninety-six well pre-coated ELISpot plates, mAb 1-D1K (Pre-coated One-step, Mabtech; 3420-2ATP-10) were washed 4 times with sterile PBS, blocked with RPMI with 10% FCS for ≥30 min at room temperature. The blocking medium was removed and the PBMC were set up with respective antigenic stimuli. Control stimuli for ELISpot included anti-CD3 mAb CD3-2 at 100 ng/ml final concentration and unstimulated wells. After incubation for 16-18 hours at 37ºC with 5% CO_2, the cell free supernatants were collected and stored at -20ºC until used. The plates were washed 5 times with PBS, and 100µl of secondary antibody, mAb 7-B6-1-ALP conjugate at 0.5mg/ml final concentration in PBS containing 0.5% FCS was added. After 2 hours of incubation at room temperature, 100µl of filtered ready to

use substrate solution (BCIP/NBT-plus) was added and developed until spots emerged, washed with tap water and allowed to dry.

IL-2 ELISpot was performed for studies described in chapter 5 and 6. 96- well polyvinylidene difluoride membrane based plates, type ELIIP (MAIPSWU10; Milipore Corp), were activated by a brief treatment for < 2 minutes with 70% ethanol, coated overnight at 4ºC with 15µg/ml of mAb IL2-I (Mabtech; 3440-2AW-Plus). The plates were then washed and blocked with R10 for ≥ 30 min. The blocking medium was removed and the PBMC were set up with respective antigenic stimuli. Control stimuli for ELISpot included anti-CD3 mAb CD3-2 at 100ng/ml final concentration and unstimulated wells. After 16-hours of incubation at 37ºC in 5% CO_2, the plates were washed, 100µl of detection antibody (IL-2-II-biotin) at 1mg/ml in PBS containing 0.5% FCS added and incubated at room temperature. After 2 hrs, 100µl of Streptavidin-ALP 1:1000 in PBS-0.5% FCS was added and incubated at room temperature. After 1 hour, 100µl of substrate solution (BCIP/NBT-plus) was added and developed until distinct spots emerged. Plates were washed with tap water and allowed to dry.

Spot forming cells were enumerated by immunospot counter (CTL, Cellular Technology Ltd) and confirmed by microscopy (X4). Results are quoted as cytokine spot forming cells (SFC)/10^6 PBMC. A positive score is defined as SFC score twice as that of the background or ± 2SD of the background response, below which is defined as a negative score. The assay is termed indeterminate when either the SFC response in the background is > 20 SFC/10^6 PBMC or SFC response of positive control is < 20 SFC/10^6 PBMC.

The cut-off for positive results for IFN-γ ELISpot were set to ≥15 SFC/10^6 PBMC (above the background) for the study described in chapter 4. A cut of ≥20 SFC/10^6 PBMC (above the background) was set for both IFN-γ and IL-2 ELISpot response in Chapter 5 and 6, however, the cut of was arbitrarily increased to ≥30 SFC/10^6 PBMC (above the background) to define immunodominance for novel antigens in the study described in Chapter 5.

The ELISpot (IFN-γ and IL-2) experiments described in Chapter 6, using live *M. tuberculosis* strains H37Rv (wild type), H37RvDRD2 (H37Rv RD2 knockout mutant) and H37RvDRD2::Rv1986 (H37Rv RD2 knockout mutant, complemented with Rv1986), were performed as described above. 200,000 PBMC/ well were co-cultured with 200,000 live bacteria / well for 16-18hrs (in duplicate wells). The counts from duplicate well were averaged and used for further analysis.

Cell Culture

In vitro cell cultures were performed for the studies presented in Chapter 3 and 6. 200,000 - 250,000 PBMC were cultured with required concentration for 1 day, 3days, 5 days and 7 days for experiments described in Chapter 3, at 37°C in 5% CO_2. Control stimuli for cell culture included PHA at 5µg/ml final concentration and unstimulated wells. After incubation for the respective time points, the plates were centrifuged at 2000 rpm for 5 minutes and cell free supernatant were collected and stored at -20°C until used.

Surface markers and Intracellular cytokine staining for flow cytometry

Intracellular cytokine assay was performed for the experiments presented in Chapter 3 and 5 (Abel et al., 2010; Hanekom et al., 2008).

For the experiments presented in Chapter 3, 200,000 PBMC were co-cultured with Purified Protein Derivative (PPD) (2µl; 1000U/ml final concentration) and controls including Staphylococcal Enterotoxin B (SEB) as positive control at 10µg/ml and unstimulated cells as negative control at 37°C with 5% CO_2. After 2 hrs, Brefeldin A at 5µg/ml (Sigma, St. Louis, MO) was added to capture the newly formed cytokines in the golgi apparatus. After 16 h incubation (in total), the cells were washed with PBS (1X), stained for the surface marker CD3 (FITC) and CD4 (PerCp) (3µl of surface Ab each) for 30 minutes at 4°C. The cells were then fixed and permeabalised with BD Cytofix/Cytoperm for 20 minutes at 4°C, after which the cells were stained with intra cellular stain for IFN-γ (APC) (10 µl of 1:10 of diluted antibody in perm wash) for 30 minutes at 4°C. After incubation, the cells were washed, and pellets re-suspended in 150µl perm wash buffer and a minimum of 20,000 cells was acquired on BD FACS Calibur. The data was analysed with Flow Jo (Tree Star).

For the experiments presented in Chapter 6, 1.5-2 x 10^6 PBMC were incubated with the two Rv1986 peptides (residues 61-80 and 161-180) at 10 µg/ml each (i.e., 20 µg/ml peptide in total) or a pool of 21 peptides from CFP-10 and ESAT-6 at 2µg/ml each (i.e, 42 µg/ml peptides in total) at 37°C. Control stimuli included SEB as positive control at 10 µg/ml and unstimulated cells as negative control. After 2 hrs, Brefeldin A at 5µg/ml (Sigma, St. Louis, MO) was added to capture the newly formed cytokines in the golgi apparatus. After 16 h incubation (in total), the cells were washed with PBS (1X). For 8 colour surface and intracellular staining the cells were first permeabilized, and fixed using Cytoperm/cytofix buffer (BD) for 20 min at 4°C, washed with BD Perm/wash and stained with antibody cocktail in BD perm/wash for 1 hr at 4°C. The antibodies used were as follows: CD3-Pacific Blue (1µl/tube), CD4 QDot605 (0.5µl/tube), CD8 Cy5.5PerCp (3µl/tube), IFN-γ Alexa700 (1µl/tube), IL-2

FITC (5µl/tube), TNF Cy7PE (5µl/tube), CD45RA- APC (3µl/tube), CD27-PE (3µl/tube), all of which were purchased from BD BioSciences.

10^6 cells were acquired on the LSR II flow cytometer (BD Bioscience). Cell doublets were excluded using forward scatter area *vs.* forward scatter height parameters. Unstained cells and single-stained mouse calibration beads were used to calculate compensations for every run. Data analysis was performed using FlowJo v 8.8.2 (Tree Star), Pestle v 1.6.1 (NIH) and Spice v 5.05013 (NIH). T cell phenotypes were defined based on the surface markers CD45RA and CD27: Central memory cells (TCM) as positive for CD27 and negative for CD45RA; effector memory (TEM) are negative for both CD27 and CD45RA and Terminally differentiated T cells (Tdiff) are negative for CD27 and positive for CD45RA. The results are expressed as the percentage of CD3+ CD4+ T cells.

Multiplex cytokine analysis

Multiplex cytokine analysis was performed as per manufacturer's instructions, for the experiment described in Chapter 6. Bioplex, mixed-to-order panel (premixed multiplex panel) from Bio-Rad was used for multiplex cytokine analysis. The assay was carried out according to the manufacturer's instructions as follows: the 96- well filter plate was pre- wet with 150µl of Bio-Rad assay buffer and the buffer removed by vacuum filtration. 50µl of multiplex bead working solution was added to the wells and the buffer removed. 100µl of Bioplex wash buffer was added to each well and washed twice and the buffer removed. 50µl of standard and sample was added to the respective wells, the plate was sealed and then covered by aluminium foil and placed over a microplate shaker. The speed of the shaker was increased to 1100 RPM for 30 sec and then reduced to 300 RPM for 30 min, incubation at room temperature. After

incubation, the plates were washed 3 times with Bioplex wash buffer. 25 µl of Bioplex detection antibody working solution was added, and incubated for 30 min as above on the microplate shaker at room temperature. The plates were washed 3 times with Bioplex wash buffer and 50 µl of streptavidin-PE was added, and incubated for 10 min, washed 3 times with Bioplex wash buffer. Beads were re-suspended with 125 µl of Bioplex assay buffer, mixed over the micro-plate shaker at room temperature at 1100 rpm for 30 sec and read on the Bioplex suspension array system.

Microbial culture conditions, hypoxia model and microbial RNA analysis

The results of experiments performed in the Seattle Biomedical Research Institute, Seattle, Washington, USA formed the basis for our studies presented in Chapter s 5-6. These techniques have previously been extensively described (Rustad et al., 2008). Briefly, exponential phase cultures grown in rolling culture to an OD600 of 0.3 were diluted to a starting OD of 0.1 with warm media. 500 mL of this starting culture was transferred to a constantly stirred one litre flask. Hypoxia was generated by introducing a constant flow of nitrogen with trace amounts of oxygen (0.2% O_2), leading to bacteriostasis. Samples were taken before hypoxia, at four hours, and after 1, 4, and 7 days of exposure to hypoxia. RNA was isolated from these samples using bead beating in the presence of Trizol, followed by chloroform extraction and precipitation of RNA. The RNA was further cleaned using an RNeasy kit (Qiagen). Approximately 3 µg of purified RNA was converted to cDNA using Superscript III (Invitrogen). Aminoallyl dUTP was included in the cDNA reaction, and subsequently conjugated to reactive Cy dye esters. The aerobically growing transcriptional profiles were directly compared to each subsequent hypoxic time point by co-hybridization on the same microarray slide. The microarray slides and protocols were provided by the

Pathogen Functional Resource Centre at the J. C. Ventner Institute as part of their NIAID contract N01-AI-15447. Slides were scanned with a GenePix 4000B purchased from Axon Technologies. Raw background subtracted intensities were normalized to *Sig*A to provide approximate measure of transcript abundance.

The lists of genes and their fold induction and normalised intensity measures at 4 hours, 24 hours, 96 hours and 168 hours were obtained collaboratively and analysed the using bio-informatic techniques with the results presented in Chapter 5.

Bioinformatics evaluation

Bioinformatics analysis formed the major part of the evaluation presented in Chapter 5 and 6. The results are discussed in detail in respective chapters.

TubercuList (http://genolist.pasteur.fr/TubercuList/ or http://tuberculist.epfl.ch/) and TB database (TBDB) (http://www.tbdb.org/) were used to interrogate Enduring Hypoxic Response genes; coding and protein sequence for each of the genes were obtained for further evaluation.

NCBI BLAST (http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi) was used to find regions of high similarity in biological sequence with the sequences in nucleotide and protein database. In addition, Local BLAST applications available at the TB database were also used for evaluation of protein sequences with the protein database (TB genome).

ProPred (Singh and Raghava, 2001) (http://www.imtech.res.in/raghava/propred/), a web-based server was used to predict MHC class II binding regions in an antigen sequence. The analysis was carried out with a threshold of 1% (high stringency) in all 51 HLA-DR alleles available in the server. Number of peptide sequences with ≥ 30%

of highest binding score any peptide could attain for that particular allele, was calculated for each allele per protein.

Pubmed (http://www.ncbi.nlm.nih.gov/pubmed/) was used for extensive literature search and to review all existing literature, if available, for the EHR genes. Information on their antigenicity, whether they are known to induce any immune response; whether the genes are essential (Sassetti et al., 2003) and the distribution of MHC class 2 allelic frequency in South African Population (du Toit et al., 1988) were obtained from literature.

Antigen Preparation

Studies described in this thesis used synthetic overlapping peptides (Chapter 3-6) and recombinant (chapter 3) proteins of *Mycobacterium tuberculosis*. The synthetic overlapping peptides used were 15-20 mers as the focus of the studies presented in this thesis were to study the CD4 response. The purity of the synthetic peptides purchased ranges between >90-95 %.

Tuberculin Purified Protein derivative (PPD) was obtained from Evans Vaccines, Liverpool, UK. 10,000U/ml stock in PBS were stored at -80°C and used at 1000 U/ml final concentration was used in experiments described in chapter 3.

Recombinant ESAT-6 and CFP-10, obtained collaboratively from Apronex, Czech Republic were used in experiments presented in Chapter 3. Recombinant proteins were reconstituted with PBS- 2% BSA and were used at a final concentration of 5 µg/ml (Table 5 and 6).

Synthetic peptides of MTB proteins were used in experiments described in Chapter 3-6.

Overlapping peptides covering, ESAT-6 and CFP-10 used in experiments described in Chapter 3 and 4 were 15-mers overlapping by 10 residues, obtained from Peptide Protein Research Ltd, Oxford, UK. Mixtures were prepared in DMSO and PBS and the final dilution of each peptide was 3 µg/ml (Table 5 and 6).

20-mers overlapping by 10 αα residues covering the entire length of Rv2654c and Rv1733c; and 15-mer truncated peptides for Rv2654c, were used in experiments presented in Chapter 4. These were synthesized by the Research Group of Peptide Chemistry, Hungarian Academy of Sciences, Eotvos Lorand University, Budapest, Hungary, within a joint, Hungary-South Africa research cooperation programme, funded by the National Research Foundation of South Africa. The peptides were dissolved in DMSO (0.1-0.3% final concentration) and PBS with 2% BSA and stored at -20°C until used (Table 9 and 10). Peptides were used at a final concentration of 3 µg/ml.

ESAT-6, CFP-10 and Acr-1 used in experiments presented in Chapter 5-6, obtained collaboratively from Veterinary Laboratories Agency, were 16-mers overlapping by 8 αα. The peptide mixtures, were prepared in DMSO (0.1% final concentration) and PBS, Stocks of 2 mg/ml were stored at -80C until used (Appendix A.1). Peptides mixtures were used at a final concentration of 10µg/ml.

All synthetic peptides used in experiments presented in Chapters 5-6 were 20-mer peptides overlapping by 10 residues and were purchased from Peptide Protein Research Ltd, Oxford, UK and from Pepscan Presto B.V, Netherlands. The peptide mixtures were prepared with DMSO (final concentration of 0.1-0.2%) and PBS and each peptide used at a final concentration of 10µg/ml (Appendix A.2).

Recombinant MTB strains

Recombinant strains described in the experiment presented in Chapter 6 were obtained collaboratively from Dr. Marcel Behr's lab at McGill University Health Centre, Montreal Canada. Briefly, the MTB H37RvΔRD2 strain (RD2 mutant) was prepared using homologous recombination and sucrose counter-selection as previously described (Sherman et al., 2001). This mutant was then electroporated with either the empty plasmid pMV306 or the same plasmid into which Rv1986 from H37Rv had been cloned. This gene was expressed under its native promoter. This resulted in the MTB H37RvΔRD2::pMV control (RD2 mutant) and MTB H37RvΔRD2::Rv1986 (RD2 mutant complemented with Rv1986) strains, which were grown in 7H9 + ADC + 0.05% Tween 80 + Kanamycin (25mg/ml) and glycerol stocks were prepared.

Colony forming units were determined. 200,000 bacilli were used at a final volume of 10µl and co-cultured with 200,000 PBMC.

Modifications include, treatment with Isoniazide treatment (Isonicotinic acid hydrazide, Sigma I-3377), at a final concentration of 1µg/ml and heat killing at 56°C for 30 minutes in a water bath.

Statistical analysis

The normality of data was assessed by the D'Agostino and Pearson omnibus test using Graphpad Prism 5.0 software (www.graphpad.com). Parametric continuous variables were assessed by student's paired and unpaired t-tests, and non-parametric by Wilcoxon matched pairs, Kruskal Wallis test with Dunn's post test correction or Mann Whitney U tests. Contingency analysis was by Fisher's exact test of probability: kappa statistics was used to test the agreement between two tests. Correlation was assessed by non-parametric spearman correlation coefficient.

Chapter 3: Comparison of various techniques and performance of peptides in *in vitro* blood cultures

Abstract

Studies on antigens of *Mycobacterium tuberculosis* are of wide interest to improve vaccine efficacy and to define a biomarker for diagnosis and monitoring of treatment. The methodologies used for these investigations differ greatly between laboratories and the discrepancies between results are inevitable. The following investigations are addressed in this chapter.

Firstly, IFN-γ response to well characterized MTB proteins was evaluated in HIV-1 infected and uninfected persons, by 3 different techniques: Overnight IFN-γ ELISpot and FACS analysis; and 7 day PBMC culture. Findings show that HIV-1 infected persons, have a lower ability to induce an immune response to MTB proteins, by all 3 techniques and suggests that the proportion of terminally differentiated activated T cells, effector memory and central memory differ between HIV-1 infected and HIV-1 uninfected. Poor correlation between the techniques suggests phenotypic heterogeneity between what is measured by specific readouts.

Secondly, discrepancies in IFN-γ response between short and prolonged *in vitro* culture systems and the performance of synthetic peptide mixtures and recombinant proteins were studied in the context of whole blood cultures. The results suggest that recombinant proteins perform well in prolonged *in vitro* whole blood cultures, whilst the peptide pools perform poorly. The activity of peptides and recombinant proteins do not differ in short term or long-term PBMC culture assays. In addition, results provide evidence that the use of DMSO (up to 0.3% final concentration) to dissolve synthetic peptides do not affect the viability or the activity of the antigens. Further findings suggest that peptides are degraded during the course of 7 days of incubation in whole blood culture system.

Introduction

In vitro studies on antigens of *M. tuberculosis* are a research priority, to discover new immunododminant targets to improve immuno diagnosis and vaccines. Various laboratories use different techniques and read out methods to compare the performance of new candidates.

The measurement of *Mycobacterium tuberculosis* antigen specific T cell Interferon-gamma (IFN-γ) production has received great attention and is used widely for determination of immune recognition in antigen discovery. First generation IFN-γ release assays used purified protein derivate of MTB, which has also been used in Mantoux test (Tuberculin Skin Test) for over one hundred years. The specificity of these assays were highly compromised, due to the cross reactivity of immune response to homologous proteins in tuberculous, non tuberculous and environmental mycobacteria.

New generation IFN-γ release assays were developed which use MTB specific ESAT-6 and CFP-10, a co-expressed immunodominant antigen encoded by the region of difference (RD)-1. These antigens were first evaluated in a 6-day lymphocyte stimulation test (LST) and found to be highly specific for the diagnosis of tuberculosis (Arend et al., 2000b; Ravn et al., 1999). IFN-γ release assays were subsequently developed that differed from the classical LST with respect to the following described in Table 1.

Studies have shown that T cell responses to MTB specific antigens: recombinant ESAT-6 and CFP-10 correlated with the respective mixture of synthetic overlapping peptides when PBMC of MTB specific cell lines were used (Arend et al., 2000b), IFN-γ results from short term (24- 72 hours) PBMC and whole blood cultures found to correlate (Scholvinck et al., 2004). However, studies also suggest a great degree of discrepancy in the measured

immune response between short term and long-term culture conditions (Beveridge et al., 2008; Leyten et al., 2007).

Table 1 Various formats of techniques used in immunodiagnosis and antigen screening

Antigen	**Form of antigen**	**Culture Media**	**Length of Incubation**	**Read out Methods**
-PPD -CFP-10 -ESAT-6 -TB7.7 (p4) -Novel candidates	-Synthetic overlapping peptides -Recombinant proteins	-Whole blood -Diluted whole blood -PBMC culture	- 24 hours - 48 hours - 72 hours - 120 hours -168 hours	-ELISA -ELISpot -FACS

In this chapter, the following investigations, were performed within the context of the techniques involved in the antigen screening and are described in two groups of experiments:

In the first group of experiments,

1. Evaluation and comparison of IFN-γ response to MTB proteins measured by 3 techniques: Overnight (16-18 hour) T cell ELISpot and FACS analysis, and 7 day PBMC culture and ELISA. This was evaluated in HIV-1 infected and HIV uninfected person with latent tuberculosis.
2. The ability to induce an IFN-γ response, between HIV-1 infected and HIV uninfected were compared.
3. Correlation of IFN-γ response detected by 3 techniques was assessed.

In the second group of experiments:

4. Evaluation and comparison of whole blood culture assays in HIV uninfected adults.

5. Comparison of short term commercial QuantiFERON TB Gold In Tube (QFT – GIT) assay with prolonged (7 day) in house *in vitro* culture assays using ESAT-6 and CFP-10 to induce an IFN-γ response.
6. Performance of MTB antigens in the form of synthetic overlapping peptide mixtures and recombinant proteins in long term whole blood culture assay system.

Due to the exploratory nature of the experiments presented in this chapter, the results and discussions are presented together.

Results and Discussion

Comparison of PPD specific IFN-γ production in HIV-1 infected and HIV uninfected persons using various techniques

A study was designed to compare IFN-γ production in HIV uninfected and HIV-1 infected persons using 3 different techniques: the overnight ELISPOT assay, measurement of CD4+IFN-γ+ T cells by intracellular flow cytometry (IC-FACS) and the 7 day PBMC culture method, using ELISA to quantitative secreted IFN-γ in the culture supernatant (7d-IFNGRA).

Twenty-four healthy patients were recruited from Khayelitsha township near Cape Town, South Africa. Thirteen were HIV uninfected (median age 25±6 years), and eleven people were HIV-1 infected (median age 30±4 years, median CD4 count 391±243) (Table 3). 23/24 patients ascertained to be latently *M. tuberculosis* infected based on a positive response to either ESAT-6 or CFP-10, or a positive tuberculin skin test reaction (TST) (defined by >5mm for HIV-1 infected and >15mm for HIV uninfected), were included in the final analysis.

Table 2 Demographic detail of persons included in Chapter 3 part 1

	HIV Uninfected	**HIV-1 Infected**
Number	12	11
Median Age (IQR)	25 (22-33)	30 (28-33)
Sex: M/F	8M/5F	4M/7F
Median TST (IQR)	18 mm (9-20)	16 mm (0-22)
Median CD4 counts	NA	391 (180-475)

IFN-γ analysis by ELISpot

The IFN-γ producing capacity of T cells in HIV uninfected and HIV-1 infected patients was first evaluated using ELISpot. The largest interferon gamma ELISpot response in the HIV uninfected group was to CFP-10 (372 SFC/10^6 PBMC, IQR 7-1151) and this response was significantly greater than that in the HIV-1 infected group (5 SFC/10^6 PBMC, IQR 0-24; p=0.03). In the HIV-1 infected group the highest response was to PPD (36 SFC / 10^6 PBMC, IQR 4-104). In addition to CFP-10, the IFN-γ responses to ESAT-6 and PPD were higher in HIV uninfected when compared to HIV-1 infected group (54 SFC/10^6 PBMC, IQR 30-90 *vs.* 20 SFC/10^6 PBMC, IQR 0-75; p=0.13 and 88 SFC / 10^6 PBMC, IQR 0-262 *vs.* 36 SFC / 10^6 PBMC, 4-104; p=0.29) respectively (Figure 7 A).

The ELISpot response was also analysed by the sum of responses to ESAT-6 and CFP-10, which was significantly higher in HIV uninfected compared to HIV-1 infected group (398 SFC / 10^6 PBMC, IQR 57-1245 *vs.* 24 SFC / 10^6 PBMC, IQR 8-90; p=0.006 respectively) (Figure 7 A).

When analysed by the frequency of response (defined as >20 SFC/10^6 PBMC of IFN-γ), there was no significant difference observed in the IFN-γ response between the two groups. ESAT-6 (83%) was highly recognized by HIV uninfected persons, while ESAT-6 and PPD (63%) were highly recognized by HIV-1 infected persons. However, the most frequently recognized response was to sum of ESAT-6 and CFP-10, by both groups (Figure 7 B).

ELISpot enumerates the number of cytokine secreting cells and the responses were predominantly attributable to the effector cells responding to the stimuli (Wilkinson et al., 2005b). These findings suggest that the HIV-1 infected persons have effector T cells, that have a reduced ability to induce an immune response, however, the percentage scored positive, do not differ between HIV-1 infected and HIV uninfected persons. However, previous report from the same population suggest that ELISpot results are unaffected in HIV-1 infection (Oni et al., 2010; Rangaka et al., 2007b).

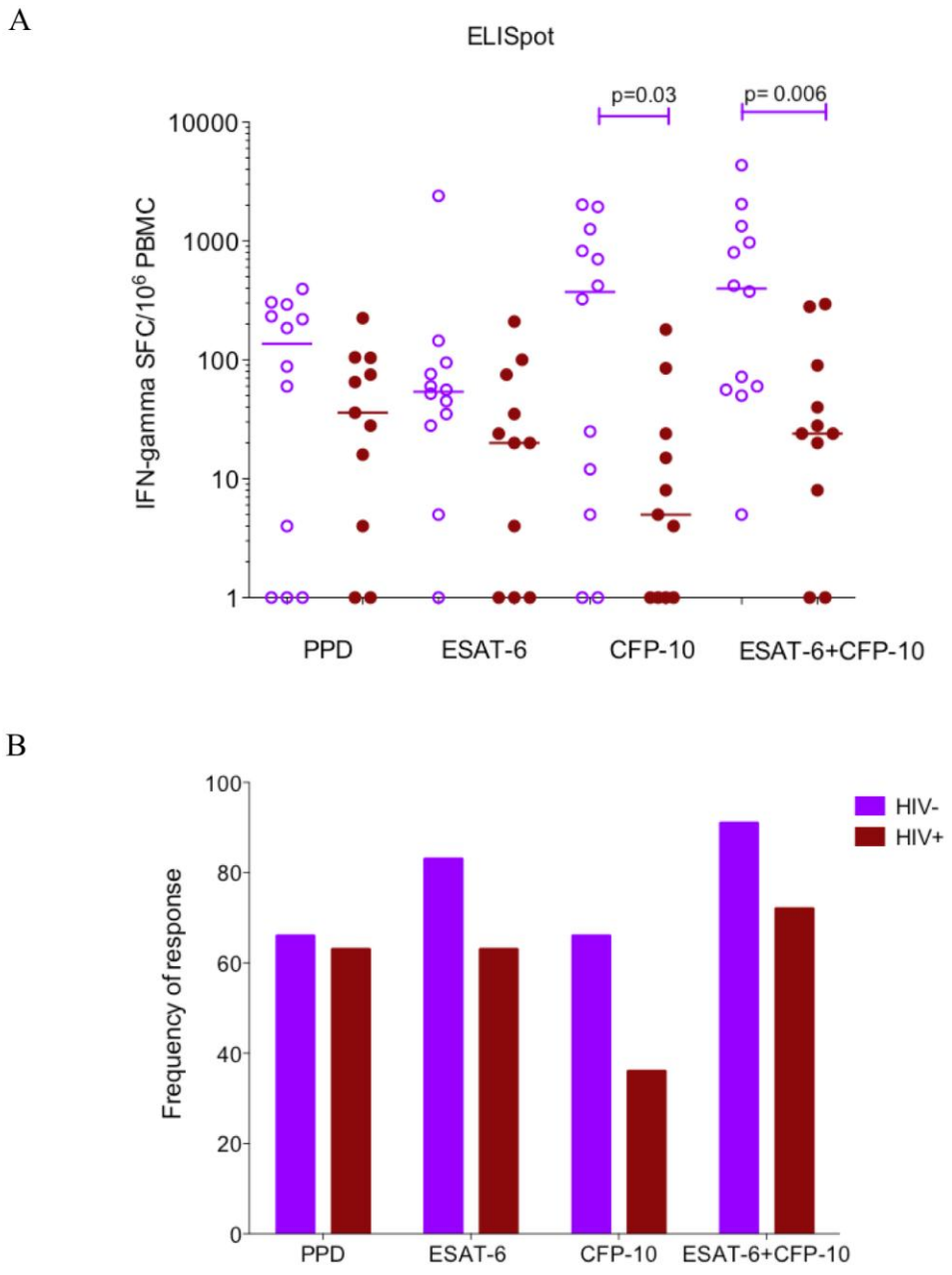


Figure 7 IFN-γ ELISpot response to antigens in HIV uninfected and HIV-1 infected persons

Panel A shows the spot forming cells (SFC) in patients with HIV uninfected (○) and HIV-1 infected (●). The strongest response in HIV uninfected group was to CFP-10. The strongest response in HIV-1 infected was to PPD. HIV uninfected persons had higher IFN-γ response than HIV-1 infected persons, and was significantly higher for CFP-10 and sum of CFP-10 and ESAT-6.

Panel B shows the frequency of response, defined by >20 SFC/10^6 PBMC. Higher frequency of response was observed for sum of ESAT-6 and CFP-10 in both groups. There was no significant difference observed between the two groups.

As there were only limited cells available, all antigen stimulation for all 3 techniques was not possible. PPD induced the largest IFN-γ response amongst the HIV-1 infected persons. Hence, PPD specific IFN-γ response was compared in FACS and 7 day in house PBMC culture - ELISA.

Intracellular FACS analysis

IFN-γ production by CD4+ T cells in response to PPD stimulation by intra cellular-FACS were analysed in both HIV-1 infected and HIV uninfected persons. PBMC were stimulated overnight (16 hrs) with BFA added 2 hours after stimulation. After incubation at 37ºC with 5% CO_2, the cells were washed, fixed, permeabilized and stained for surface markers (CD3-FITC and CD4-PerCp) and intracellular stain for IFN-γ (APC). Stained cells were acquired on FACS calibur (minimum of 20,000 cells/test) and results were analysed with Flow Jo (Tree Star). The gating strategies are shown in Figure 8 .

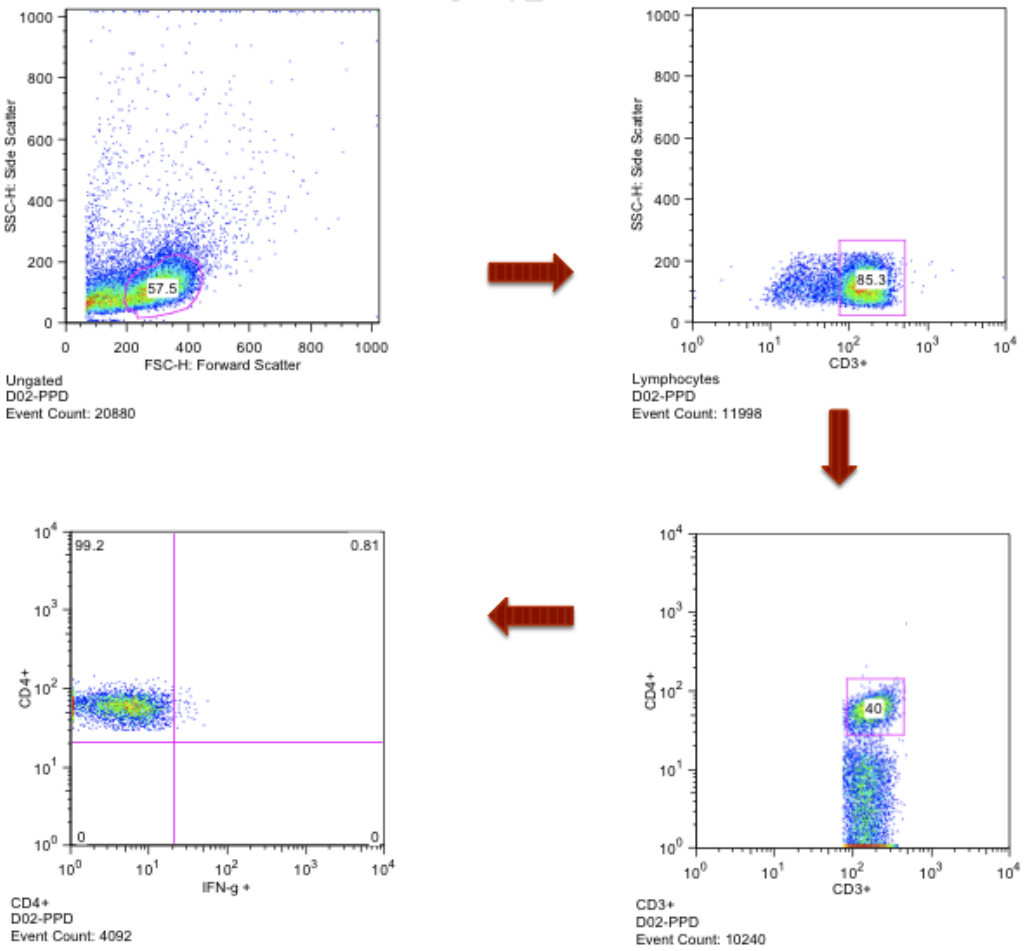


Figure 8 Gating strategies

Results indicated a higher percentage of lymphocytes and PPD specific IFN-γ producing CD4+ T cells in HIV uninfected group when compared to the HIV-1 infected group: median percentage of lymphocytes 75% *vs.* 46 %; p=0.004 and median CD4+ IFN-γ+ T cells 0.73% *vs.* 0.11%; p=0.0002 respectively. Similar trend was observed when analyzed by the number of cells per million: a median $1189/10^6$ (IQR 833-1679) CD4+IFN-γ+ T cells in the HIV uninfected group, compared to $400/10^6$ (IQR 263-935) CD4+IFN-γ+ T cells in the HIV-1 infected group (p=0.05, Mann-Whitney U) (Figure 9 A-C).

IC- FACS measures the cytokine expressing T cells. The CD4 phenotypes that are responsible for these responses can be both memory and effector T cells. These findings suggest that HIV-1 infected persons might have lower proportions of functional memory T cells.

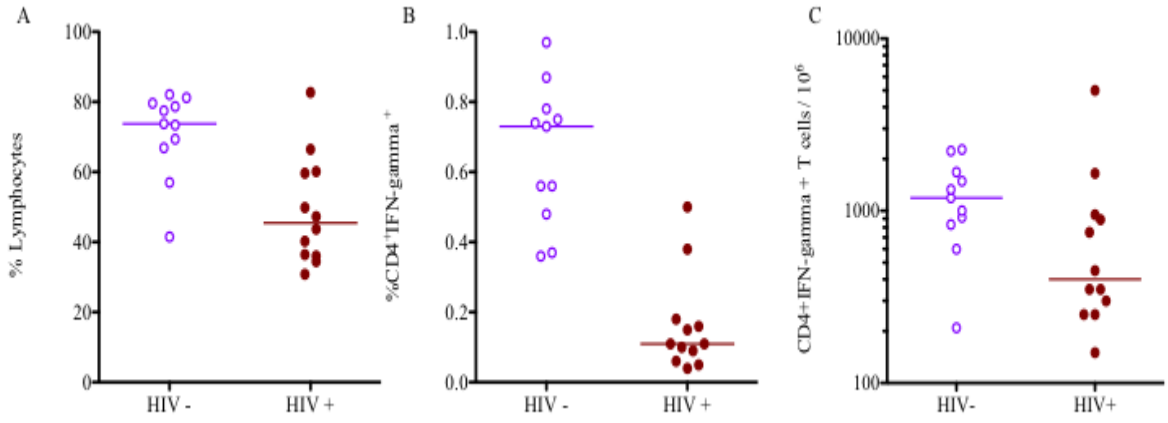


Figure 9 FACS Analysis

Panel A shows percentage of lymphocytes of the total PBMC acquired. HIV uninfected in blue and HIV-1 infected in red. Total lymphocytes were defined on the basis of side scatter that defines the granularity or the complexity of the cells and forward scatter that defines approximate size of cells.

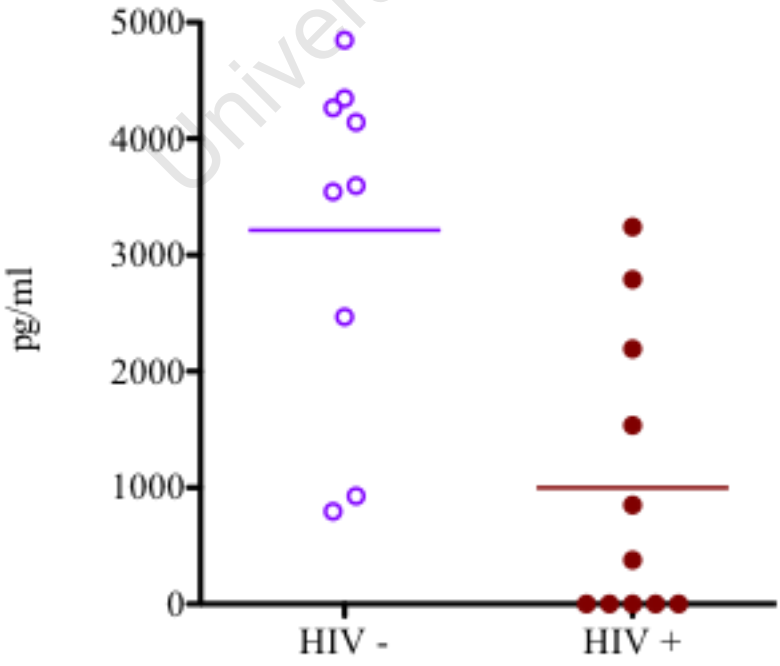
Panel B shows the percentage of CD4+T cell that produce IFN-γ to PPD stimulation.

Panel C shows the number of IFN-γ producing CD4+T cells in HIV-1 uninfected and infected groups.

7 day PBMC culture and ELISA

One way of evaluating the functionality of memory T cells is to assess their proliferative capacity. An indirect way to assess this is the prolonged (7 day) cell culture of PBMC combined with evaluation of the IFN-γ response of the proliferating memory cells. In this assay, the stimulations included PPD, the control antigen: PHA as the positive control and media as the negative control. The cell free supernatants were collected at the end of 7 days of incubation and IFN-γ was measured using in house ELISA, as per the standard laboratory protocol.

PPD specific IFN-γ levels were higher in HIV uninfected (mean 3214 SEM±497pg/ml IFN-γ) compared to the HIV-1 infected group (999±373 pg/ml IFN-γ; p=0.002) (Figure 10). The results indicate that HIV-1 infected memory T cells have lesser IFN-γ producing capacity and probably lower proliferating capacity, supporting the findings observed by FACS.



Shows the IFN-γ response in pg/ml in HIV uninfected and infected groups. Bar indicates mean responses.

Figure 10 IFN-γ response by 7 day PBMC culture

Correlation of responses between ELISpot, FACS and ELISA

The correlation of PPD specific IFN-γ response measured by the three techniques was analysed. In the HIV uninfected group, no correlation was observed amongst the three techniques. The correlation co-efficient (rho) for ELISpot *vs.* ELISA was 0.24 (p=0.24); FACS *vs.* ELISpot - 0.228 (p=0.5) and ELISA *vs.* FACS was -0.010 (p=0.97).

In the HIV-1 infected group also, no correlation was observed amongst the three techniques: ELISpot *vs.* ELISA -0.01 (p=0.97), ELISpot *vs.* FACS -0.5 (p=0.12) and FACS *vs.* ELISA -0.38 (p=0.24). These finding suggest that there was poor correlation between the techniques irrespective of the immune status of the patients and might also reflect on the heterogeneity of the patient population tested.

In conclusion, these investigations provide further insight in addition to the findings available in the literature. There is poor correlation between the IFN-γ read out by various techniques in both HIV uninfected and HIV-1 infected persons, suggests phenotypic heterogeneity between what is measured by specific readouts. In addition, HIV-1 infected persons have reduced ability, to induce an immune response to MTB protein as measured by all three methods.

Evaluation and comparison of whole blood culture assays in HIV uninfected adults

Comparison of short and prolonged in vitro Whole blood culture

This study was designed to compare the commercially available overnight QuantiFERON TB Gold in Tube (QFT-GIT) with the widely used 7 day diluted whole blood assay (Beveridge et al., 2008; Hanekom et al., 2008). In this evaluation, 29 healthy laboratory and clinical staff (volunteers) who reported some exposure to tuberculosis during their lives were included. No TST or HIV-1 testing was performed. None of the volunteers had symptoms of active TB or of being HIV-1 seropositive. The demographic details of those included in this evaluation are described in Table 3. Of the volunteers, 10 were male and 19 were female; median age was 27 years and 25 were BCG vaccinated. 17 of the 29 persons reported work related contact with tuberculosis patients (clinicians).

Table 3 Demographic details of persons included in the study described in Chapter 3 part 2

	Latent TB
Number	29
Sex	10M/19F
Median Age	27 years
BCG vaccinated	25 vaccinated 1 not vaccinated 3 Unknown

Blood sample was collected by venepuncture in Sodium Heparin tube and 3 QFT tubes [(1ml of blood per tube (Nil, Antigen and Mitogen tubes)]. QFT tubes were incubated overnight and the plasma was harvested and stored at -80°C until assayed. 1 ml of 1:10 diluted whole blood (in RPMI) was used to set up the in house 7day culture assay (Beveridge et al., 2008; Black et al., 2002; Weir et al., 1994). MTB specific antigens (ESAT-6 and CFP-10) were used as stimulants either as peptide pools (3µg/ml of each peptide) or recombinant proteins (5 µg/ml). After 7 days of Incubation at 37°C in the presence of 5% CO_2, the supernatants were harvested and an in-house IFN-γ ELISA was performed. This assay has a dynamic range of approximately 10-3000 pg/ml. The lowest sensitivity of the assay was 14 pg/ml and calculated values below this were reassigned 'zero'.

Of 29 QFT-GIT tests, 14 were positive (48%). 10/29 donors (34%) responded (> 14 pg/ml) to any of the two-peptide mixtures in the 7 day diluted whole blood assay. However, when recombinant antigens were used as stimulants, 22/29 (90%) donors responded to either recombinant ESAT-6 or CFP-10 (Figure 11).

Kappa statistics were calculated to compare agreement between short term QFT-GIT and 7day (peptide and recombinant stimulated) cultures. When QFT-G IT was compared to any response by peptides in 7 day culture, agreement was 59%; Kappa 0.16 (i.e. SLIGHT) and to response by recombinant antigen, agreement was 55%; Kappa 0.13 (i.e. SLIGHT).

Appendix Figure C.1 describes direct comparison of 7 day whole blood assay response to ESAT-6 in pg and QFT-GIT in IU/ml.

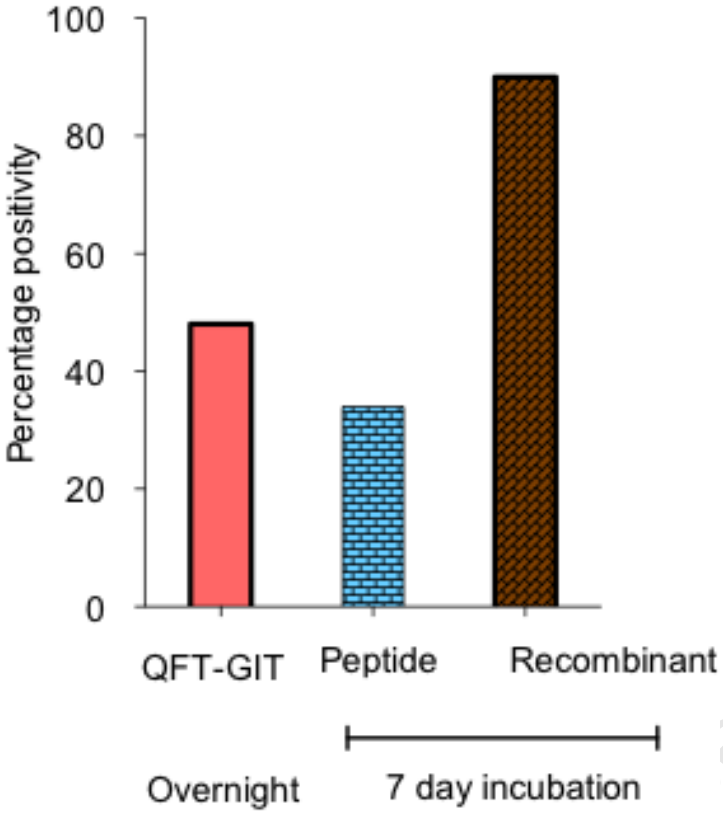


Figure 12 Percentage positivity for IFN-γ response

Shows frequency of response to IFN-γ, as defined by ≥0.035 IU/ml for QFT-GIT; ≥14pg/ml for peptide pool or recombinant antigen.

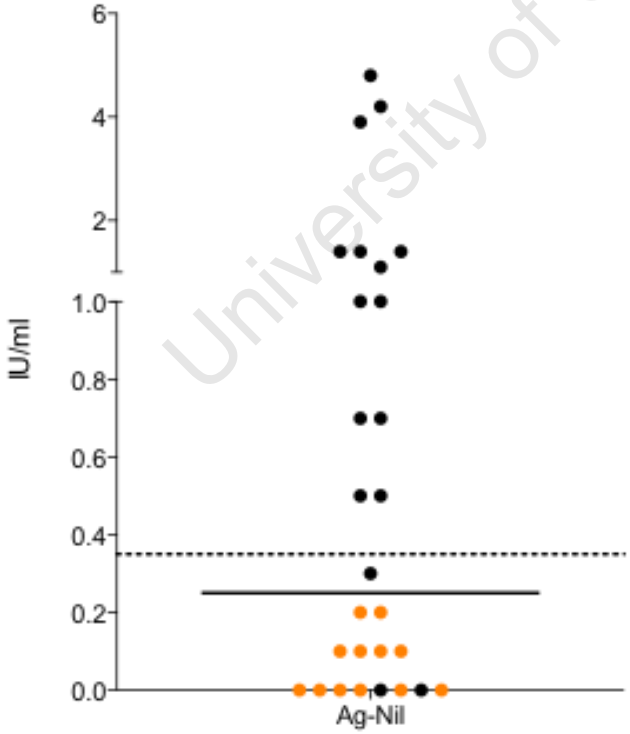


Figure 11 IFN-γ response measured by QFT-GIT

Shows the IFN-γ response measured by QFT-GIT and represented as IU/ml.

Subjects with a negative QFT –GIT and a postive response in ELISpot with recombinant protein are shown in orange.

Dotted line indicates the cut off for a positive response (0.35IU/ml). Solid line indicates the median response.

Table 4 Correlation between QFT -GIT and IFN-γ response in 7 day Whole blood culture

Parameter	**PP ESAT-6**	**PP CFP-10**	**Rec ESAT-6**	**Rec CFP-10**
n*	28	28	28	28
Spearman r	0.4491	0.1475	0.3306	0.5861
95% confidence interval	0.07974 to 0.7100	-0.2497 to 0.5022	-0.06012 to 0.6335	0.2617 to 0.7915
P value (two-tailed)	0.0165	0.4537	0.0857	0.001
* 1 person had QFT-GIT IFN-γ response >10 IU/ml and therefore excluded from this analysis (because of the arbitrary value from the QFT-GIT assay result).				

I further analyzed the qualitative responses (figure 12 and 14) by correlating the absolute values in pg/ml with the IU/ml in the QFT-GIT. Although there was only slight agreement between QFT-G IT and In house IFN-γ release assay, there was significant correlation between QFT-GIT with In house IFN-γ response to peptide pool of ESAT-6 (r=0.45; p=0.02) and the IFN-γ response to the recombinant CFP-10 (r=0.6; p=0.001). The spearman rho and associated p values are as follows (Table 4).

The performance of peptide mixtures are inferior to that of recombinant proteins in prolonged whole blood culture

A large difference in the IFN-γ response was observed between synthetic peptides and recombinant ESAT-6 and CFP-10 (Figure 14) in the 7 day whole blood culture assay. However, the peptide mixtures and recombinant proteins perform at a similar manner in overnight assays (as depicted in Figure 13) and long term cultures using PBMC as found in the literature (Arend et al., 2000b).

Figure 13 IFN-γ response to peptide pools and recombinant protein in overnight ELISpot

Shows IFN-γ SFC/10^6 PBMC measured by overnight ELISpot. "PP" indicates the peptide pools and R indicates recombinant proteins of ESAT-6 and CFP-10. When analysed by Wilckoxons matched pair T test, there was no significant difference found between the peptide pool and recombinant protein.

By contrast, the IFN-γ response to recombinant proteins: ESAT-6 and CFP-10 were significantly higher than to the corresponding peptide pools. (ESAT-6 peptide pool median of 0 pg/ml *vs.* recombinant 1878 pg/ml; p= <0.0001 and CFP-10 peptide pool 0 pg/ml *vs.* recombinant 305pg/ml) (Figure 14). 10/29 donors (34%) responded (> 14 pg/ml) to any of the two peptide mixtures. 22/29 (90%) donors responded > 100 pg/ml to either recombinant ESAT-6 or CFP-10 and this difference was significant (p≤0.0001 Fisher's test) (Figure 12).

Figure 14 IFN-γ response to peptide pool and recombinant ESAT-6 and CFP-10

Shows IFN-γ response measured by ELISA on a 7 day diluted whole blood culture supernatant. "PP" indicates peptide pool and "R" indicates the recombinant proteins. Recombinant proteins induced significantly higher response than the peptide pools in the 7 day diluted whole blood culture ($p<0.0001$ for both ESAT-6 and CFP-10). IFN-γ levels are shown as pg/ml. Lines indicate median response.

In order to further investigate, first, I re-assayed IFN-γ response to peptide pools in a high sensitivity ELISA (lowest standard that was measurable was 6 pg/ml). With the high sensitivity ELISA, the percentage positivity (response >6 pg/ml) to any peptide mixture increased to 66% (19/29). However, the median IFN-γ response to either ESAT-6 (peptide pool 4 pg/ml *vs.* recombinant 1878 pg/ml; p= <0.001) or CFP-10 (14 pg/ml *vs.* 304 pg/ml; p=0.001) was significantly lower than recombinant proteins (Figure 15).

Figure 15 IFN-γ responses to peptide pool and recombinant ESAT-6 and CFP-10 (high sensitivity ELISA)

Shows IFN-γ response measured by ELISA on a 7 day diluted whole blood culture supernatant, assayed on a high sensitivity ELISA. "PP" indicates peptide pool and "R" indicates the recombinant proteins. Recombinant proteins induced significantly higher response than the peptide pools in the 7 day diluted whole blood culture ($p<0.0001$ for both ESAT-6 and CFP-10). IFN-γ levels are shown as pg/ml. Lines indicate median response.

Considering the high purity of the recombinant proteins, and clear negative responses observed, the possibility of contamination in the recombinant protein to induce higher IFN-γ response was ruled out. In addition, there was no significant difference observed, in overnight ELISpot assay using PBMC from the same persons.

Since this observation seemed to be biologically very interesting, I further investigated the possible causes in a subset of donors.

DMSO in peptide mixture is not toxic

The peptide mixtures were prepared in 0.1% to 0.3% final concentration of DMSO and PBS/ 2% BSA, whereas the recombinant proteins were dissolved either in PBS or PBS / 2% BSA. In order to rule out the possible toxic effect of DMSO in the peptide mixture to the cells in the whole blood culture, I tested PBMC from 4 volunteers, stimulated with recombinant ESAT-6 and CFP-10 in parallel with and without DMSO in overnight ELISpot assay. 3% DMSO was added to recombinant proteins. Positive controls included were anti CD-3 (MabTech) and unstimulated well. The experiment was set in duplicate well and the average spot forming cell counts were analyzed.

The median SFC response to recombinant ESAT-6 (without DMSO) (67 SFC/10^6 PBMC, IQR 17-70) and recombinant ESAT-6 with DMSO were similar (61 SFC/10^6 PBMC, IQR 16-66; p=0.2 by Wilcoxon signed rank paired t test). Similar results were observed for recombinant CFP-10 as well (61 SFC/10^6 PBMC, IQR 13-298 *vs.* 60 SFC/10^6 PBMC, IQR 13-280 for CFP-10 with DMSO; p=0.5) (Figure 16). These finding confirmed that DMSO upto 0.3% final concentration used in the peptide mixture as not toxic to the culture system and ruled out the toxic effect of DMSO as the reason for the reduced level of interferon gamma observed with the peptides.

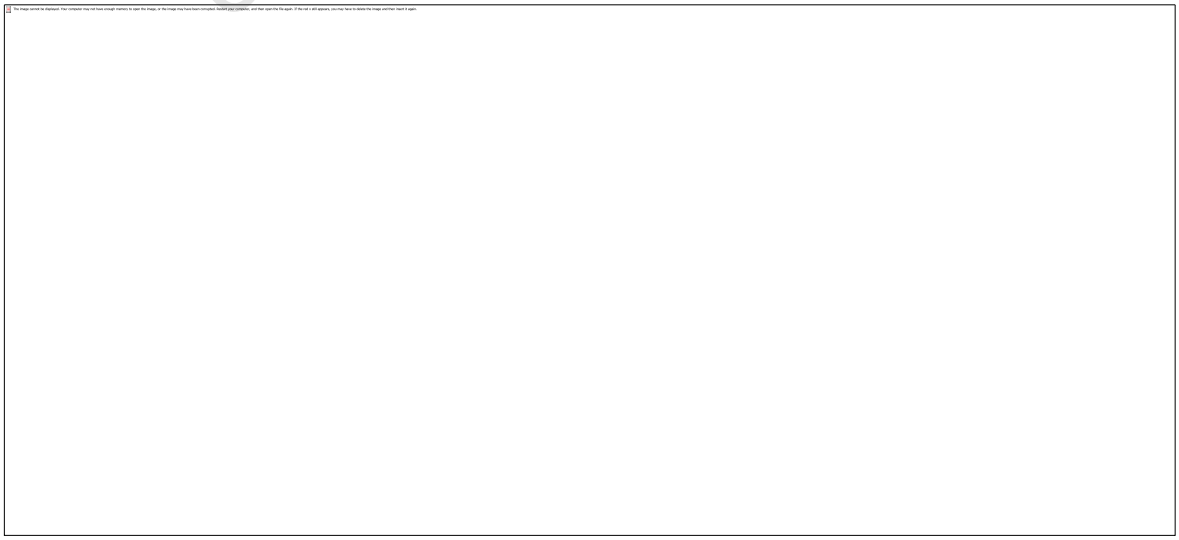


Figure 16 Effect of DMSO

Figure 16 Effect of DMSO

PBMC from 4 persons were tested for IFN-γ response to recombinant proteins: ESAT-6 and CFP-10, with or without DMSO. E and C refer to recombinant ESAT-6 or CFP-10 respectively; ED and CD refers to DMSO in addition to recombinant ESAT-6 and CFP-10 respectively. Bars indicate IFN-γSFC response per person. The IFN-γ levels were at similar levels.

Peptide and recombinant proteins perform similarly in prolonged PBMC cultures

In addition, I evaluated the PBMC cell culture with various time points of incubation: 1 day, 3 days, 5 days and 7 days, to study its performance. I hypothesized that there will be no / less effect of peptide degradation by protease activity, as in the whole blood culture.

250,000 PBMC were cultures with appropriate stimuli (peptide pool or recombinant protein) and controls in four sets. These were incubated at 37°C with 5%CO_2 for 1 day, 3 days, 5 days and 7 days. At the end of each incubation time point, the cell free supernatant was collected and stored at -20°C until ELISA was performed.

The results (shown in Figure 17) suggest, that both peptide and recombinant form of ESAT-6 and CFP-10 induce similar levels of IFN-γ production during the time course of the experiments. Therefore, these results are comparable in terms of response between peptide and recombinant protein to the overnight ELISpot (Figure 15), and suggest that peptides and recombinant proteins have similar activity in prolonged PBMC based cultures. This might also suggest that the PBMC probably have less proteases compared to whole blood cultures.

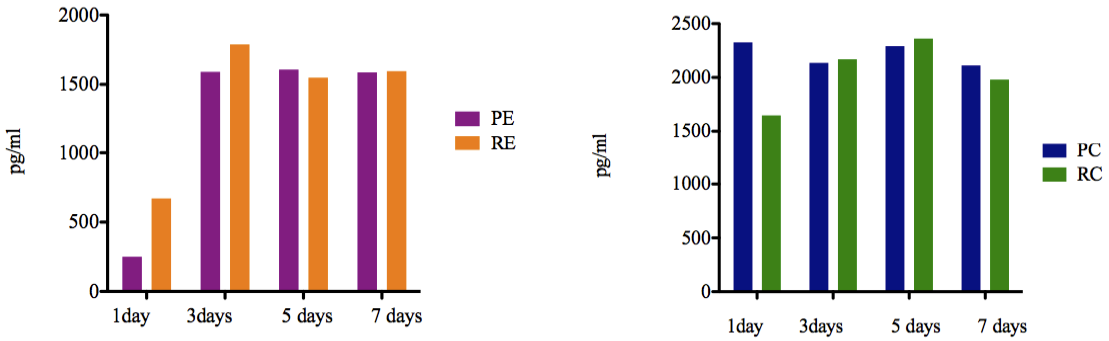


Figure 17 IFN-γ response to peptide and recombinant proteins in PBMC culture

PBMC from 2 persons were cultured with peptide and recombinant proteins at 250,000 cells/well (200ml) and incubated for 1, 3, 5 and 7 days. PE and PC denote, response from peptides of ESAT-6 and CFP-10 respectively; RE and RC denotes, recombinant proteins. There was no significant difference in the response between the activity of peptide and recombinant proteins.

IFN-γ response after seven days results from the daughter cells of those cells that initially proliferated rather than from the initially activated cells. For this purpose, the prolonged culture assay would depend on enough intact peptide being around during the course of the culture. The peptides have a greater sensitivity than recombinant proteins to degradation by proteases present in the whole blood (Nusslein et al., 2006; Yeaman et al., 2002). Therefore, I hypothesized that in-sufficient peptide was present to last for 7 days of culture to stimulate the proliferating T cells.

Peptides degrade in prolonged whole blood culture

In order to further test the hypothesis of possible peptide degradation or protease activity in the whole blood culture, we collaboratively investigated the nature of peptide mixtures and recombinant proteins in the supernatants of whole blood cultures at the point of setting up (day 0) and at the end of 7 days of incubation using SELDI analysis at Imperial College London.

1:10 diluted whole blood cultures were set up, and the supernatants were harvested at the respective time points and evaluated by SELDI. Protein Chip N20 arrays which mimic normal-phase chromatography with silicate functionality, broad protein binding chip were used. 15 µl of the culture supernatant was used for the analysis.

The peptides ESAT-6 and CFP-10 produced peaks at Day 0 (Figure 18) but the peaks were absent or dramatically reduced by Day 7. This finding further supports the hypothesis that degradation of these peptides is taking place and therefore might contribute to the reduced activity and reduced IFN-γ response after 7 days of incubation.

Table 5 ESAT-6 peptide and recombinant sequence and predicted molecular mass

Peptide number	Sequence	Molecular Mass
ESAT-6 1	MTEQQWNFAGIEAAA	1667
ESAT-6 2	WNFAGIEAAASAIQG	1506
ESAT-6 3	IEAAASAIQGNVTSI	1445
ESAT-6 4	SAIQGNVTSIHSLLD	1555
ESAT-6 5	NVTSIHSLLDEGKQS	1628
ESAT-6 6	HSLLDEGKQSLTKLA	1640
ESAT-6 7	EGKQSLTKLAAAWGG	1517
ESAT-6 8	LTKLAAAWGGSGSEA	1419
ESAT-6 9	AAWGGSGSEAYQGVQ	1467
ESAT-6 10	SGSEAYQGVQQKWDA	1654
ESAT-6 11	YQGVQQKWDATATEL	1738
ESAT-6 12	QKWDATATELNNALQ	1703
ESAT-6 13	TATELNNALQNLART	1630
ESAT-6 14	NNALQNLARTISEAG	1572
ESAT-6 15	NLARTISEAGQAMAS	1520
ESAT-6 16	ISEAGQAMASTEGNV	1465
ESAT-6 17	QAMASTEGNVTGMFA	1515

ESAT-6 Recombinant protein Sequence: Predicted weight: 11800 Da

M T V H M T E Q Q W N F A G I E A A A S A I Q G N V T S I H

S L L D E G K Q S L T K L A A A W G G S G S E A Y Q G V Q Q

K W D A T A T E L N N A L Q N L A R T I S E A G Q A M A S T

E G N V T G M F **A H V Q L E H H H H H H**

Table 6 CFP-10 peptide and recombinant protein sequence and predicted molecular mass

Peptide number	Sequence	Molecular
CFP-10 1	MAEMKTDAATLAQEA	1581
CFP-10 2	TDAATLAQEAGNFER	1594
CFP-10 3	LAQEAGNFERISGDL	1620
CFP-10 4	GNFERISGDLKTQID	1693
CFP-10 5	ISGDLKTQIDQVEST	1634
CFP-10 6	KTQIDQVESTAGSLQ	1605
CFP-10 7	QVESTAGSLQGQWRG	1604
CFP-10 8	AGSLQGQWRGAAGTA	1431
CFP-10 9	GQWRGAAGTAAQAAV	1415
CFP-10 10	AAGTAAQAAVVRFQE	1490
CFP-10 11	AQAAVVRFQEAANKQ	1631
CFP-10 12	VRFQEAANKQKQELD	1804
CFP-10 13	AANKQKQELDEISTN	1689
CFP-10 14	KQELDEISTNIRQAG	1702
CFP-10 15	EISTNIRQAGVQYSR	1722
CFP-10 16	IRQAGVQYSRADEEQ	1750
CFP-10 17	VQYSRADEEQQQALS	1752
CFP-10 18	ADEEQQQALSSQMGF	1669

CFP-10 Recombinant protein Sequence Predicted weight: 12600 Da

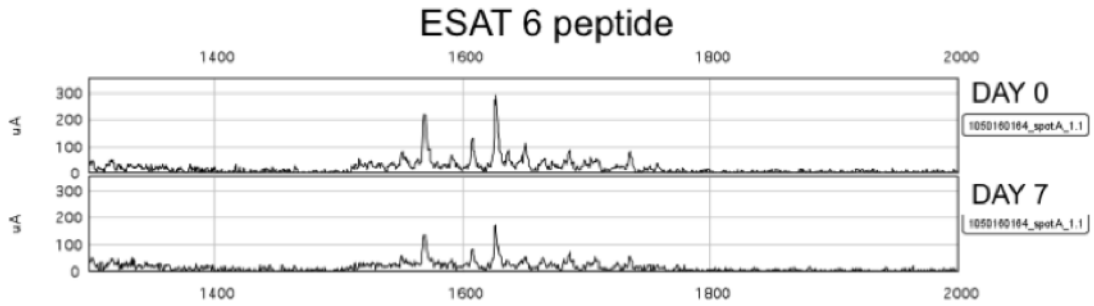
M T V H M A E M K T D A A T L A Q E A G N F E R I S G D L K T Q I D

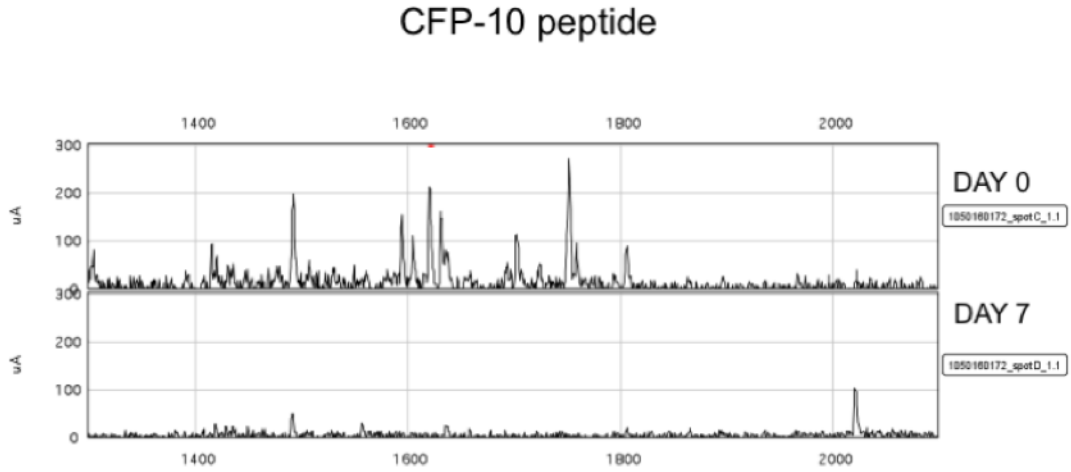
Q V E S T A G S L Q G Q W R G A A G T A A Q A A V V R F Q E

A A N K Q K Q E L D E I S T N I R Q A G V Q Y S R A D E E Q

Q Q A L S S Q M G F **H V Q L E H H H H H H**

A





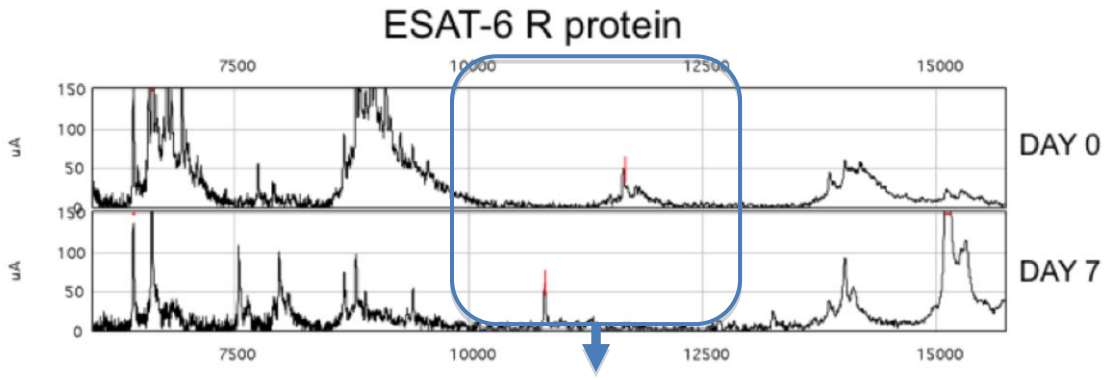
B

Figure 18 SELDI analysis of peptide mixture of ESAT-6 and CFP-10

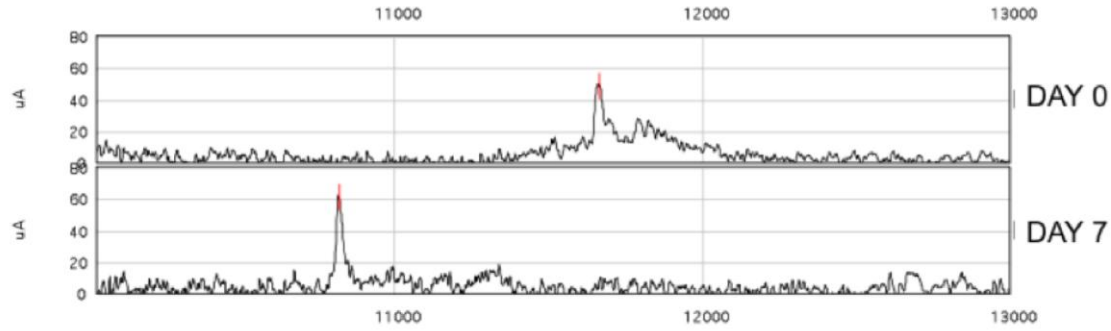
Shows the SELDI analysis of ESAT-6 (A) and CFP-10 (B) peptides in the supernatants at the time of set up (day 0) and at the end of the incubation (day 7). Day 0 results shows peaks with respective masses in both ESAT-6 and CFP-10 (table 6 and 7). The peaks were either reduced or not present at day 7.

By contrast, recombinant ESAT-6 and CFP-10 produced peaks both at Day 0 and Day 7 (Figure 19 and 20). Interestingly, the peaks produced by recombinant ESAT-6 and CFP-10 at Day 7 showed a shift to a lower molecular mass suggestive of some protease activity or cleavage of the protein.

A



B



C

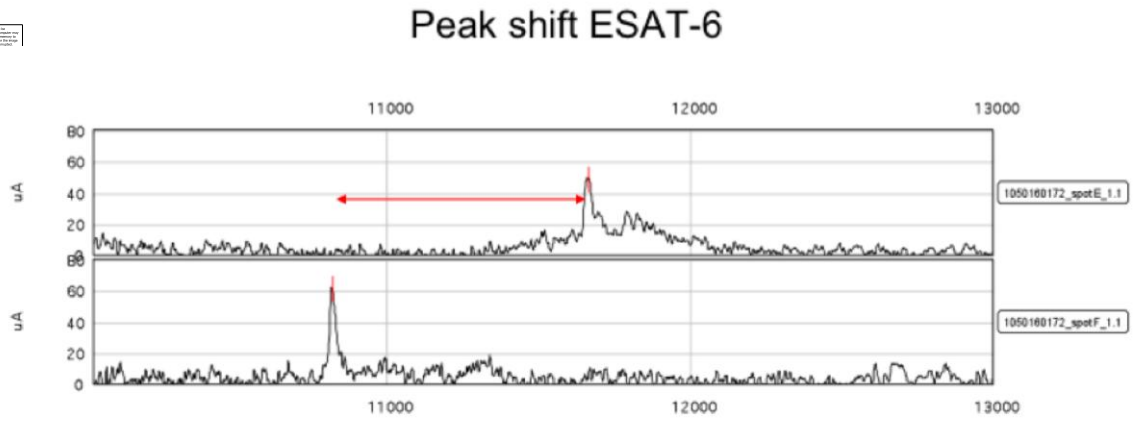


Figure 19 Recombinant ESAT-6 analysis by SELDI

Shows the SELDI analysis of ESAT-6 recombinant protein in the supernatants at the time of set up (day 0) and at the end of the incubation (day 7). Day 0 results shows peaks with respective masses in both ESAT-6 (table 6). At day 7 a peak was detected. However these peaks showed a shift towards a lower molecular weight.

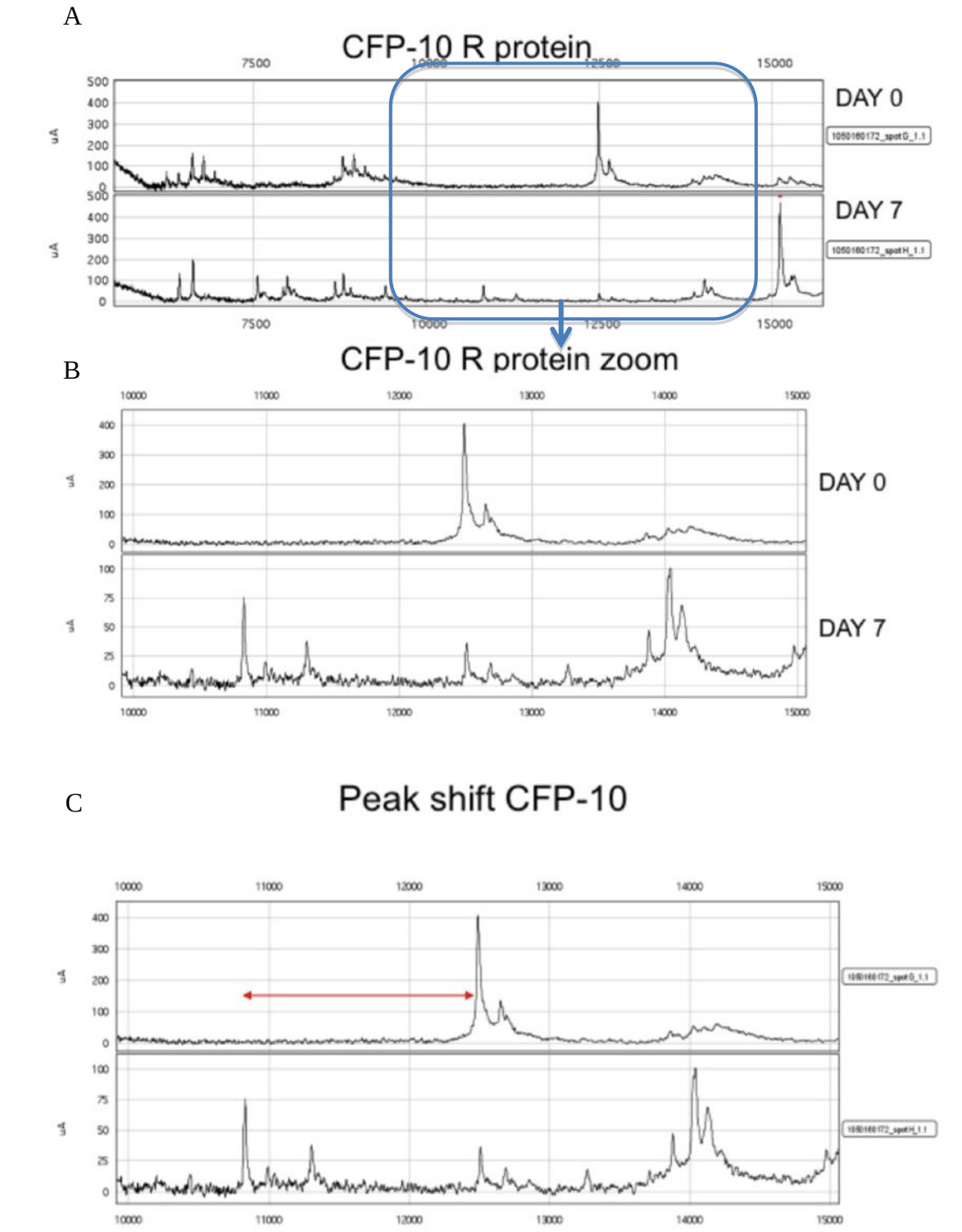


Figure 20 Recombinant CFP-10 analysis by SELDI

Shows the SELDI analysis of CFP-10 recombinant protein in the supernatants at the time of set up (day 0) and at the end of the incubation (day 7). Day 0 results shows peaks with respective masses in both ESAT-6 (table 6). At day 7 a peak was detected. However these peaks showed a shift towards a lower molecular weight.

The recombinant proteins are of higher molecular weight than that of the native protein. The theoretical molecular weight of ESAT-6 is 9.9-kDa and the recombinant CFP-10 is 11.8-kDa. The differences in the molecular weight are due to the flanking sequences from the expression vector including His tag (marked in red in figure 21, Table 16 and 17) in the recombinant protein.

A

ESAT-6 Recombinant protein Sequence

MTV H MT E Q Q W N F A G I E A A A S A I Q G N V T S I H
S L L D E G K Q S L T K L A A A W G G S G S E A Y Q G V Q Q
K W D A T A T E L N N A L Q N L A R T I S E A G Q A M A S T
E G N V T G M F A HV Q L E H H H H H H

B

CFP-10 Recombinant protein Sequence

MTV H MA E M K T D A A T L A Q E A G N F E R I S G D L K T Q I D
Q V E S T A G S L Q G Q W R G A A G T A A Q A A V V R F Q E
A A N K Q K Q E L D E I S T N I R Q A G V Q Y S R A D E E Q
Q Q A L S S Q M G F HV Q L E H H H H H H

Figure 21 Flanking region and His tag in recombinant ESAT-6 and CFP-10

For recombinant ESAT-6 the predicted molecular weight is 11kDa. The Day 0 SELDI m/z showed peaks at 11.7 kDa (3000 laser) and day 7 SELDI showed peak at 10.8kDa, a shift of 841 Da. Further analysis reveled that cleavage of " H H H H H H " produces a protein that would shift a protein peak by 841 Da.

Similarly recombinant CFP-10 (predicted molecular mass:12.6 kDa), Day 0 SELDI m/z showed peak at 12.5kDa (2500 laser) and Day 7 showed a peak at 10.8kDa. The cleavage of " G F H V Q L E H H H H H H " produces a protein that would shift a protein peak by 1652 Da.

These findings provide evidence that the peptides are degraded by day 7, and therefore not available to stimulate the proliferating T cells and therefore, most probably responsible for the reduced response observed. Similarly, this investigation also provide evidence suggestive of that even recombinant proteins may also be subject to exo-peptidase activity during or at 7 days of whole blood culture. This empirical data whereby the recombinant proteins stimulate a recall response after 7 days suggests that these cleavages do not affect the antigenic activity. However, further confirmation of these findings need to be performed with antibody depletion of the respective proteins.

In conclusion, the discrepancy in response between short and prolonged *in vitro* culture systems using whole blood cultures was investigated. The commercially available QFT-GIT assay was found to be less sensitive than the 7day whole blood culture with recombinant antigens for detection of latent tuberculosis (48 % vs. 90% positivity respectively) (Figure 12). This could be entirely attributed to the variation in the length of incubation of these assays: where the long term assay being more sensitive, due to the IFN-γ response from both effector and memory T cells, while the short term incubation assay being less sensitive as it detects only effector T cell response.

In addition, these investigations led to uncover the mystery of reduced performance of the peptide mixtures over the recombinant proteins of RD1-encoded antigens in prolonged whole blood culture, which has not been reported elsewhere. In long-term whole blood culture system, the recombinants perform better. The activity of peptide and recombinant proteins do not differ in short term or long-term PBMC cultures. The use of DMSO (up to 0.3% final concentration) was found not to affect the viability or the activity of the antigens. The decreased performance of peptides was found to be due to degradation during the course of incubation in whole blood culture system. In addition, these finding also suggest the possibility of exo-peptidase activity on recombinant proteins in the prolonged whole blood culture.

Chapter 4: Mapping T cell antigenic determinants of novel antigens: species specific Rv2654c and dosR regulated Rv1733c

Abstract

Species specific and stress-induced genes of MTB are considered as excellent targets to improve the diagnostic tests for tuberculosis and develop post exposure vaccine candidates. RD11 encoded Rv2654c and dosR encoded Rv1733c were previously described as dominant targets in active and latent tuberculosis respectively. T cell antigenic determinants of Rv2654c and Rv1733c were mapped in a high tuberculosis endemic area. Rv2654c and Rv1733c induced T cells IFN-γ response from persons with active and latent tuberculosis. Rv2654c showed focused dominant epitope region, while Rv1733c, a degenerate pattern of epitope spread throughout the length of the antigen. The most striking discovery however was that commercially used p38-55 peptide of Rv2654c is not immunodominant in this population, and the αα region covering p51-65 of Rv2654c was the dominant target. In addition, preliminary results show that adding p51-65 to the QFT-GIT TB Ag tube, increases the percentage positive score.

Introduction

The lack of a quick and reliable clinical laboratory test for early stage and latent tuberculosis has long represented a diagnostic dilemma. The tuberculin skin test (TST), using purified protein derivative (PPD), is largely utilized for both diagnosis and screening. The greatest drawback of using PPD is its broad cross-reactivity with antigens derived from several mycobacterial species (Borsuk et al., 2009) and greatly decreased specificity in patients with advanced disease or with immunodeficiency such as HIV-1 infection. The control strategies of tuberculosis, however depend largely on the partially effective vaccine, BCG and diagnosis and treatment of those who are at high risk of developing active disease (Maartens and Wilkinson, 2007). Therefore a diagnostic test, which is specific to MTB infection, is urgently needed (Bakir et al., 2009; Roupie et al., 2007).

Recent genome driven advances in MTB antigen discovery have identified genes that are absent from the vaccine strain of *M. bovis* BCG described as the regions of difference (RD) on the MTB chromosome (Behr et al., 1999; Brosch et al., 2002; Mahairas et al., 1996). These genes offer the advantage to distinguish between infection with MTB and prior vaccination with BCG. Two antigens encoded by RD1 (ESAT-6 and CFP-10) were found to be immunodominant and studied in much detail and have been included in the two recently commercialised blood based immunodiagnostic tests for TB (T-SPOT.TB and QuantiFERON (QFT) (Lalvani and Pareek, 2009b; Pai et al., 2008; Richeldi et al., 2009).

Recent studies of genes other than RD1-encoded immunodominant targets revealed RD11-encoded, Rv2654c. Rv2654c has been previously described and mapped in a prolonged proliferative assay (Aagaard et al., 2004), and one epitope was found to be

dominant for IFN-γ response in 7 out of 9 persons with active tuberculosis as opposed to those with latent infection or BCG vaccinated controls in the tested Danish population (Aagaard et al., 2004; Brock et al., 2004).

Based on these promising yet limited data, one peptide of Rv2654c (termed p4) was included in the whole blood based IFN-γ release assay QuantiFERON TB-Gold In-tube (QFT-G-IT) in addition to ESAT-6 and CFP-10 (Aagaard et al., 2004; Brock et al., 2004), and shown to increase the sensitivity and specificity of those 3rd generation tests to detect active tuberculosis (Parkash et al., 2009).

In vitro studies exploring the conditions which MTB is thought to encounter *in vivo*, during the course of latent infection have described the dosR regulon, which is induced and regulates a group of genes when MTB cultures were exposed to 2 hours of Hypoxia (Roberts et al., 2004; Rustad et al., 2009; Sherman et al., 2001). Hypoxia has been shown to be a condition that is present within the granulomas, which are the hallmark of the tuberculosis infection (Via et al., 2008). DosR encoded proteins have been well evaluated, and there are claims that these antigens are preferentially recognized in those with latent infection (Black et al., 2009; Goletti et al., 2009; Leyten et al., 2006; Schuck et al., 2009). One such well evaluated candidate Rv1733c has been shown to be promising in animal models to induce protective immune response (Roupie et al., 2007). Therefore, Rv1733c has the potential to be developed into a post exposure vaccine (Bivas-Benita et al., 2009; Black et al., 2009; Leyten et al., 2006; Roupie et al., 2007) but has never been mapped for human T cell epitopes.

Rv2654c and Rv1733c have been previously described, to be preferentially recognized in persons with active and latent disease respectively based on a prolonged proliferation based assay. However, short-term assays are attractive and preferred in

the clinical and research settings for operational reasons, as the results could be available faster (Pai et al., 2009). These candidates (Rv2654c and Rv1733c) have neither been mapped nor evaluated in short term culture conditions, despite Rv2654c (p4) being predominantly used in short-term whole blood assay (Brock et al., 2004).

Therefore, I was interested to evaluate the pattern of T cell determinants of species specific Rv2654c and dosR (infection stage specific) Rv1733c antigens in a high tuberculosis endemic setting, to evaluate the performance in a short term *in vitro* cell culture assay and to determine the pattern of immunodominant epitopes, which could potentially be used to improve the immunodiagnosis of tuberculosis.

Results

29 Healthy laboratory and clinical volunteers, with latent TB infection, and no symptoms of active tuberculosis, who agreed to participate, were included. Participants self reported on BCG, TST and HIV-1 infection status. Latent tuberculosis (LTBI) was defined by an interferon-gamma enzyme linked immunospot (ELISpot) response to ESAT-6 or CFP-10 of > 20 spot forming cells (SFC)/10^6 PBMC. In addition, 10 persons with active tuberculosis (ATB) were recruited from the Ubuntu clinic in Khayeltisha and were defined by a positive culture of *M. tuberculosis* from one or more sputum specimens. ATB patients were sampled prior to commencing antitubercular chemotherapy. Known immunosuppression for HIV-1 and other reasons, age < 18 years and pregnancy formed other exclusion criteria. The baseline characteristics of subjects enrolled to the study are shown in Table 7.

	Latent TB	Active TB
Number	29	10
Sex	10M:21F	6M:4F
Median Age	27 years	27 years
BCG vaccinated	27 vaccinated 1 not vaccinated 3 Unknown	2 vaccinated 4 not vaccinated 4 Unknown
TB smear/culture positive	NA	5 smear and culture +; 1 smear -, culture +; 4 smear +, culture results unknown.

Table 7 Demographic details of persons included Chapter 4

IFN-γ analysis of Rv2654c in active and latent tuberculosis

To determine if there was evidence of preferential recognition, I analyzed the pattern of recognition as a pool of peptides covering the whole protein in those with active TB (n=10) and latent TB (n=26). The median background IFN-γ ELISpot response were 5 SFC/10^6 PBMC (IQR: 0-10) in those with latent TB and 0 SFC/10^6 PBMC (IQR: 0-9) in those with active TB. A positive IFN-γ response was defined as >15 SFC/10^6 PBMC.

When analyzed as the SFC response, the IFN-γ responses were weak and were similar between the two groups. Although persons with active TB tend to have slightly higher median response when compared to those with latent TB (10 SFC/10^6 PBMC, IQR: 6-25 *vs.* 4 SFC/10^6 PBMC, IQR:0-25; p=0.3194), however, this difference in the median response was not statistically significant (Figure 22A). When analyzed by the proportion of responders, Rv2654c was recognized (>15 SFC/10^6 PBMC) in 4/10 persons with ATB (40%) while 8/26 persons with LTBI (31%) (Figure 22 B).

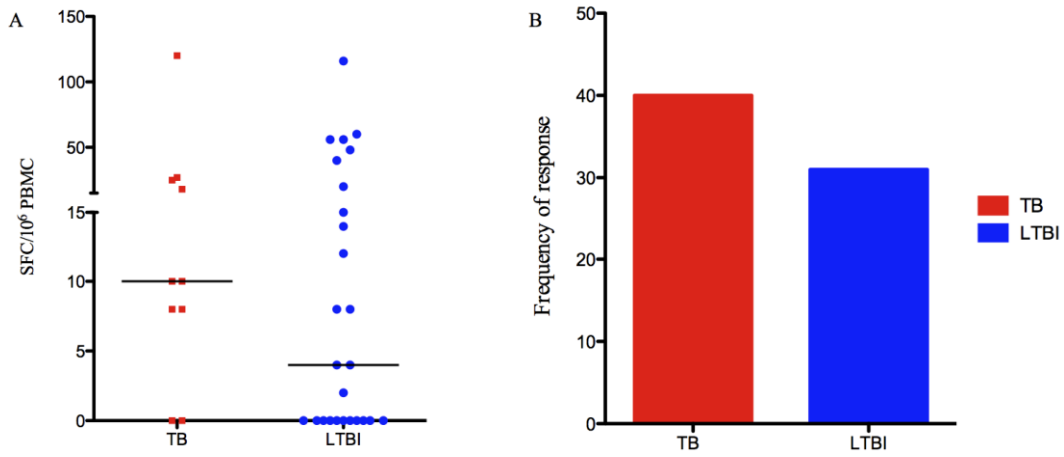


Figure 22 IFN-γ response in Active and Latent TB

Panel A shows the IFN-γ SFC/10^6 PBMC to Rv2654c peptide pool in persons with latent and active tuberculosis. The median response was not significantly different between the two groups. Panel B shows the frequency of response to Rv2654c in latent and active tuberculosis.

Peptide map of IFN-γ response to Rv2654c

PBMC from 26 persons with LTBI was used to map seven peptides, covering the entire region of Rv2654c (Figure 23). In addition 10 persons with active tuberculosis were also mapped (Figure 24) to evaluated differential patterns were to be found. The peptide sequences used in this evaluation are described in Table 9.

All persons with latent tuberculosis responded to ESAT-6 or CFP-10 with >15 SFC/10^6 PBMC of IFN-γ response. Overall, the responses to ESAT-6 and CFP-10 were higher than that of any peptide or peptide pool in both active and latent tuberculosis ($p<0.0001$).

When analysed by the proportion of responders, >15% responded (>15 SFC/10^6 PBMC) with IFN-γ response to every peptide tested in both groups with the exception of peptide covering the αα region 11-30 (latent and active) and 41-60 (only active disease). Over all the responses were moderate with the highest median response being 8 SFC/10^6 PBMC in latent tuberculosis and 18 SFC/10^6 PBMC in active tuberculosis ($p= 0.179$) for the peptide covering the region 61-81 and the peptide pool of Rv2654c (Figure 23A and 24 A).

When analyzed by the frequency of response (defined by >15 SFC/10^6 PBMC), Rv2654c pooled peptides were recognized by 31% and 40% in latent and active tuberculosis respectively.

Peptides covering the region 61-81 were the most recognized in 35% and 50%; 51-70 (27% and 30%) in latent and active tuberculosis respectively; peptides covering region 1-20 was recognized by 31% only in latent tuberculosis. All other peptides had a lower frequency of recognition (Figure 23B and 24B).

Table 8 Peptide Sequence of Rv2654c

Rv2654c sequence	
P1 (1-20)	MSGHALAARTLLAAADELVG
P2 (11-30)	LLAAADELVGGPPVEASAAA
P3 (21-40)	GPPVEASAALAGDAAGAWR
P4 (31-50)	LAGDAAGAWRTAAVELARAL
P5 (41-60)	TAAVELARALVRAVAESHGV
P6 (51-70)	VRAVAESHGVAAVLFAATAA
P7 (61-81)	AAVLFAATAAAAAAVDRGDPP

Figure 23 Peptide map of Rv2654c in Latent tuberculosis

Panel A shows IFN-γ SFC response to the 20 mer peptide. Bars indicate median response.

Panel B shows Frequency of responders for a positive IFN-γ response, defined by >15 SFC/10^6 PBMC.

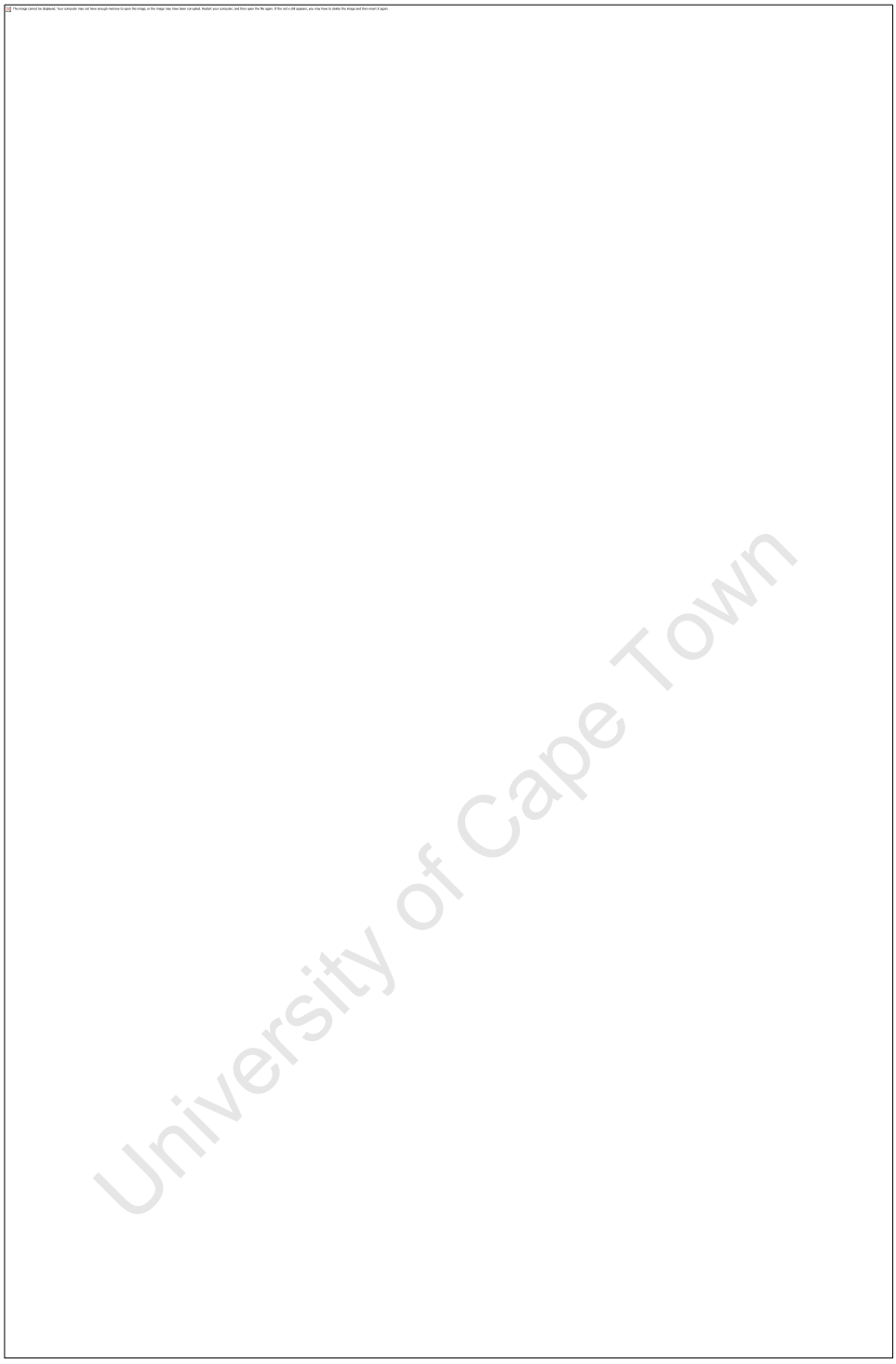


Figure 24 Peptide map of Rv2654c in Active tuberculosis

Panel A shows IFN-γ SFC response to the 20 mer peptide. Bars indicate median response.

Panel B shows Frequency of responders for a positive IFN-γ response, defined by >15 SFC/10^6 PBMC.

In order to further study the epitopic targets within the highly recognized peptide regions, 15-mer truncated peptides with 10 αα overlap were synthesized. In addition, the previously described, immunodominant and preferentially recognized (in ATB) peptide covering the 38-55 (p4) (Aagaard et al., 2004), was also synthesized to be tested in this setting. Origins of the truncated peptides are described in Figure 25.

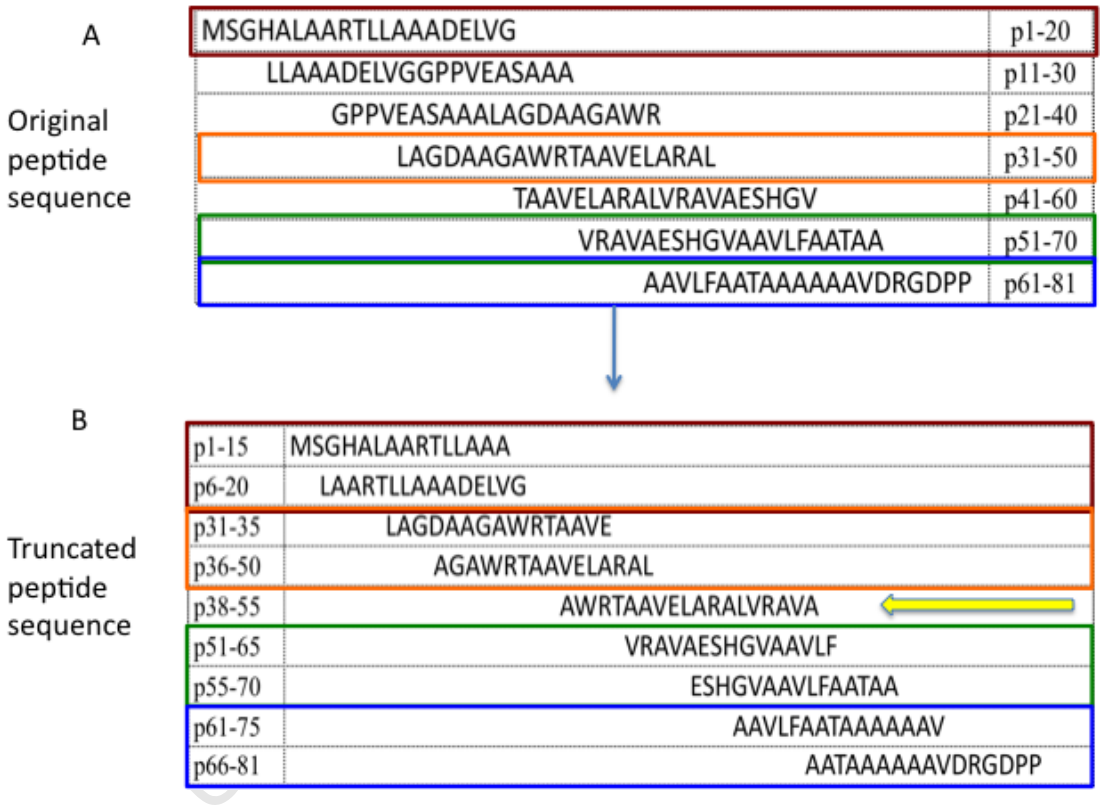


Figure 25 Sequence of Truncated peptide sequence from original sequence of Rv2654c

Panel A shows the original peptide sequence, with highlights on highly recognized sequence
Panel B shows truncated peptides originating from the highly recognized sequence. Arrow (yellow) indicates previously described immunodominant peptide covering αα region 38-55 of Rv2654c.

The IFN-γ responses to the truncated peptides in 10 persons with active tuberculosis in addition to 7 further persons (truncated peptides) with latent tuberculosis. Due to the limited availability of cells, peptide 1-15, 6-20, 31-45 and 36-50 were tested in 10; 38-58 and 51-65 in 14 and 55-70 in 26 LTBI persons.

The response in those with active tuberculosis was tending to be slightly higher than those with latent tuberculosis, although the overall response was weak. The highest IFN-γ SFC response was to peptide sequence covering the αα region 51-65 in persons with ATB and LTBI (n=9) (25 SFC/10^6 PBMC *vs.* 16 SFC/10^6 PBMC respectively; p=0.6) but the responses were not significantly different (Figure 26 A and B). All other peptides had lower response in both clinical groups with the exception of 61-81 and 55-70 (18 SFC/10^6 PBMC and 10 SFC/10^6 PBMC) in ATB and 61-81 (8 SFC/10^6 PBMC) in LTBI.

When analyzed by the frequency of responders, the peptide sequence covering the region 51-65 had the highest response in both active (67%) and latent (50%) disease, the difference being not significant (p=0.7) (Figure 26. C and D). All other peptide sequences were less recognized. The peptide sequence covering the region 38-55 was recognized by only 20% and 14% in active and latent tuberculosis respectively.

Figure 26 IFN-γ analysis of Rv2654c in active and latent tuberculosis

Panel A and B shows the IFN-γ SFC response in LTBI (Blue) and active TB (red) to the peptides of Rv2654c, represented as SFC/10^6 PBMC. Bars indicate median response. Dotted lines indicate 15 SFC/10^6 PBMC, the cut-off for a positive response.

Panel C and D shows the frequency of response in LTBI (blue) and Active TB (red) to the peptides of Rv2654c (>15 SFC/10^6 PBMC).

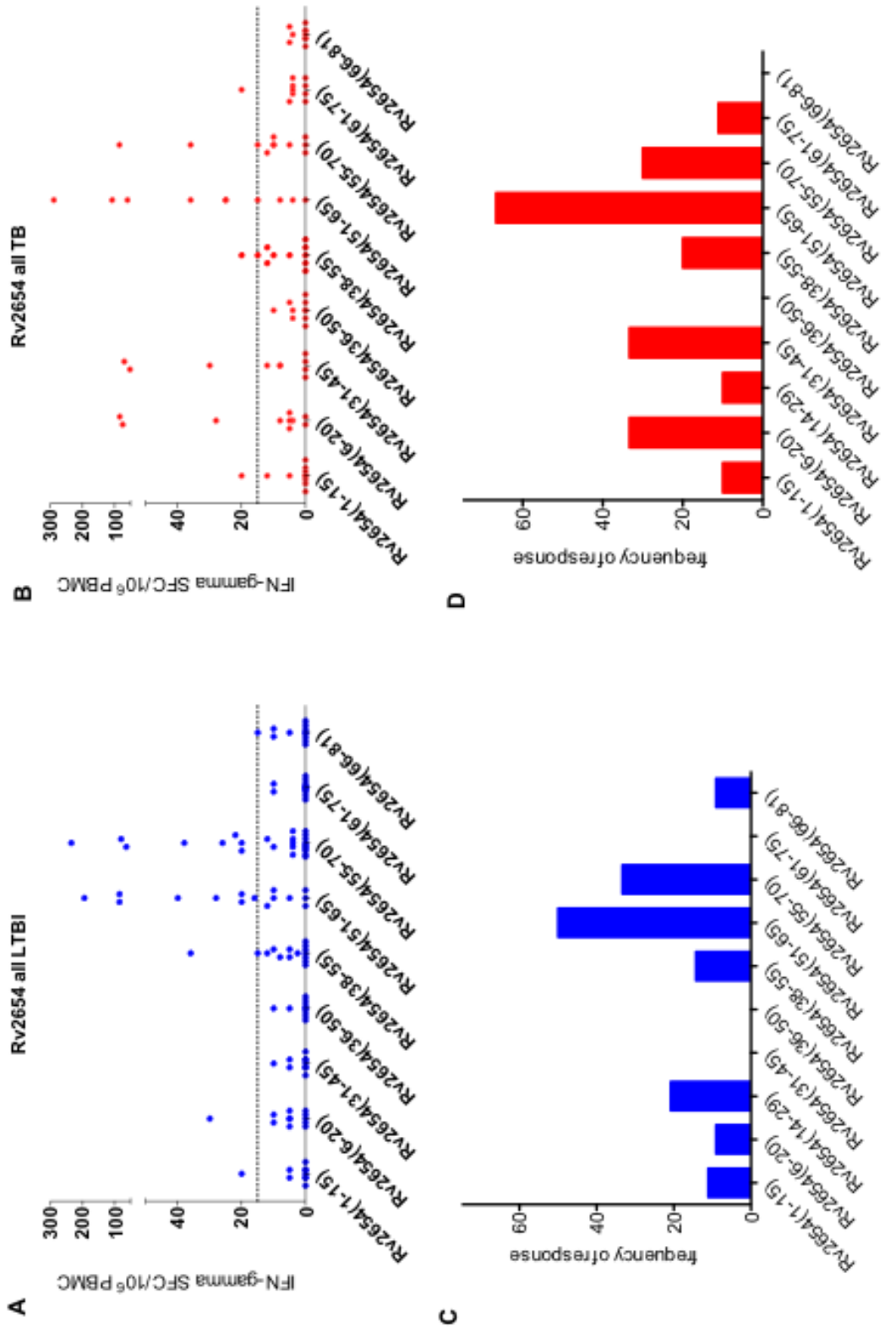


Figure 26 IFN-γ analysis of Rv2654 in active and latent Tuberculosis

Comparison of responses between 4 peptides covering the αα region 38-55, 51-70, 51-65 and 55-70

The striking finding in this study was the dominant IFN-γ response to peptide region covering the region 51-65 in both active and latent disease. It was also found that adding 5αα (p51-70) at the C terminal reduces the response, while removing 5αα from the N-terminal (p55-70) also reduces the response (Figure 27 and 28), suggesting that the epitope core of this peptide is present within the region p55-65. The previously described immunodominant region covering peptide sequence 38-55 performed poorly with only a moderate median response of 5 and 3 SFC/10^6 PBMC (p=0.7) and recognized in 20% and 14% in ATB and LTBI respectively.

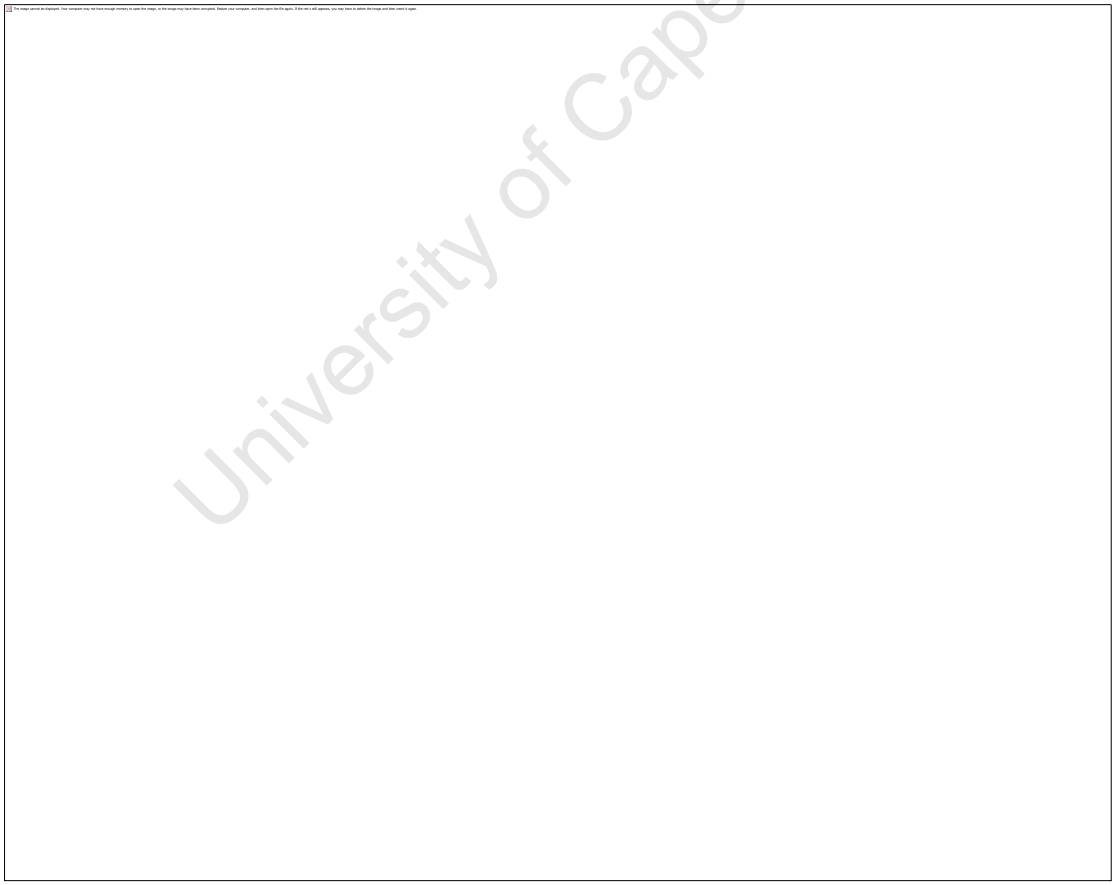


Figure 27 IFN-γ response to peptides of Rv2654c in active and latent tuberculosis

Panel A and C shows the SFC/10^6 PBMC response in LTBI and ATB for peptide regions 38-55, 51-70, 51-65 and 55-70.
Panel B and D shows frequency of responders. Peptide covering the region 51-65 had both the dominant IFN-γ SFC response and most frequently recognized in both ATB and LTB.

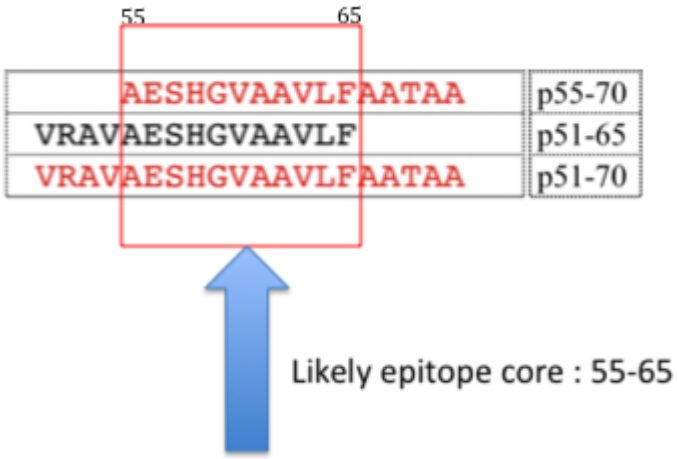


Figure 28 Likely epitope core of Rv2654c.

IFN-γ response of ESAT-6, CFP-10 and Rv2654c

In light of these findings, I compared the IFN-γ response between commercial QFT-GIT (obtained at recruitment) with the in house ELISpot response to ESAT-6, CFP-10 and Rv2654c peptide pool. By QFT- GIT, 14 persons scored a positive response (defined by ≥0.35 IU/ml) for IFN-γ. However, 26 of the 29 recruits had a positive response (defined by ≥20 SFC/10^6 PBMC) by ELISpot. The frequency of response to sum of ESAT-6, CFP-10 and Rv2654c by ELISpot was significantly higher than the frequency of positive response by QFT- GIT assay (90 % *vs.* 45%; p=0.001), although the test for agreement (kappa statistics) between these two assays was fair (k=0.310, 95% CI: 0.07 to 0.6) (Figure 29).

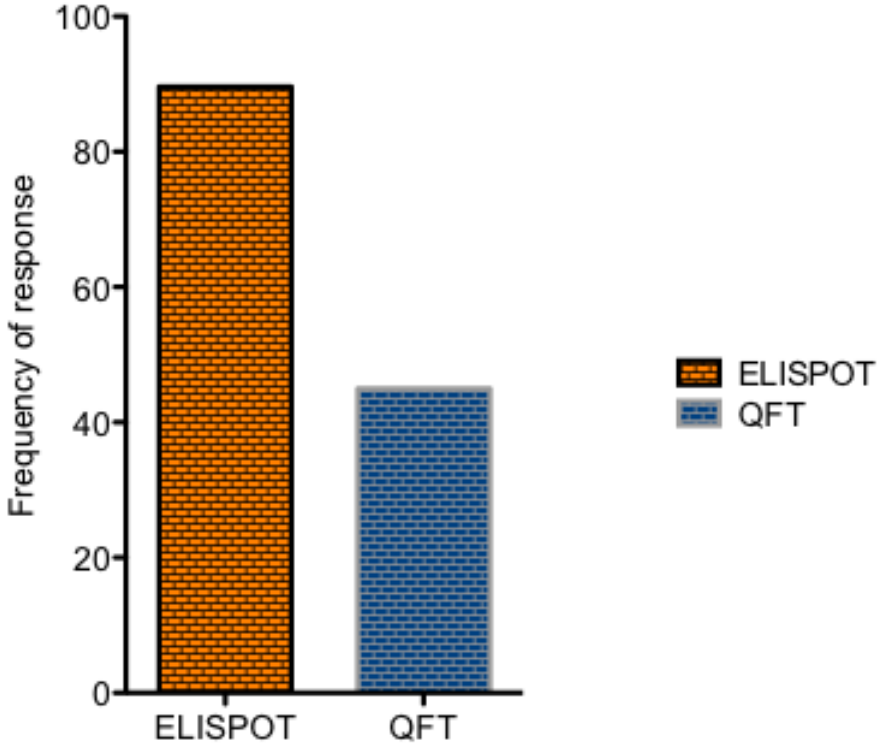


Figure 29 Frequency of response (IFN-γ) by ELISpot and QFT

QFT boosting experiment

I reasoned that the lower positive responses in QFT as opposed to the ELISpot with ESAT-6, CFP-10 and Rv2654c (Figure 26) might be due to the absence of the immunodominat peptide of Rv2654c p51-65 in the QFT, which contains only p38-55 of Rv2654c. Therefore, I reasoned that it would be interesting to investigate whether adding p51-65 of Rv2654c to QFT-GIT tubes would increase the proportion of results positive of these assay.

Blood from 15 volunteers was obtained, and were set up with normal (Nil, TB Ag and Mitogen) tubes and in addition, Nil and TB antigen tube to which 5 µg/ml of p51-65

was added and incubated as that of normal QFT-GIT. Supernatant were collected and QFT ELISA were performed to measure IFN-γ (IU/ml).

8 out of 15 (53%) had a positive response to TB Ag- Nil. However, boosting with p51-65 increased the proportion of responders to 12 out of 15 (80%). 4 people remained negative (Figure 30). This finding suggests, recognition of a different region of the antigen Rv2654c in different populations, and that boosting QuantiFERON Gold In Tube with the immunodominat peptide (p51-65) increases the proportion scored positive and thus potential sensitivity to detect TB infection.

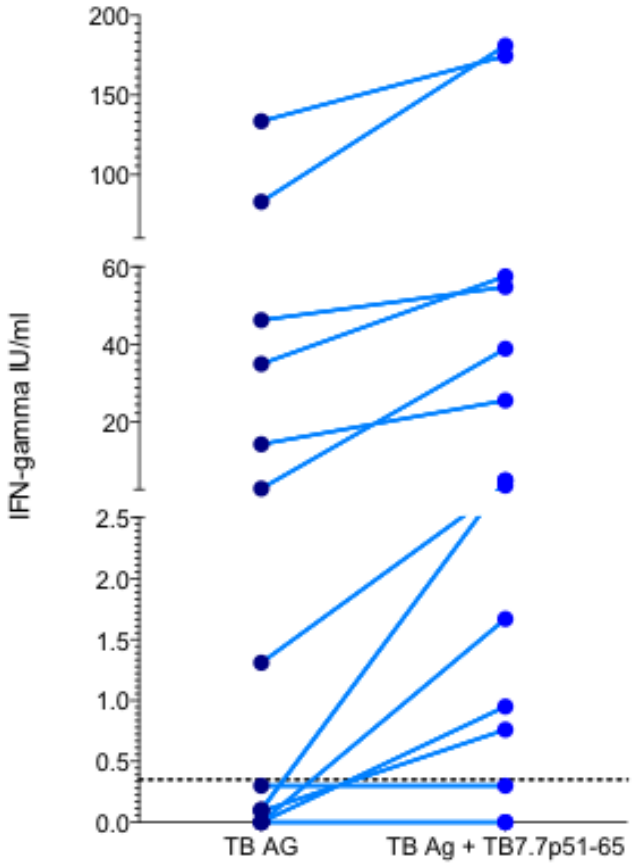


Figure 30 QFT GIT boosting experiment: TB antigen with and without p51-65 of Rv2654c

Shows the IFN-γ IU/ml before (TB AG) and after adding peptide 51-65 (TB AG +TB 7.7 p51-65) to the TB antigen tube of QFT –GIT. The results are corrected for the background. The QFT ELISA was repeated for those who score >10 IU/ml with diluted sample, and the values were extrapolated using the standard curve.

IFN-γ analysis of Rv1733c in active and latent tuberculosis

Rv1733c, a dosR encoded antigen has been described to be preferentially recognised in those with latent infection as opposed to those with active tuberculosis disease (Leyten et al., 2006). Its potential as a post exposure vaccine has also been exploited in animal models and found to be promising (Bivas-Benita et al., 2009; Roupie et al., 2007).

The ability of pooled peptides of Rv1733c to restimulate IFN-γ in 10 persons with active tuberculosis was analyzed, and compared it with the responses in those with latent tuberculosis (n=25).

The median SFC response induced by Rv1733c was tended to be slightly higher in those with latent disease when compared to those with active disease (12 SFC/10^6 PBMC, IQR 5-38 *vs.* 7 SFC/10^6 PBMC, IQR 0-34; p=0.29), however this difference was not significantly different (Figure 31). The reverse pattern in RD1-encoded antigens (ESAT6+CFP-10) was also noted, where the SFC responses were higher in those with active disease (109 SFC/10^6 PBMC, IQR 36-163 *vs.* 20 SFC/10*6* PBMC, IQR 24-176; p=0.71). Although there was no evidence of significant difference in the preferential recognition in those with latent tuberculosis, these findings are similar to the pattern of recognition of Rv1733c as previously reported (Leyten et al., 2006).

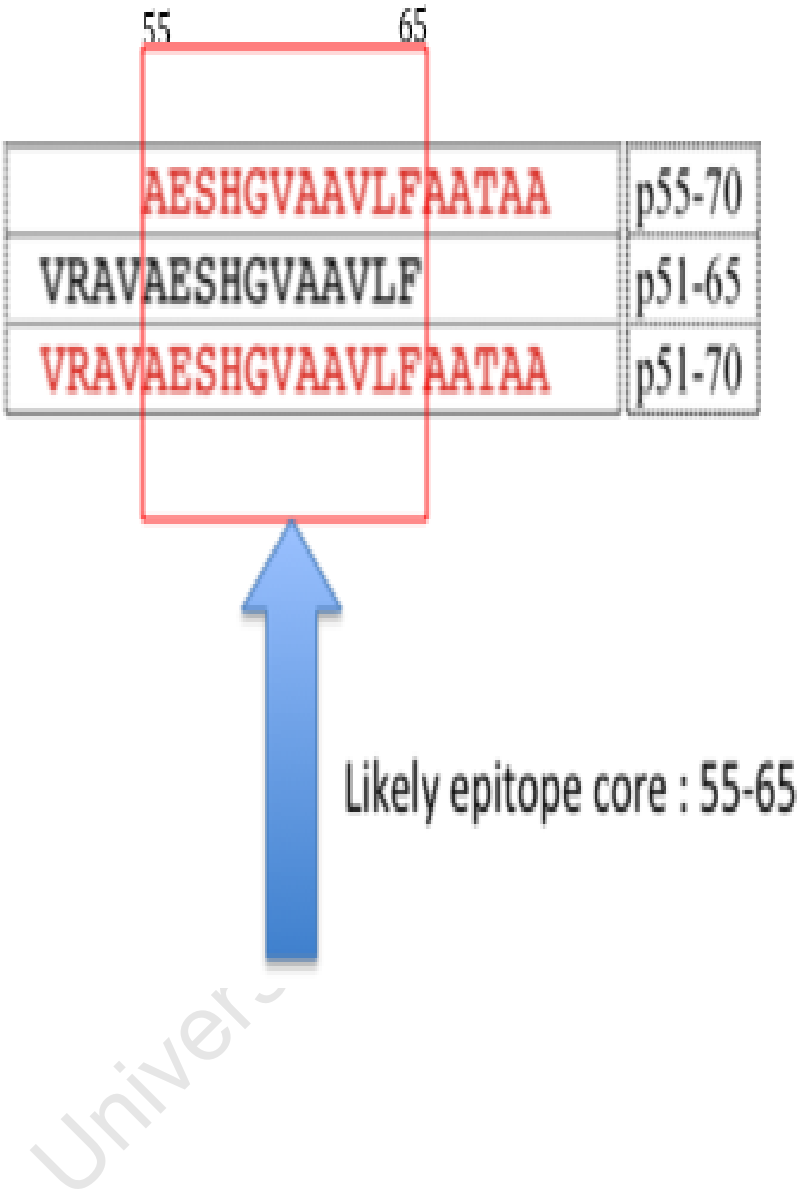


Figure 31 IFN-γ response to Rv1733c in active and latent tuberculosis

Shows the IFN-γ SFC/ 10^6 PBMC response to persons with active (n=10) or latent (n=25) tuberculosis to Rv1733c and ESAT-6+CFP-10. The lines indicate median response. Active TB are shown in red (●) and latent tuberculosis in blue (●)

Rv1733c induced a slightly higher (not different) median response in those with latent tuberculosis, whereas ESAT-6 and CFP-10 induced a higher (not different) response in those with active disease. However, these differences were not statistically significant.

Peptide map of Rv1733c

Therefore, I proceeded to determine dominant peptides in Rv1733c using peptide mapping. As this antigen was described to be preferentially recognized in latent infection, persons with active TB were not recruited to map T cell epitopes for Rv1733c. IFN-γ responses to the peptides of Rv1733c, was evaluated in 26 healthy, RD1 positive volunteers using ELISpot assay. Every peptide stimulated a response in at least 5 donors.

When analysed by the median IFN-γ response, the highest response was to the pooled peptides of Rv1733c (median 33 SFC/10^6 PBMC, IQR 8-70), which was greater (not significantly) than that of the ESAT-6 (15 SFC/10^6 PBMC, IQR 0-70; p= 0.7) and CFP-10 (17 SFC/10^6 PBMC, IQR 0-146; p=0.7).

Amongst the peptide sequences, the highest median responses were from αα region covering 171-190 (11 SFC/10^6 PBMC, IQR 1-37) and 41-60 (10 SFC/10^6 PBMC, IQR 1-20). However there was no peptide residue with dominant SFC response (Figure 32 A).

When analyzed by the frequency of response (>15 SFC/ 10^6 PBMC), all peptides were recognized by >20% of the donors. The peptides with αα region 51-60, 101-121 and 151-170 were recognized by >40% of the donors and 71-90,171-190 and 191-210 were recognized by 50% of the donors. The highest frequency of response was however to the peptide pool of Rv1733c (63%) (Figure 32 B). There was no dominant pattern of recognition, and the epitope targets seem to be dispersed throughout the protein sequence.

Table 9 Peptide sequence of Rv1733c

Rv1733c sequence	
P1 (1-20)	MIATTRDREGATMITFRLRL
P2 (11-30)	ATMITFRLRLPCRTILRVFS
P3 (21-40)	PCRTILRVFSRNPLVRGTDR
P4 (31-50)	RNPLVRGTDRLEAVVMLLAV
P5 (41-60)	LEAVVMLLAVTVSLLTIPFA
P6 (51-70)	TVSLLTIPFAAAAGTAVQDS
P7 (61-81)	AAAGTAVQDSRSHVYAHQAQT
P8 (71-90)	RSHVYAHQAQTRHPATATVI
P9 (81-100)	TRHPATATVIDHEGVIDSNT
P10 (91-110)	DHEGVIDSNTTATSAPPRTK
P11 (101-121)	TATSAPPRTKITVPARWVVNG
P12 (111-130)	ITVPARWVVNGIERSGEVNA
P13 (121-140)	GIERSGEVNAKPGTKSGDRV
P14 (131-150)	KPGTKSGDRVGIWVDSAGQL
P15 (141-160)	GIWVDSAGQLVDEPAPPARA
P16 (151-170)	VDEPAPPARAIADAALAALG
P17 (161-180)	IADAALAALGLWLSVAAVAG
P18 (171-190)	LWLSVAAVAGALLALTRAIL
P19 (181-200)	ALLALTRAILIRVRNASWQH
P20 (191-210)	IRVRNASWQHDIDSLFCTQR

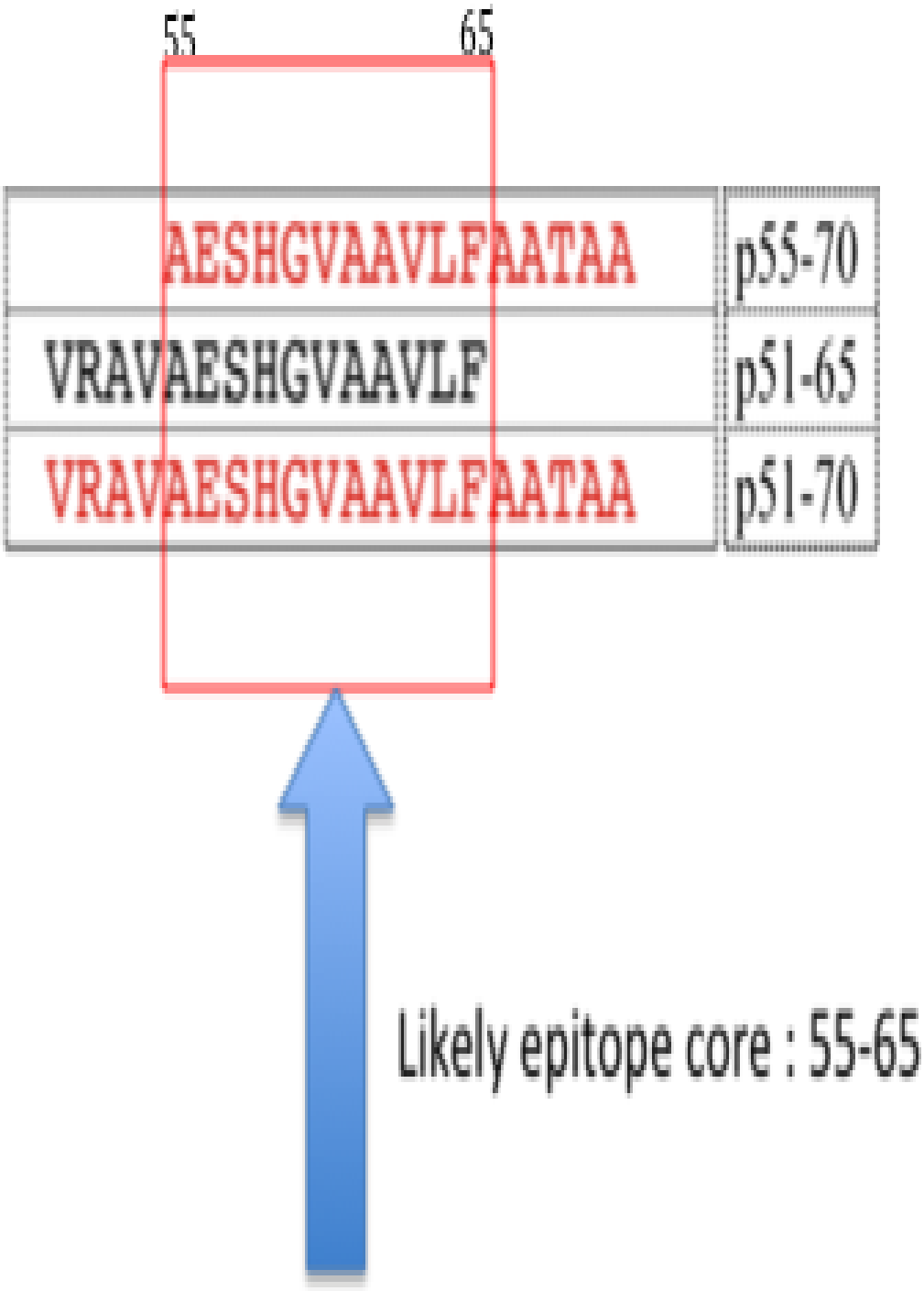


Figure 32 Peptide Map of Rv1733c

Panel A shows the IFN-γ SFC/106 PBMC for peptide of Rv1733c, ESAT-6 and CFP-10. Dotted line indicates the cut off for a positive response15 SFC / 10^6 PBMC.

Panel B shows the frequency of recognition of peptides of Rv1733, and peptide pools of Rv1733, and sum of ESAT-6 and CFP-10.

Discussion

As species specific and infection stage specific immunodominant targets of *M. tuberculosis* are important in considering potential candidates to improve immunodiagnosis of tuberculosis and developing post exposure vaccines, T cell targets of Rv2654c and Rv1733c were evaluated, in a setting of high tuberculosis endemicity. Rv2654c and Rv1733 re-stimulate T cells from persons in active and latent tuberculosis. Although there was slightly higher median SFC response in persons with active and latent tuberculosis for Rv2654c and Rv1733c respectively, these differences were not statistically significant.

Rv2654 had focused dominant epitope region, whereas, in Rv1733, there was a degenerate pattern of epitopes spread through-out the length of the antigen. The most striking finding however is the immunodominant region covering p51-65 of Rv2654c, and that the previously described dominant epitope covering the region p38-55 was recognized poorly by those with active and latent disease in this population. Preliminary data suggesting a boosting effect of p51-65 in the QFT TB Gold in tube assay (QFT-GIT) is also described.

Synthetic overlapping peptides are as efficient in detecting T-cell responses as recombinant proteins, with the advantage of faster production and lower costs and in the same time allow mapping of the epitope regions, enabling a focus on the immunodominat regions as potential candidates for vaccine development (Jurcevic et al., 1996). Evaluation of T cell epitopes is highly influenced by the MHC II allelic polymorphism present within different populations, and the infection pressure in people living in low and high endemic areas may contribute to the different patterns of recognition in different populations (Ovsyannikova et al., 2008). There is also evidence of discrepancy in the responses between short and prolonged incubation time in assays, due to difference in the phenotype of T cells responsible for the response (Beveridge et al., 2008; Leyten et al., 2007) and as decribed in

Chapter 3. Studies on species-specific genes of MTB have been very attractive for their specificity and are considered to be appropriate candidates for development of diagnostics. Recent interest in defining LTBI, based on immune response to the antigens which are differentially expressed during stress related conditions, led to the discovery of interesting immunodominat antigens. These formed the basis of selection of Rv2654c and Rv1733c for evaluation.

Rv2654c, is RD-11 encoded *phiRv2* phage insert, which are absent in strains other than *M. tuberculosis* (especially in *M. bovis* and in all BCG vaccine strains). The specificity of Rv2654c is greater than RD1-encoded CFP-10 or ESAT-6 such that it is only present within the *M. tuberculosis* species, whereas RD1-encoded genes are also shared amongst some environmental *Mycobacteria*. Therefore, an immune response to this antigen suggests exposure and infection purely by MTB, and not merely due to BCG vaccination or exposure to environmental *Mycobacteria*. The limitation however was that no group of person without tuberculosis disease or latent infection was included. This is due to the high tuberculosis endemic setting in which this study was done.

The promising evidence of the immunodominance of Rv2654c is based on prolonged proliferative assay, and based on this evidence, a peptide has been included in the commercial whole blood based IFN-γ response assay QFT-GIT (Aagaard et al., 2004; Brock et al., 2004; Lalvani and Pareek, 2009b; Leyten et al., 2007; Pai et al., 2008) which is an overnight, short term assay. The sensitivity and specificity of QFT has been evaluated with much variable results on the positive predictive value. In my own experience, there was relatively poor agreement for detection of prevalent TB when QFT-GIT or Mantoux (TST) was compared to sputum culture for MTB in HIV-infected people recruited from the same population (Gideon HP et al, abstract presented at the 5th IAS Conference on HIV-1 Pathogenesis, Treatment and Prevention, Cape Town, July 2009) (data not shown). This might be explained, based on our finding that the immunodominant

epitope in this population is different (p51-65 as opposed to 38-55 in QFT-G IT), and that there is a boosting effect in response upon addition of p51-65.

Rv1733c is a DosR-encoded immunodominat antigen, with promising experimental evidence in mice (Roupie et al., 2007). The pattern of T cell epitope recognition is found to be similar to that of the well known RD-1 encoded antigen ESAT-6 (Pathan et al., 2001). Human and animal studies have described the potential of this antigen to be included in development of post exposure vaccine candidate. However, concerns have also been raised due to the cross reactive immune response to this antigen in individuals infected with other *Mycobacteria* (Lin et al., 2009a). Further studies are required to validate its contribution and influence in the potential novel TB vaccines.

In summary, both Rv2654c induced an IFN-γ response in both active and latent tuberuclosis. No evidence of stage specific recognition was observed. Rv2654c has a focused T cell epitope, which is different from what is previously described. This may be due to the MHC polymorphism present in different population, rendering binding of different regions of the protein. The out come of this study would be to evaluate whether adding p51-65 to QFT-GIT TB Ag tube would increase the percentage positivity of the test in HIV-1 infected persons. Rv1733c is a potential antigen; further studies are required to evaluate its potential as a diagnostic candidate.

Chapter 5: Bioinformatic and empirical analysis of novel hypoxia-inducible targets of the human anti-tubercular T cell response

Abstract

Hypoxia is a relevant stimulus that *Mycobacterium tuberculosis* encounters *in vivo*. Therefore whole genome based transcriptional profiles of *M. tuberculosis* subject to prolonged hypoxia (described as the Enduring hypoxic response (EHR)) were analyzed, to guide the discovery of novel potential antigens, by a combined bioinformatic and empirical approach. Analysis included fold induction of 100 highly induced genes at 7 days of hypoxia, their transcript abundance, population specific MHC class II-peptide binding prediction (ProPred), and a literature search. 26 candidate genes were selected by bioinformatic prediction and evaluated empirically using IFN-γ and IL-2 ELISpot using well-described immunodominant antigens (Acr-1, CFP-10, ESAT-6) as references. 23 of 26 proteins induced an IFN-γ response in PBMC of persons with active or latent tuberculosis; of that 10 proteins induced a dominant IFN-γ response in >50% of the donors. Five novel immunodominant proteins: Rv1957, Rv1954c, Rv1955, Rv2022c and Rv1471, showed responses similar to CFP-10 and ESAT-6 in both magnitude and frequency. Only moderate evidence of infection stage specific recognition of antigens was observed. IL-2 responses were of lower magnitude than that of IFN-γ and the antigenic repertoire far more focused. Reconciliation of bioinformatic and empirical hierarchies of immunodominance revealed that antigens could, to an extent, be predicted providing transcriptomic data was coupled to population specific MHC-peptide binding prediction. These findings revealed that a number of genes induced by hypoxia are in fact potent T-cell targets and therefore offers general support to the important role of hypoxia in the natural course of tuberculosis infection. The approach described for antigen discovery could be potentially be employed when predicting antigens from pathogens other than *M. tuberculosis*.

Introduction

Mycobacterium tuberculosis remains a major threat to global health with an estimated 10 million incident cases of tuberculosis (TB) that causes around 1.3 million deaths per year (WHO, 2010). Furthermore, one third of the world's population is considered latently infected: a reservoir from which active TB disease will continue to develop for the foreseeable future and thus forming a major obstacle to achieve global control. The existence of so many with latent infection highlights that the partially effective vaccine: *Mycobacterium bovis* Bacille Calmette Guérin (BCG) does not prevent such infection (Maartens and Wilkinson, 2007).

Studies of MTB genomic deletions (encoded by the region of difference 1, RD-1) led to the discovery of highly immunodominant, co-regulated, secreted, species specific proteins CFP-10 and ESAT-6 (Berthet et al., 1998; Sorensen et al., 1995). This use of genomic information has paved the way to the development of *in vitro* assays that measure T cell release of Interferon-gamma (IFN-γ) which have operational advantages and improve the specificity and possibly sensitivity of tuberculosis immunodiagnosis (Pai et al., 2008). An alternative genome based strategy is to relate what is highly expressed by bacilli *in vivo* or *in vitro* (and thereby potentially available as an antigen) (Sidders et al., 2008) to what is recognized *in vivo*.

Multiple *in vitro* and *in vivo* studies increasingly indicate adaptation of *M. tuberculosis* via oxygen limitation into a non-replicating persistent state (Rustad et al., 2009). These observations have led to the models of latency and reactivation that suggests a prominent role of oxygen status (Boshoff and Barry, 2005; Gomez and McKinney, 2004; Rao et al., 2008; Via et al., 2008). These *in vitro* culture models represent proxies of conditions that tubercle bacilli are thought to encounter *in vivo* during persistence in immunocompetent host. This is supported by the observations that the genes of DosR regulon, which are

expressed as a part of the initial adaptation to hypoxia in MTB, are up regulated *in vitro* in IFN-γ activated MTB infected murine macrophages (Schnappinger et al., 2003; Shi et al., 2003; Timm et al., 2003), have increased level of RNA in human lung tissue in patients with active TB (Timm et al., 2003), and are found to encode a number of MTB antigens with the potential to induce a strong T cell and IFN-γ responses (Black et al., 2009; Demissie et al., 2006; Leyten et al., 2006; Roupie et al., 2007; Wilkinson et al., 1998b).

The dosR regulon is also induced in response to nitric oxide, in standing cultures (which generates a hypoxic gradient), and following infection of macrophage, mice and guinea pigs (Sharma et al., 2006; Voskuil et al., 2003; Voskuil et al., 2004). Some of these conditions are marked by significant bacterial replication, suggesting that the role of dosR may not be specific to latency and that other factors may also be involved. This lead to further characterization of the hypoxic response in more depth revealing that the initial hypoxic conditions regulated by dosR contributes modestly to the survival under hypoxic conditions *in vitro* and it is dispensable for the virulence in mice (Rustad et al., 2008).

Recent whole genome based transcriptional analysis of MTB under similar in vitro hypoxic conditions, revealed that induction of the dosR regulon is somewhat transient, with expression of nearly half of the genes returning to baseline by 24 hours (1 day). An additional more prolonged transcriptional response to hypoxia was noted, which increased until 96 hours (4 days) and then stabilized around an enduring hypoxic response (EHR) at 168 hours (7 days) (Rustad et al., 2008).

Based on these findings 100 highly induce EHR genes were considered for evaluation. Since empirical screening of all the 100 genes is a considerable undertaking, a bioinformatic approach was used for selection of interesting candidates.

In this chapter describes evaluation of EHR encoded genes as potential antigens. I present the analysis of whole genome based transcriptional profiles of *M. tuberculosis* subject to prolonged hypoxia by a combined bioinformatic and immunological approach.

The hypothesise addressed in this chapter includes the following:

1. Molecules highly induced when MTB was subject to hypoxia *in vitro*, may be expressed *in vivo* and there by could be targets of the immune response.
2. These molecules might provide further understanding in the biology of latent infection and may be preferentially recognized by T cells during latent tuberculosis infection.
3. Genomic information of a pathogen might aid antigen prediction.

Results

Bioinformatic evaluation and selection of candidate antigens

Fold Expression and SigA normalized intensity in EHR

Whole genome based transcriptional data of MTB was analyzed, using fold induction (mean >2 log fold induction) and transcript levels (normalized to median *Sig*A intensity) during the hypoxic time course (4 hours, 1 day, 4 days and at 7 days) as described previously (Rustad et al., 2008) (Appendix B.1 and B.2). EHR genes were ranked based on the fold induction at 7 days of hypoxia (Table 10): 100 genes with highest fold induction levels were selected for further evaluation. A qualitative score was assigned to aid prediction of candidates. The genes were ranked, ranging from rank 1 (highest fold induction) to rank 100 (lowest) and referred as R(FI) (Table 10). The transcript levels at 4 and 7 days were also ranked from 1 to 100, and the average of ranks at 4 and 7 days were calculated; the average rank was further ranked and referred to as R(NI) (Table 10).

Table 10 *M. tuberculosis* transcript levels: Fold induction at day 7, Normalized intensity at day 4 and 7 with respective ranks

			Fold Induction		Normalised Intensity (NI)					
S.No	**Name**	**Gene name**	**7*day**	**Rank at 7 day (RFI)**	**96h***	**168h***	**Rank at 96hr**	**Rank at 168hrs**	**Average of ranks**	**Rank (RNI)**
1	Rv0140	Rv0140	3.180	47	1.26	5.10	46	13	29.	26
2	Rv0188	Rv0188	2.638	76	2.10	2.78	26	26	26	22
3	Rv0233	nrdB	2.492	86	1.21	1.76	51	51	51	50
4	Rv0251	hsp	4.117	16	11.8	13.2	3	3	3	2
5	Rv0268	Rv0268c	2.800	69	0.83	1.39	66	59	62.	67
6	Rv0327	cyp135A1	2.924	65	0.14	0.18	98	10	99	10
7	Rv0350	dnaK	2.335	96	5.58	4.17	4	17	10.	7
8	Rv0384	clpB	2.941	63	1.25	1.58	47	54	50.	49
9	Rv0474	Rv0474	3.390	33	4.37	4.84	7	15	11	9
10	Rv0520	Rv0520	3.296	38	0.48	0.63	83	84	83.	85
11	Rv0521	Rv0521	2.317	99	0.46	0.53	86	89	87.	88

12	Rv0754	PE_PGRS1	3.056	54	0.54	0.65	80	82	81	80
13	Rv0766	cyp123	3.040	55	1.32	1.07	41	66	53.	57
14	Rv0767	Rv0767c	4.019	19	3.49	2.63	10	28	19	18
15	Rv0791	Rv0791c	3.040	56	0.89	0.96	61	70	65.	71
16	Rv0793	Rv0793	2.730	72	0.47	0.44	84	92	88	89
17	Rv0826	Rv0826	3.477	31	3.43	3.07	11	25	18	16
18	Rv0846	Rv0846c	3.538	30	0.51	0.65	82	83	82.	83
19	Rv0847	lpqS	5.564	1	21.7	9.97	1	5	3	3
20	Rv0848	cysK2	4.040	17	2.68	1.82	18	48	33	29
21	Rv0849	Rv0849	2.592	83	0.46	0.46	87	91	89	90
22	Rv0940	Rv0940c	2.684	73	0.85	2.01	65	41	53	55
23	Rv0967	Rv0967	4.806	5	14.3	22.7	2	2	2	1
24	Rv0976	Rv0976c	3.105	51	0.65	0.52	73	90	81.	81
25	Rv0986	Rv0986	2.526	85	0.26	0.63	96	85	90.	92
26	Rv0990	Rv0990c	3.902	21	1.16	2.16	52	39	45.	41
27	Rv0991	Rv0991c	5.230	2	5.16	6.96	6	11	8.5	5
28	Rv1169	PE11	3.224	44	0.70	2.52	70	33	51.	51
29	Rv1221	sigE	2.771	70	3.98	26.1	8	1	4.5	4
30	Rv1284	Rv1284	3.107	50	0.58	0.62	77	86	81.	82
31	Rv1285	cysD	2.403	92	1.27	1.03	44	68	56	61
32	Rv1403	Rv1403c	3.984	20	0.92	1.34	60	62	61	64
33	Rv1405	Rv1405c	3.370	34	1.55	1.90	34	44	39	36
34	Rv1471	trxB1	4.556	8	1.85	2.18	31	38	34.	31
35	Rv1587	Rv1587c	2.313	100	2.49	2.05	20	40	30	27
36	Rv1738	Rv1738	4.987	4	0.54	1.40	79	58	68.	75
37	Rv1806	PE20	4.393	9	0.39	2.22	89	37	63	68
38	Rv1813	Rv1813c	4.029	18	2.02	1.38	30	60	45	40
39	Rv1875	Rv1875	3.745	24	0.86	9.15	64	6	35	32
40	Rv1909	furA	2.989	59	0.64	0.79	74	75	74.	78
41	Rv1954	Rv1954c	2.678	74	0.42	0.22	88	99	93.	94
42	Rv1955	Rv1955	4.383	10	5.36	3.84	5	19	12	12
43	Rv1956	Rv1956	3.115	49	2.34	2.53	22	32	27	23
44	Rv1957	Rv1957	3.542	28	1.04	1.80	56	49	52.	53
45	Rv1986	Rv1986	3.058	53	2.10	1.35	25	61	43	39
46	Rv1990	Rv1990c	3.151	48	1.25	1.75	48	52	50	48
47	Rv1994	Rv1994c	3.220	45	1.46	8.08	38	8	23	20
48	Rv1995	Rv1995	2.317	98	0.27	0.29	95	97	96	96
49	Rv1996	Rv1996	2.945	62	0.38	0.60	92	88	90	91
50	Rv2007	fdxA	2.623	81	0.65	1.15	72	64	68	74
51	Rv2011	Rv2011c	2.906	66	1.11	1.67	53	53	53	56
52	Rv2012	Rv2012	3.317	36	1.10	1.77	54	50	52	52
53	Rv2021	Rv2021c	2.483	87	1.29	1.17	42	63	52.	54
54	Rv2022	Rv2022c	2.948	61	0.74	0.85	69	73	71	76
55	Rv2025	Rv2025c	2.878	68	1.54	1.41	35	56	45.	42
56	Rv2030	Rv2030c	2.984	60	0.13	0.36	99	96	97.	99
57	Rv2031	hspX	4.217	14	0.63	1.40	75	57	66	72
58	Rv2032	acg	3.210	46	0.17	0.38	97	95	96	97
59	Rv2034	Rv2034	4.997	3	3.37	7.18	14	9	11.	11
60	Rv2035	Rv2035	3.828	23	2.24	3.68	23	20	21.	19
61	Rv2036	Rv2036	3.310	37	0.87	1.89	63	45	54	58
62	Rv2050	Rv2050	3.238	41	3.17	4.57	17	16	16.	15

63	Rv2465	rpiB	2.994	57	1.00	2.24	58	36	47	45
64	Rv2466	Rv2466c	4.153	15	3.29	10.8	16	4	10	6
65	Rv2504	scoA	2.669	75	0.75	3.29	68	24	46	44
66	Rv2517	Rv2517c	2.633	78	1.50	4.94	36	14	25	21
67	Rv2558	Rv2558	3.636	27	2.03	1.82	29	47	38	35
68	Rv2617	Rv2617c	2.482	88	1.23	3.63	50	21	35.	33
69	Rv2623	TB31.7	3.234	42	0.58	0.99	78	69	73.	77
70	Rv2626	Rv2626c	4.755	6	0.88	2.30	62	35	48.	46
71	Rv2627	Rv2627c	3.280	40	0.12	0.40	10	94	97	98
72	Rv2628	Rv2628	2.427	91	0.52	0.81	81	74	77.	79
73	Rv2642	Rv2642	4.741	7	3.42	7.14	12	10	11	10
74	Rv2643	arsC	3.729	25	0.60	2.38	76	34	55	59
75	Rv2658	Rv2658c	4.286	13	1.25	0.66	49	80	64.	70
76	Rv2659	Rv2659c	3.286	39	1.47	0.72	37	79	58	62
77	Rv2660	Rv2660c	2.471	89	0.33	0.28	93	98	95.	95
78	Rv2662	Rv2662	3.348	35	0.30	0.95	94	71	82.	84
79	Rv2663	Rv2663	2.901	67	0.76	2.57	67	31	49	47
80	Rv2664	Rv2664	2.762	71	0.69	1.56	71	55	63	69
81	Rv2693	Rv2693c	2.348	94	1.70	1.96	32	43	37.	34
82	Rv2694	Rv2694c	2.624	79	3.60	5.60	9	12	10.	8
83	Rv2699	Rv2699c	2.633	77	0.39	0.66	90	81	85.	86
84	Rv2780	ald	2.448	90	3.35	3.58	15	22	18.	17
85	Rv2913	Rv2913c	3.540	29	1.03	0.78	57	76	66.	73
86	Rv2962	Rv2962c	2.334	97	0.38	0.42	91	93	92	93
87	Rv2963	Rv2963	3.069	52	1.40	0.74	40	78	59	63
88	Rv3130	Rv3130c	3.712	26	2.07	2.65	27	27	27	24
89	Rv3131	Rv3131	2.368	93	0.47	0.62	85	87	86	87
90	Rv3269	Rv3269	2.930	64	1.44	0.90	39	72	55.	60
91	Rv3288	usfY	2.609	82	2.12	1.06	24	67	45.	43
92	Rv3290	lat	3.224	43	2.37	8.68	21	7	14	13
93	Rv3334	Rv3334	3.399	32	2.51	2.01	19	42	30.	28
94	Rv3406	Rv3406	4.310	11	1.56	1.85	33	46	39.	37
95	Rv3503	fdxD	2.552	84	1.26	0.75	45	77	61	65
96	Rv3515	fadD19	2.990	58	0.97	1.09	59	65	62	66
97	Rv3536	Rv3536c	3.887	22	1.28	3.40	43	23	33	30
98	Rv3597	lsr2	2.345	95	2.04	2.58	28	30	29	25
99	Rv3681	whiB4	2.623	80	1.06	2.58	55	29	42	38
10	Rv3862	whiB6	4.301	12	3.38	4.14	13	18	15.	14

* data obtained from (Rustad et al., 2008) (Appendix B.1 and B.2)

MHC class II- binding peptide prediction

Using ProPred (Singh and Raghava, 2001), MHC class II (MHCII)-binding peptide prediction was performed for the following HLA 51 allelic variants as described in the methods (Chapter 2): DR1 (2 allelic variants), DR3 (7), DR4 (9), DR 7 (6), DR8 (9), DR11(9), DR13 (11) and DR15(3) (Table 11). Instead of selecting only the predicted binding regions of the proteins, a rational approach was undertaken. A novel protocol was developed for scoring, which allowed ranking of the prediction, and prioritization of the gene products.

ProPred analysis defines the binding peptide sequences for all 51 HLA allelic variants and provides the binding score attained by that peptide sequence. In addition, it also provides the highest binding score that could be attained by any peptide to that allele. In the scoring protocol that was developed, the number of peptide regions with a score higher than 30% of the highest binding score attainable by any peptide for all the 51 HLA allelic variants was calculated. This score was then summed and corrected for the number of allelic variants of each allele, and referred as the MHC II binding peptide prediction score (Table 11).

Table 11 MHC II binding peptide prediction: number of peptide sequence with >30 % of highest binding score

		MHC class II binding peptide prediction score							
S.No	**Name**	**DR1**	**DR3**	**DR4**	**DR7**	**DR8**	**DR11**	**DR13**	**DR15**
1	Rv0140	0.00	2.00	1.90	2.00	0.60	0.20	1.18	1.30
2	Rv0188	3.00	2.20	3.20	5.00	3.50	3.20	4.07	4.90
3	Rv0233	0.50	1.64	2.00	5.00	1.30	1.90	2.47	2.60
4	Rv0251c	1.50	2.28	0.30	1.00	1.66	2.00	2.36	3.90
5	Rv0268c	0.50	1.40	0.00	1.00	1.50	1.60	2.17	1.60
6	Rv0327c	3.00	4.95	4.70	5.00	5.20	4.90	4.68	4.90
7	Rv0350	1.50	4.54	3.20	5.00	3.10	2.90	3.60	2.70
8	Rv0384c	4.00	4.94	5.36	5.00	4.30	4.00	4.13	4.90
9	Rv0474	0.00	2.57	0.10	2.00	1.00	0.76	0.16	1.60
10	Rv0520	0.00	1.70	0.30	3.50	1.46	0.60	0.63	2.00
11	Rv0521	0.50	1.70	2.40	5.00	2.10	3.40	3.69	3.60
12	Rv0754	2.50	4.90	4.10	5.00	4.10	2.80	3.00	2.80

13	Rv0766c	3.00	4.70	3.20	5.00	4.76	4.20	4.70	4.90
14	Rv0767c	2.50	3.08	1.80	5.00	1.80	2.84	3.45	1.09
15	Rv0791c	3.50	4.10	4.30	5.00	3.30	2.80	3.40	5.00
16	Rv0793	0.50	1.14	0.30	1.00	3.40	1.20	1.70	2.00
17	Rv0826	2.50	3.14	1.60	5.00	4.30	3.40	4.08	4.90
18	Rv0846c	1.00	4.78	4.50	5.00	5.80	3.30	4.60	4.20
19	Rv0847	0.50	1.84	1.80	3.00	1.60	0.90	1.50	2.30
20	Rv0848	0.50	4.28	4.10	5.00	3.32	3.30	3.56	1.07
21	Rv0849	1.50	2.70	2.70	6.00	3.60	3.60	3.63	5.00
22	Rv0940c	1.50	0.12	2.30	5.00	3.72	3.10	4.18	5.00
23	Rv0967	0.50	0.57	2.10	3.00	0.60	0.20	1.18	3.00
24	Rv0976c	3.00	4.99	6.10	5.00	4.46	3.60	2.80	5.00
25	Rv0986	3.00	4.40	3.20	5.00	1.66	2.50	2.98	4.60
26	Rv0990c	4.00	4.54	3.80	5.00	2.80	3.90	4.03	1.00
27	Rv0991c	0.00	1.14	1.30	2.00	0.90	0.50	0.27	0.60
28	Rv1169c	0.50	0.50	2.50	5.00	0.00	0.20	0.50	3.00
29	Rv1221	0.50	4.60	3.34	5.00	4.10	3.00	3.40	4.30
30	Rv1284	2.00	3.54	2.50	5.00	2.50	2.60	3.30	4.00
31	Rv1285	2.00	4.00	3.10	5.00	4.10	3.40	2.49	4.60
32	Rv1403c	3.00	3.55	2.10	4.00	0.60	1.70	2.09	1.00
33	Rv1405c	2.00	3.99	3.40	3.00	3.10	3.10	2.90	4.30
34	Rv1471	1.50	1.14	1.15	2.00	0.80	0.60	0.89	3.30
35	Rv1587c	0.50	1.70	2.80	5.00	2.30	3.40	3.69	3.60
36	Rv1738	0.00	0.14	0.70	3.00	1.30	0.30	1.09	1.30
37	Rv1806	4.00	1.70	2.60	8.00	1.60	1.00	1.08	4.00
38	Rv1813c	0.50	3.00	0.80	4.00	3.10	2.30	2.56	1.09
39	Rv1875	1.00	2.70	2.50	4.00	2.10	2.90	3.10	1.18
40	Rv1909c	1.50	2.80	1.40	2.00	1.83	1.00	1.80	1.30
41	Rv1954c	1.00	1.20	1.10	3.00	3.32	1.40	2.18	3.30
42	Rv1955	1.00	3.95	1.00	2.00	1.96	2.29	2.76	3.00
43	Rv1956	0.50	3.40	0.70	0.00	2.10	1.50	1.48	1.30
44	Rv1957	3.00	3.28	1.80	5.00	2.50	3.40	3.79	4.30
45	Rv1986	5.00	4.80	4.40	3.00	4.76	4.10	4.50	5.20
46	Rv1990c	0.00	2.54	1.70	4.00	0.83	0.60	1.72	1.00
47	Rv1994c	0.50	1.14	1.30	3.00	1.16	0.60	1.30	0.30
48	Rv1995	1.00	4.10	2.50	4.00	3.00	2.20	2.80	4.30
49	Rv1996	3.00	4.80	3.90	5.00	2.10	2.00	2.60	3.00
50	Rv2007c	0.50	2.60	0.10	2.00	0.00	0.30	0.00	0.00
51	Rv2011c	0.00	4.43	1.74	2.00	2.50	1.44	3.55	2.33
52	Rv2012	0.00	2.14	0.70	0.00	3.80	2.90	3.17	2.20
53	Rv2021c	0.00	3.24	1.70	5.00	2.46	2.50	2.68	4.00
54	Rv2022c	1.50	1.70	0.70	3.00	1.30	1.30	1.81	1.30
55	Rv2025c	5.00	5.60	4.60	5.00	4.90	4.70	3.89	4.90
56	Rv2030c	4.00	4.70	4.20	5.00	4.90	4.00	4.40	3.96
57	Rv2031c	1.00	1.90	1.10	3.00	0.80	1.40	1.17	4.30
58	Rv2032	2.00	4.98	3.70	5.00	3.60	3.20	3.18	4.30
59	Rv2034	1.00	2.20	0.80	0.00	1.30	0.60	0.70	0.60
60	Rv2035	1.00	1.00	0.80	1.00	1.60	0.80	1.60	0.72
61	Rv2036	0.00	1.28	0.90	3.00	4.76	2.90	3.90	4.60
62	Rv2050	0.00	1.70	0.30	1.00	0.60	0.30	0.09	1.00
63	Rv2465c	0.00	1.50	0.40	2.00	1.80	0.10	0.60	2.00

64	Rv2466c	0.00	1.57	1.80	4.50	2.16	2.16	2.45	2.20
65	Rv2504c	1.00	2.80	2.20	5.00	2.16	1.50	2.56	4.30
66	Rv2517c	0.00	0.00	0.60	3.00	0.60	0.80	0.09	2.00
67	Rv2558	1.50	2.80	1.70	5.00	1.70	1.40	1.79	3.30
68	Rv2617c	4.00	1.78	3.10	5.00	3.16	2.10	3.36	4.90
69	Rv2623	2.00	4.57	2.50	5.00	2.00	1.10	1.79	3.30
70	Rv2626c	1.50	2.28	2.60	2.00	2.46	2.60	3.20	1.60
71	Rv2627c	3.50	3.28	2.80	5.00	3.90	3.90	4.60	4.60
72	Rv2628	0.00	0.14	0.40	5.00	0.76	0.60	1.69	1.00
73	Rv2642	1.00	2.00	2.00	1.00	2.20	2.00	3.80	0.80
74	Rv2643	3.50	3.14	3.30	0.00	3.60	4.60	5.36	1.45
75	Rv2658c	0.50	1.57	1.20	1.00	0.50	0.20	0.80	0.60
76	Rv2659c	1.50	2.57	3.10	5.00	3.66	3.20	3.08	4.00
77	Rv2660c	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00
78	Rv2662	0.50	0.42	0.90	0.00	0.76	0.50	0.80	1.30
79	Rv2663	0.00	1.20	0.40	1.00	2.10	1.00	1.45	1.30
80	Rv2664	0.00	0.94	4.40	0.00	0.30	0.70	0.76	1.00
81	Rv2693c	3.00	5.20	6.50	5.00	5.00	4.80	5.93	5.00
82	Rv2694c	1.00	2.00	1.70	2.00	3.00	2.93	3.50	1.20
83	Rv2699c	0.00	1.27	0.40	3.00	3.80	1.20	2.07	1.60
84	Rv2780	3.50	5.10	3.70	7.00	5.20	4.30	5.40	6.30
85	Rv2913c	5.50	4.40	3.60	8.00	2.76	7.40	4.88	9.30
86	Rv2962c	3.50	5.00	4.14	4.50	4.90	5.40	5.39	4.20
87	Rv2963	1.50	4.28	3.40	1.00	4.30	2.10	2.58	1.20
88	Rv3130c	4.00	5.30	5.34	6.50	4.80	2.90	3.62	4.90
89	Rv3131	3.00	4.04	4.60	5.00	3.33	4.71	3.45	4.00
90	Rv3269	0.00	0.43	0.77	1.00	3.17	1.28	1.64	0.67
91	Rv3288c	2.00	2.40	2.00	5.00	1.32	2.30	2.85	5.00
92	Rv3290c	1.00	3.70	2.70	2.00	3.43	4.50	4.66	7.30
93	Rv3334	1.00	1.70	2.10	6.00	2.96	1.20	1.20	3.60
94	Rv3406	1.50	4.50	3.90	5.00	3.80	4.40	4.24	4.30
95	Rv3503c	1.00	1.80	0.10	1.00	0.50	0.40	0.09	2.60
96	Rv3515c	2.50	0.43	3.40	4.00	4.96	4.20	4.35	3.00
97	Rv3536c	1.00	3.27	1.80	5.00	2.66	1.80	2.18	0.90
98	Rv3597c	0.00	0.85	0.30	0.00	0.96	0.50	1.18	0.00
99	Rv3681c	0.50	0.28	0.00	0.00	1.60	0.50	1.18	0.00
100	Rv3862c	0.50	0.70	1.70	3.00	0.33	0.80	1.18	1.60

[Represented as the average number of peptide sequence with >30% of the highest peptide binding score achievable for each allele]

HLA allelic polymorphism in the human population, are known to influence the peptide recognition by MHC II. To account for this, the MHCII binding peptide prediction score was complemented by a correction for allelic frequency *(af*) of the South African Xhosa

population (in which the empirical evaluation was to be carried out), available from the literature (Appendix B.3)(du Toit et al., 1988). The correction was performed, by multiplying the prediction score for each allele (eg: score for DR1= 2) with its allelic frequency (*af*) (eg: *af* of DR1= 0.0839) within the population. The resulting score is referred as the population specific (corrected) MHCII binding peptide as presented in the Table 12.

In order to rank and prioritize gene products: minimum, median and maximum score (of all alleles) attained by each protein was calculated. The median score was then ranked (highest to lowest) and assigned rank numbers 1-100. This population specific MHCII binding prediction rank was referred as R(cPP) (Table 12).

Table 12 MHC II binding peptide prediction corrected to *af* in Xhosa population

S.No	Name	DR1	DR3	DR4	DR7	DR8	DR11	DR13	DR15	Minimum	Maximum	Median	Rank At prediction. R(cPP)
1	Rv0140	0.0000	0.8386	0.2537	0.2858	0.0074	0.0627	0.3885	0.2826	0.0000	0.8386	0.2681	71
2	Rv0188	0.2517	0.9225	0.4272	0.7145	0.0434	1.0038	1.3398	1.0653	0.0434	1.3398	0.8185	19
3	Rv0233	0.0420	0.6877	0.2670	0.7145	0.0161	0.5960	0.8131	0.5652	0.0161	0.8131	0.5806	42
4	Rv0251c	0.1259	0.9560	0.0401	0.1429	0.0206	0.6274	0.7769	0.8479	0.0206	0.9560	0.3852	58
5	Rv0268c	0.0420	0.5870	0.0000	0.1429	0.0186	0.5019	0.7144	0.3478	0.0000	0.7144	0.2454	74
6	Rv0327c	0.2517	2.0755	0.6275	0.7145	0.0645	1.5371	1.5407	1.0653	0.0645	2.0755	0.8899	10
7	Rv0350	0.1259	1.9036	0.4272	0.7145	0.0384	0.9097	1.1851	0.5870	0.0384	1.9036	0.6507	36
8	Rv0384c	0.3356	2.0713	0.7156	0.7145	0.0533	1.2548	1.3596	1.0653	0.0533	2.0713	0.8904	6
9	Rv0474	0.0000	1.0776	0.0134	0.2858	0.0124	0.2384	0.0533	0.3478	0.0000	1.0776	0.1459	85
10	Rv0520	0.0000	0.7128	0.0401	0.5002	0.0181	0.1882	0.2074	0.4348	0.0000	0.7128	0.1978	80
11	Rv0521	0.0420	0.7128	0.3204	0.7145	0.0260	1.0666	1.2147	0.7826	0.0260	1.2147	0.7137	31
12	Rv0754	0.2098	2.0546	0.5474	0.7145	0.0508	0.8784	0.9876	0.6087	0.0508	2.0546	0.6616	35
13	Rv0766c	0.2517	1.9707	0.4272	0.7145	0.0590	1.3175	1.5472	1.0653	0.0590	1.9707	0.8899	9
14	Rv0767c	0.2098	1.2914	0.2403	0.7145	0.0223	0.8909	1.1357	0.2370	0.0223	1.2914	0.4774	48
15	Rv0791c	0.2937	1.7191	0.5741	0.7145	0.0409	0.8784	1.1193	1.0870	0.0409	1.7191	0.7964	21
16	Rv0793	0.0420	0.4780	0.0401	0.1429	0.0422	0.3764	0.5596	0.4348	0.0401	0.5596	0.2597	73
17	Rv0826	0.2098	1.3166	0.2136	0.7145	0.0533	1.0666	1.3431	1.0653	0.0533	1.3431	0.8899	7
18	Rv0846c	0.0839	2.0043	0.6008	0.7145	0.0719	1.0352	1.5143	0.9131	0.0719	2.0043	0.8138	20

19	Rv0847	0.0420	0.7715	0.2403	0.4287	0.0198	0.2823	0.4938	0.5000	0.0198	0.7715	0.3555	63
20	Rv0848	0.0420	1.7946	0.5474	0.7145	0.0412	1.0352	1.1720	0.2326	0.0412	1.7946	0.6309	38
21	Rv0849	0.1259	1.1321	0.3605	0.8574	0.0446	1.1293	1.1950	1.0870	0.0446	1.1950	0.9722	4
22	Rv0940c	0.1259	0.0503	0.3071	0.7145	0.0461	0.9725	1.3761	1.0870	0.0461	1.3761	0.5108	46
23	Rv0967	0.0420	0.2390	0.2804	0.4287	0.0074	0.0627	0.3885	0.6522	0.0074	0.6522	0.2597	72
24	Rv0976c	0.2517	2.0923	0.8144	0.7145	0.0553	1.1293	0.9218	1.0870	0.0553	2.0923	0.8681	12
25	Rv0986	0.2517	1.8449	0.4272	0.7145	0.0206	0.7843	0.9810	1.0000	0.0206	1.8449	0.7494	29
26	Rv0990c	0.3356	1.9036	0.5073	0.7145	0.0347	1.2234	1.3267	0.2174	0.0347	1.9036	0.6109	40
27	Rv0991c	0.0000	0.4780	0.1736	0.2858	0.0112	0.1569	0.0889	0.1304	0.0000	0.4780	0.1436	87
28	Rv1169c	0.0420	0.2097	0.3338	0.7145	0.0000	0.0627	0.1646	0.6522	0.0000	0.7145	0.1871	82
29	Rv1221	0.0420	1.9288	0.4459	0.7145	0.0508	0.9411	1.1193	0.9348	0.0420	1.9288	0.8247	17
30	Rv1284	0.1678	1.4843	0.3338	0.7145	0.0310	0.8156	1.0864	0.8696	0.0310	1.4843	0.7651	27
31	Rv1285	0.1678	1.6772	0.4139	0.7145	0.0508	1.0666	0.8197	1.0000	0.0508	1.6772	0.7671	26
32	Rv1403c	0.2517	1.4885	0.2804	0.5716	0.0074	0.5333	0.6874	0.2174	0.0074	1.4885	0.4068	54
33	Rv1405c	0.1678	1.6730	0.4539	0.4287	0.0384	0.9725	0.9547	0.9348	0.0384	1.6730	0.6944	33
34	Rv1471	0.1259	0.4780	0.1535	0.2858	0.0099	0.1882	0.2930	0.7174	0.0099	0.7174	0.2370	75
35	Rv1587c	0.0420	0.7128	0.3738	0.7145	0.0285	1.0666	1.2147	0.7826	0.0285	1.2147	0.7137	32
36	Rv1738	0.0000	0.0587	0.0935	0.4287	0.0161	0.0941	0.3588	0.2826	0.0000	0.4287	0.0938	94
37	Rv1806	0.3356	0.7128	0.3471	1.1432	0.0198	0.3137	0.3555	0.8696	0.0198	1.1432	0.3513	64
38	Rv1813c	0.0420	1.2579	0.1068	0.5716	0.0384	0.7215	0.8428	0.2370	0.0384	1.2579	0.4043	55
39	Rv1875	0.0839	1.1321	0.3338	0.5716	0.0260	0.9097	1.0205	0.2565	0.0260	1.1321	0.4527	51
40	Rv1909c	0.1259	1.1740	0.1869	0.2858	0.0227	0.3137	0.5926	0.2826	0.0227	1.1740	0.2842	68
41	Rv1954c	0.0839	0.5032	0.1469	0.4287	0.0412	0.4392	0.7177	0.7174	0.0412	0.7177	0.4339	52
42	Rv1955	0.0839	1.6562	0.1335	0.2858	0.0243	0.7184	0.9086	0.6522	0.0243	1.6562	0.4690	49
43	Rv1956	0.0420	1.4256	0.0935	0.0000	0.0260	0.4706	0.4872	0.2826	0.0000	1.4256	0.1880	81
44	Rv1957	0.2517	1.3753	0.2403	0.7145	0.0310	1.0666	1.2477	0.9348	0.0310	1.3753	0.8247	16
45	Rv1986	0.4195	2.0126	0.5874	0.4287	0.0590	1.2862	1.4814	1.1305	0.0590	2.0126	0.8589	13
46	Rv1990c	0.0000	1.0650	0.2270	0.5716	0.0103	0.1882	0.5662	0.2174	0.0000	1.0650	0.2222	76
47	Rv1994c	0.0420	0.4780	0.1736	0.4287	0.0144	0.1882	0.4280	0.0652	0.0144	0.4780	0.1809	83
48	Rv1995	0.0839	1.7191	0.3338	0.5716	0.0372	0.6901	0.9218	0.9348	0.0372	1.7191	0.6309	39
49	Rv1996	0.2517	2.0126	0.5207	0.7145	0.0260	0.6274	0.8559	0.6522	0.0260	2.0126	0.6398	37
50	Rv2007c	0.0420	1.0902	0.0134	0.2858	0.0000	0.0941	0.0000	0.0000	0.0000	1.0902	0.0277	98
51	Rv2011c	0.0000	1.8569	0.2329	0.2858	0.0310	0.4531	1.1672	0.5073	0.0000	1.8569	0.3695	60
52	Rv2012	0.0000	0.8973	0.0935	0.0000	0.0471	0.9097	1.0436	0.4783	0.0000	1.0436	0.2859	67
53	Rv2021c	0.0000	1.3585	0.2270	0.7145	0.0305	0.7843	0.8823	0.8696	0.0000	1.3585	0.7494	28
54	Rv2022c	0.1259	0.7128	0.0935	0.4287	0.0161	0.4078	0.5959	0.2826	0.0161	0.7128	0.3452	66
55	Rv2025c	0.4195	2.3481	0.6141	0.7145	0.0608	1.4744	1.2806	1.0653	0.0608	2.3481	0.8899	8
56	Rv2030c	0.3356	1.9707	0.5607	0.7145	0.0608	1.2548	1.4485	0.8609	0.0608	1.9707	0.7877	24
57	Rv2031c	0.0839	0.7967	0.1469	0.4287	0.0099	0.4392	0.3852	0.9348	0.0099	0.9348	0.4069	53
58	Rv2032	0.1678	2.0881	0.4940	0.7145	0.0446	1.0038	1.0469	0.9348	0.0446	2.0881	0.8247	18
59	Rv2034	0.0839	0.9225	0.1068	0.0000	0.0161	0.1882	0.2304	0.1304	0.0000	0.9225	0.1186	92
60	Rv2035	0.0839	0.4193	0.1068	0.1429	0.0198	0.2510	0.5267	0.1565	0.0198	0.5267	0.1497	84

61	Rv2036	0.0000	0.5367	0.1202	0.4287	0.0590	0.9097	1.2839	1.0000	0.0000	1.2839	0.4827	47
62	Rv2050	0.0000	0.7128	0.0401	0.1429	0.0074	0.0941	0.0296	0.2174	0.0000	0.7128	0.0671	95
63	Rv2465c	0.0000	0.6290	0.0534	0.2858	0.0223	0.0314	0.1975	0.4348	0.0000	0.6290	0.1255	90
64	Rv2466c	0.0000	0.6583	0.2403	0.6431	0.0268	0.6776	0.8065	0.4783	0.0000	0.8065	0.5607	43
65	Rv2504c	0.0839	1.1740	0.2937	0.7145	0.0268	0.4706	0.8428	0.9348	0.0268	1.1740	0.5925	41
66	Rv2517c	0.0000	0.0000	0.0801	0.4287	0.0074	0.2510	0.0296	0.4348	0.0000	0.4348	0.0549	96
67	Rv2558	0.1259	1.1740	0.2270	0.7145	0.0211	0.4392	0.5893	0.7174	0.0211	1.1740	0.5142	44
68	Rv2617c	0.3356	0.7464	0.4139	0.7145	0.0392	0.6588	1.1061	1.0653	0.0392	1.1061	0.6866	34
69	Rv2623	0.1678	1.9162	0.3338	0.7145	0.0248	0.3451	0.5893	0.7174	0.0248	1.9162	0.4672	50
70	Rv2626c	0.1259	0.9560	0.3471	0.2858	0.0305	0.8156	1.0534	0.3478	0.0305	1.0534	0.3475	65
71	Rv2627c	0.2937	1.3753	0.3738	0.7145	0.0484	1.2234	1.5143	1.0000	0.0484	1.5143	0.8573	14
72	Rv2628	0.0000	0.0587	0.0534	0.7145	0.0094	0.1882	0.5563	0.2174	0.0000	0.7145	0.1235	91
73	Rv2642	0.0839	0.8386	0.2670	0.1429	0.0273	0.6274	1.2510	0.1739	0.0273	1.2510	0.2205	77
74	Rv2643	0.2937	1.3166	0.4406	0.0000	0.0446	1.4430	1.7645	0.3152	0.0000	1.7645	0.3779	59
75	Rv2658c	0.0420	0.6583	0.1602	0.1429	0.0062	0.0627	0.2634	0.1304	0.0062	0.6583	0.1367	89
76	Rv2659c	0.1259	1.0776	0.4139	0.7145	0.0454	1.0038	1.0139	0.8696	0.0454	1.0776	0.7921	23
77	Rv2660c	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0593	0.0000	0.0000	0.0593	0.0000	100
78	Rv2662	0.0420	0.1761	0.1202	0.0000	0.0094	0.1569	0.2634	0.2826	0.0000	0.2826	0.1385	88
79	Rv2663	0.0000	0.5032	0.0534	0.1429	0.0260	0.3137	0.4773	0.2826	0.0000	0.5032	0.2128	79
80	Rv2664	0.0000	0.3941	0.5874	0.0000	0.0037	0.2196	0.2502	0.2174	0.0000	0.5874	0.2185	78
81	Rv2693c	0.2517	2.1804	0.8678	0.7145	0.0620	1.5058	1.9519	1.0870	0.0620	2.1804	0.9774	3
82	Rv2694c	0.0839	0.8386	0.2270	0.2858	0.0372	0.9191	1.1522	0.2609	0.0372	1.1522	0.2733	69
83	Rv2699c	0.0000	0.5325	0.0534	0.4287	0.0471	0.3764	0.6814	0.3478	0.0000	0.6814	0.3621	61
84	Rv2780	0.2937	2.1384	0.4940	1.0003	0.0645	1.3489	1.7777	1.3696	0.0645	2.1384	1.1746	2
85	Rv2913c	0.4615	1.8449	0.4806	1.1432	0.0342	2.3214	1.6065	2.0218	0.0342	2.3214	1.3748	1
86	Rv2962c	0.2937	2.0965	0.5527	0.6431	0.0608	1.6940	1.7744	0.9131	0.0608	2.0965	0.7781	25
87	Rv2963	0.1259	1.7946	0.4539	0.1429	0.0533	0.6588	0.8493	0.2609	0.0533	1.7946	0.3574	62
88	Rv3130c	0.3356	2.2223	0.7129	0.9289	0.0595	0.9097	1.1917	1.0653	0.0595	2.2223	0.9193	5
89	Rv3131	0.2517	1.6952	0.6141	0.7145	0.0413	1.4779	1.1372	0.8696	0.0413	1.6952	0.7921	22
90	Rv3269	0.0000	0.1797	0.1024	0.1429	0.0393	0.4008	0.5387	0.1449	0.0000	0.5387	0.1439	86
91	Rv3288c	0.1678	1.0063	0.2670	0.7145	0.0164	0.7215	0.9382	1.0870	0.0164	1.0870	0.7180	30
92	Rv3290c	0.0839	1.5514	0.3605	0.2858	0.0425	1.4117	1.5341	1.5870	0.0425	1.5870	0.8861	11
93	Rv3334	0.0839	0.7128	0.2804	0.8574	0.0367	0.3764	0.3950	0.7826	0.0367	0.8574	0.3857	57
94	Rv3406	0.1259	1.8869	0.5207	0.7145	0.0471	1.3803	1.3958	0.9348	0.0471	1.8869	0.8247	15
95	Rv3503c	0.0839	0.7547	0.0134	0.1429	0.0062	0.1255	0.0296	0.5652	0.0062	0.7547	0.1047	93
96	Rv3515c	0.2098	0.1803	0.4539	0.5716	0.0615	1.3175	1.4320	0.6522	0.0615	1.4320	0.5128	45
97	Rv3536c	0.0839	1.3711	0.2403	0.7145	0.0330	0.5647	0.7177	0.1957	0.0330	1.3711	0.4025	56
98	Rv3597c	0.0000	0.3564	0.0401	0.0000	0.0119	0.1569	0.3885	0.0000	0.0000	0.3885	0.0260	99
99	Rv3681c	0.0420	0.1174	0.0000	0.0000	0.0198	0.1569	0.3885	0.0000	0.0000	0.3885	0.0309	97
100	Rv3862c	0.0420	0.2935	0.2270	0.4287	0.0041	0.2510	0.3885	0.3478	0.0041	0.4287	0.2722	70

Final Ranking for informatics prediction

The overall qualitative score was calculated by summing the ranks: R(FI), R(NI) and R(cPP). The summed score was in turn ranked (R) from the lowest to highest and the gene with the lowest score was assigned the highest rank (rank 1) for being the best candidate and the highest score with the lowest rank (rank 100), as the worst candidate (Table 13).

Table 13 Overall score and ranking

S.No	Name	Gene name	R(FI)	R(cPP)	R(NI)	Over all Score	R
1	Rv0140	Rv0140	47	71	26	144	46
2	**Rv0188**	**Rv0188**	**76**	**19**	**22**	**117**	**27**
3	Rv0233	nrdB	86	42	50	178	73
4	Rv0251c	hsp	16	58	3	77	7
5	Rv0268c	Rv0268c	69	74	67	210	87
6	Rv0327c	cyp135A1	65	7	100	172	67
7	Rv0350	dnaK	96	36	7	139	40
8	Rv0384c	clpB	63	6	49	118	30
9	Rv0474	Rv0474	33	85	10	128	36
10	Rv0520	Rv0520	38	80	85	203	82
11	Rv0521	Rv0521	99	31	88	218	91
12	Rv0754	PE_PGRS11	54	35	80	169	63
13	Rv0766c	cyp123	55	8	57	120	33
14	**Rv0767c**	**Rv0767c**	**19**	**48**	**18**	**85**	**10**
15	Rv0791c	Rv0791c	56	21	71	148	49
16	Rv0793	Rv0793	72	73	89	234	95
17	**Rv0826**	**Rv0826**	**31**	**9**	**16**	**56**	**1**
18	Rv0846c	Rv0846c	30	20	83	133	38
19	**Rv0847**	**lpqS**	**1**	**63**	**2**	**66**	**3**
20	Rv0848	cysK2	17	38	29	84	9
21	**Rv0849**	**Rv0849**	**83**	**4**	**90**	**177**	**72**
22	Rv0940c	Rv0940c	73	46	55	174	69
23	**Rv0967**	**Rv0967**	**5**	**72**	**1**	**78**	**8**
24	Rv0976c	Rv0976c	51	12	81	144	47
25	Rv0986	Rv0986	85	28	92	205	84
26	**Rv0990c**	**Rv0990c**	**21**	**40**	**41**	**102**	**16**
27	**Rv0991c**	**Rv0991c**	**2**	**87**	**5**	**94**	**13**
28	Rv1169c	PE11	44	82	51	177	70
29	Rv1221	sigE	70	15	4	89	11
30	**Rv1284**	**Rv1284**	**50**	**27**	**82**	**159**	**58**
31	Rv1285	cysD	92	26	61	179	74
32	Rv1403c	Rv1403c	20	54	64	138	39
33	Rv1405c	Rv1405c	34	33	36	103	17
34	**Rv1471**	**trxB1**	**8**	**75**	**31**	**114**	**26**
35	Rv1587c	Rv1587c	100	32	27	159	57
36	Rv1738	Rv1738	4	94	75	173	68
37	Rv1806	PE20	9	64	68	141	43
38	Rv1813c	Rv1813c	18	55	40	113	25
39	Rv1875	Rv1875	24	51	32	107	22
40	Rv1909c	furA	59	68	78	205	83

41	**Rv1954c**	**Rv1954c**	**74**	**52**	**94**	**220**	**93**
42	**Rv1955**	**Rv1955**	**10**	**49**	**12**	**71**	**6**
43	**Rv1956**	**Rv1956**	**49**	**81**	**24**	**154**	**52**
44	**Rv1957**	**Rv1957**	**28**	**16**	**53**	**97**	**15**
45	**Rv1986**	**Rv1986**	**53**	**13**	**39**	**105**	**19**
46	Rv1990c	Rv1990c	48	76	48	172	65
47	Rv1994c	Rv1994c	45	83	20	148	48
48	Rv1995	Rv1995	98	39	97	234	96
49	Rv1996	Rv1996	62	37	91	190	77
50	Rv2007c	fdxA	80	98	74	252	98
51	Rv2011c	Rv2011c	66	60	56	182	75
52	Rv2012	Rv2012	36	67	52	155	55
53	**Rv2021c**	**Rv2021c**	**87**	**29**	**54**	**170**	**64**
54	**Rv2022c**	**Rv2022c**	**61**	**66**	**76**	**203**	**81**
55	Rv2025c	Rv2025c	68	10	42	120	32
56	Rv2030c	Rv2030c	60	24	99	183	76
57	**Rv2031c**	**hspX**	**14**	**53**	**72**	**139**	**42**
58	Rv2032	acg	46	17	96	159	59
59	Rv2034	Rv2034	3	92	11	106	20
60	Rv2035	Rv2035	23	84	19	126	35
61	Rv2036	Rv2036	37	47	58	142	44
62	Rv2050	Rv2050	41	95	15	151	50
63	Rv2465c	rpiB	57	90	45	192	78
64	Rv2466c	Rv2466c	15	43	6	64	2
65	Rv2504c	scoA	75	41	44	160	60
66	Rv2517c	Rv2517c	78	96	21	195	80
67	Rv2558	Rv2558	27	44	35	106	21
68	Rv2617c	Rv2617c	88	34	33	155	53
69	Rv2623	TB31.7	42	50	77	169	62
70	Rv2626c	Rv2626c	6	65	46	117	29
71	Rv2627c	Rv2627c	40	14	98	152	51
72	Rv2628	Rv2628	91	91	79	261	99
73	Rv2642	Rv2642	7	77	9	93	12
74	Rv2643	arsC	25	59	59	143	45
75	**Rv2658c**	**Rv2658c**	**13**	**89**	**70**	**172**	**66**
76	**Rv2659c**	**Rv2659c**	**39**	**22**	**62**	**123**	**34**
77	**Rv2660c**	**Rv2660c**	**89**	**100**	**95**	**284**	**100**
78	**Rv2662**	**Rv2662**	**35**	**88**	**84**	**207**	**85**
79	**Rv2663**	**Rv2663**	**67**	**79**	**47**	**193**	**79**
80	**Rv2664**	**Rv2664**	**71**	**78**	**69**	**218**	**90**
81	**Rv2693c**	**Rv2693c**	**94**	**3**	**34**	**131**	**37**
82	Rv2694c	Rv2694c	79	69	8	156	56
83	Rv2699c	Rv2699c	77	61	86	224	94
84	**Rv2780**	**ald**	**90**	**2**	**17**	**109**	**24**
85	Rv2913c	Rv2913c	29	1	73	103	18
86	Rv2962c	Rv2962c	97	25	93	215	88
87	Rv2963	Rv2963	52	62	63	177	71
88	Rv3130c	Rv3130c	26	5	87	118	31
89	Rv3131	Rv3131	93	23	23	139	41
90	Rv3269	Rv3269	64	86	60	210	86
91	Rv3288c	usfY	82	30	43	155	54
92	Rv3290c	lat	43	11	13	67	5
93	**Rv3334**	**Rv3334**	**32**	**57**	**28**	**117**	**28**
94	**Rv3406**	**Rv3406**	**11**	**18**	**37**	**66**	**4**
95	Rv3503c	fdxD	84	93	65	242	97
96	Rv3515c	fadD19	58	45	66	169	61
97	Rv3536c	Rv3536c	22	56	30	108	23

98	Rv3597c	lsr2	95	99	25	219	92
99	Rv3681c	whiB4	81	97	38	216	89
100	Rv3862c	whiB6	12	70	14	96	14

*selected proteins are in bold

Selection of candidates for empirical evaluation

I selected genes that were:

(i) Species specific: especially if they are absent from *M. bovis* and the commonly used BCG vaccine strains (described and presented in Chapter 6)

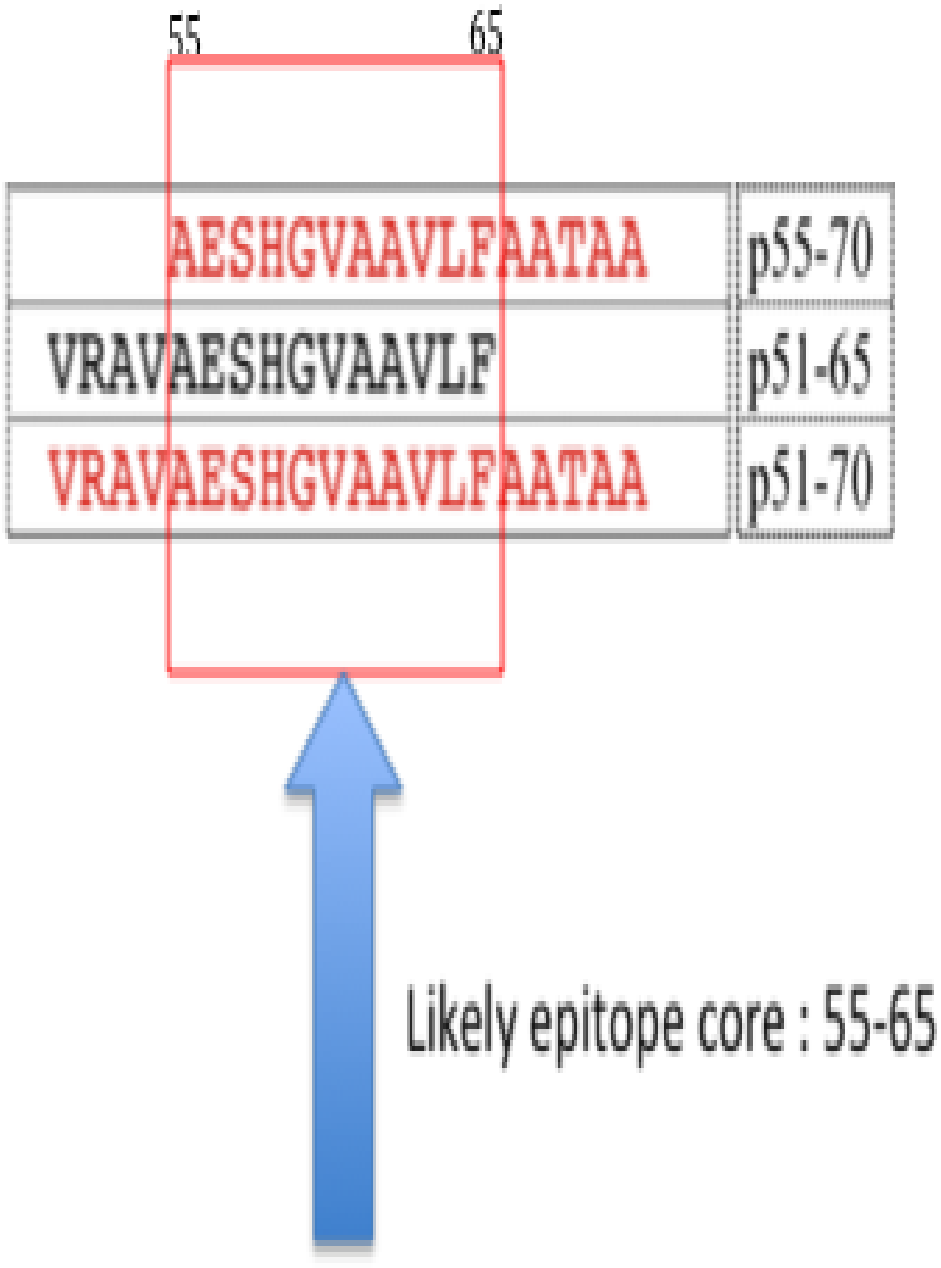
(ii) Based on the overall prediction score (R) (Table 14): both high and low ranking candidates were included in order to aid relating bioinformatic and empirical hierarchies.

(iii) Genomic organisation (e.g. predicted operons)

(iv) Genes that are also induced in other stress conditions, such as nutrition starvation and heat shock.

Analysis of transcript levels of bacilli exposed to 7 days of hypoxia

The fold induction and *sigA* normalized transcript intensity over a time course of 168 hrs hypoxia for the selected EHR genes (and of Acr1, CFP-10 and ESAT-6) were analyzed. The fold induction and the *sig*A normalized transcript intensity for the EHR and dosR regulated Acr1 genes increased over a time course of 168 hours (Figure 33 A and B).



Interestingly, whilst the fold induction (Figure 33 A) for the RD1 encoded genes fell, the normalized intensity remained at a similar absolute level to that of both EHR and the dosR regulated Acr1 gene (Figure 33 B).

Figure 33 was generated from the data obtained from (Rustad et al., 2008) (Appendix B.1 and B.2)

Figure 33 *M. tuberculosis* transcript levels from bacilli exposed to 7 days of hypoxia

Figure 33 *M. tuberculosis* transcript levels from bacilli exposed to 7 days of hypoxia

Panel A Shows the fold induction over hypoxic time course with reference to phase of aerobic cultures. There is a relative increase in the transcripts of EHR genes and Acr1 (Rv2031c) and a fall for those of CFP-10 (Rv3874) and ESAT-6 (Rv3875).
Panel B Shows the median transcript intensity normalized to *Sig*A. There is a very rapid peak in Acr1 levels followed by a decline; a sustained increase in EHR genes; and a fall in CFP-10 and ESAT-6. However overall abundance is similar at 168 hours for CFP-10 and ESAT-6 as that of other EHR genes.

Evaluation of regions of similarity between biological sequences

The nucleotide and protein sequence of selected EHR candidate genes (n=26) were further analysed for regions of similarity with sequences in the nucleotide and protein databases (and with reference to TB genome). Similarity was represented as the percentage identity with respect to MTB H37Rv strain. Results are summarised as the percentage similarity (≥50%) with the H37Rv sequence as the reference in Table14 (protein) and Table 15 (nucleotide). Bacterial strains presented in the table can be largely classified as strains of tuberculous and non-tuberculous mycobacteria and environmental bacteria that are phylogenically related to mycobacteria. When analysed by the nucleotide and protein

sequence, 18 and 11 EHR gene candidates shared regions of similarity only within the *M. tuberculosis* complex (i.e., MTB and *M. bovis* including BCG) respectively.

Table 14 BLASTp analysis of EHR genes (n=26) using protein sequence with Protein database. Represented as percentage similarity to MTB H37Rv

	Rv0188	Rv0767c	Rv0826	Rv0847	Rv0849	Rv0967	Rv0990c	Rv0991c	Rv1284	Rv1471	Rv1954c	Rv1955	Rv1956	Rv1957	Rv2021c	Rv2022c	Rv2466c	Rv2517c	Rv2660c	Rv2662	Rv2663	Rv2664	Rv2693c	Rv2780	Rv3334	Rv3406
M. tuberculosis H37Ra	100	100	100	100	100	100	100	100	100	100		100	100	100	100	100		100	100	100	100	100	100	100	100	100
M. tuberculosis CDC1551	100	100	100	99		100	100	99	100	100		100	100	99	100	100		99	100		100	100	100	100	100	100
M. tuberculosis F11	100	100	100	99	100	100	100	100	100	100		100	100	100	100	100	100		100		100	100	100	100	100	100
M. tuberculosis C	100	100	100	99	100	100	100	99	100			100	100	100	100	60	97		100		100	100	100	100	100	100
*M. tuberculosis Haarlem (*draft)	100	100	100	99		100	100	100	100	100		100	100	99	100	100	100		100	100	100	100	100	100	100	100
M. bovis AF2122/97	100	100	99	99	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	64	100	100
M. bovis BCG	100	100	100	99	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	64	100	100
M. leprae TN								87		57							80							85		
M. avium 104		74	84				73	80	80	71							82						71	83		78
M. avium k10	100	74	84				73	80	80	71							82	66					71	82		
M. abscessus		60	68		56		57	68	66	62							77						54	78		52
M. gilvum PYR-GCK		62	66		62		63	71	77	63							77						65	78		56
M. marinum		74	87	63	75	80	72	83	84	72							86						81	83		84
M. smegmatis MC2 155		70	72				63	72	79	68							78						61	81		60
M. sp. KMS		66	70		64		62	75	78	73							81						67	80		57
M. sp. MCS		66	66		64		62	75	78	73							81						67	80		57
M. ulcerans Agy99			87	62	73	79	70	83	85														80	83		84
M. vanbaalenii PYR-1		60	69		63		61	72	78	70							75						66	79		
C. diphtheriae NCTC 13129										53							57									
C. efficiens YS-314																	52									
C. glutamicum ATCC 13032																	55									
C. jeikeium K411																	52									
R. sp. RHA1			61					63	77	60							72						53	68		57
R. sphaeroides 2.4.1																								57		
S. avermitilis MA-4680								58	52								57							68		
S. coelicolor A3(2)								58	53								55							66		
N. farcinica IFM 10152								68		63							74							62		54
P. acnes KPA171202								55																60		
A. cellulolyticus 11B								60									56							60		

Table 15 BLASTn analysis of EHR genes (n=26) using nucleotide sequence with nucleotide database. Represented as percentage similarity to MTB H37Rv

	Rv0188	Rv0767c	Rv0826	Rv0847	Rv0849	Rv0967	Rv0990c	Rv0991c	Rv1284	Rv1471	Rv1954c	Rv1955	Rv1956	Rv1957	Rv2021c	Rv2022c	Rv2466c	Rv2517c	Rv2660c	Rv2662	Rv2663	Rv2664	Rv2693c	Rv2780	Rv3334	Rv3406
M. tuberculosis H37Ra	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
M. tuberculosis CDC1551	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
M. tuberculosis F11	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
M. tuberculosis KZN 1435	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100							100	100
M. tuberculosis sequence from clone y223																			100	100	100	100				
M. bovis subsp. bovis AF2122/97	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	96	100	100	100	100	100	100	100
M. bovis BCG Pasteur 1173P2	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	96	100	100	100	100	100	100	100	
M. bovis BCG str. Tokyo 172 DNA	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	96	100	100	100	100	100	100	100
M. leprae Br 4923																								100		
M. leprae strain TN																								100		
M. avium 104																	95						89			99
M. avium subsp. paratuberculosis str. K10																	95						89			
M. gilvum PYR-GCK		82																								
M. kansasii strain																										
M. marinum M																	99							99		98
M. riyadhense strain																										
M. smegmatis str. MC2 155	100									81																
M. sp KMS										59																
M. sp. MCS										59																
M. sp.JLS										59																
M. ulcerans Agy99																	99		99					98		83
M. vanbaalenii PYR-1		83			53												92									
N. farcinica IFM 10152 DNA					82																					
Expression vector mce1	100																									

Functional Categories of Enduring hypoxic genes

The genes with the highest fold induction during hypoxia time course were assigned a functional category. As previously described (Rustad et al., 2008), the 100 highly induced genes featured a larger percentage of putative regulatory and a high proportion of genes of unknown function when compared with the whole genome of *Mycobacterium tuberculosis*. Table 16 describes the percentage of genes according to the functional category. Additional properties of these candidate genes are summarized in Appendix B.4. Figure 34 presents a graphical representation of percentage gene distribution according to the functional category, in the repertoire of EHR (Panel B, C, D) in relation to the whole genome (Panel A) of MTB.

Figure 34: Distribution in MTB according to functional category

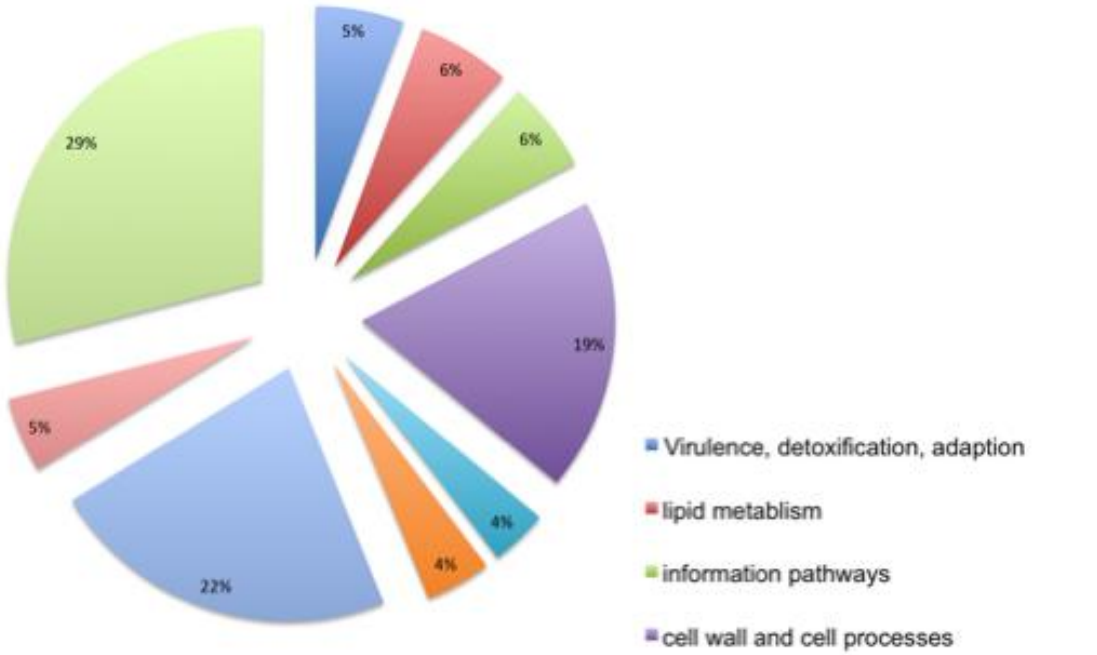
Shows the percentage distribution of genes according to the functional categories in whole genome of MTB and in various repertoire of EHR.

Panel A represents the whole genome of MTB; Panel B represents the described 230 EHR genes (Rustad et al., 2008); Panel C represents highly induced 100 EHR genes, that were selected for the study and Panel D represents 30 EHR genes (includes Acr1) selected for empirical evaluation.

The functional categories were obtained for the whole genome (panel A) and the 100 highly induced EHR genes (Panel C and including panel D) from the TubercuList. For the complete EHR list (230 genes) (panel B), the functional categories were adapted from (Rustad et al., 2008).

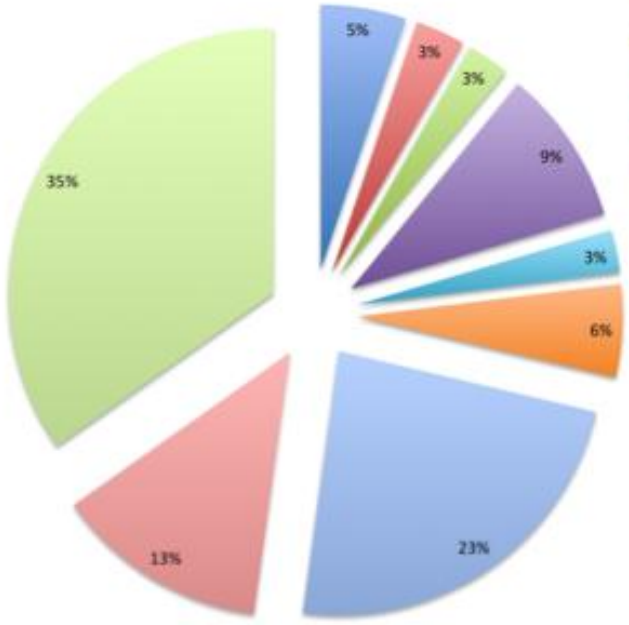
A

Whole genome



B

EHR complete list (230)



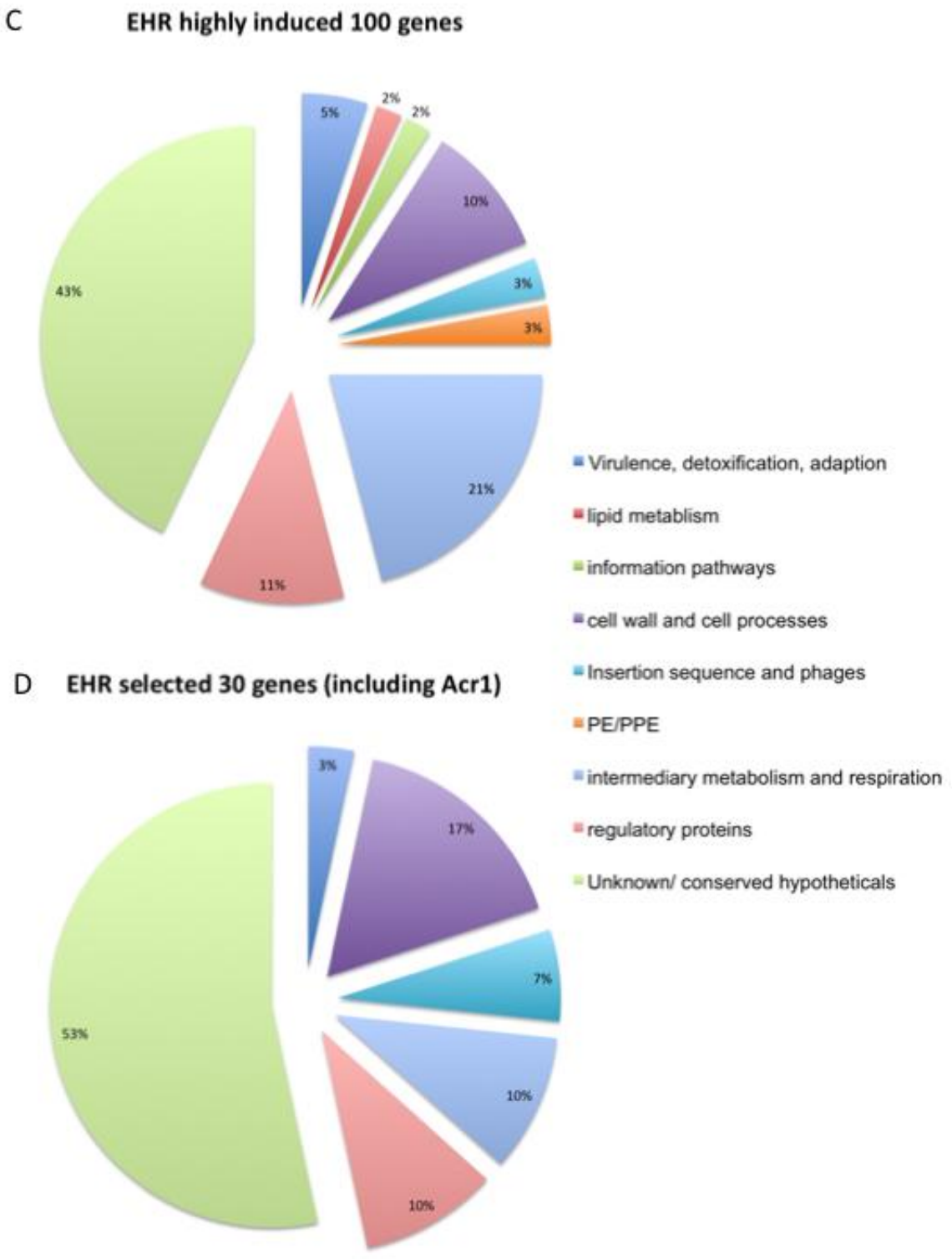


Figure 34 Distribution of MTB annotated genes according to functional category

Table 16 Percentage of MTB gene distribution according to functional category

CODE	Functional category	% of genes in the respective functional category				
		Whole genome	EHR complete list (230)	EHR top 100 genes	EHR selected 29 EHR genes (plus Acr-1)	Genes tested empirically (29 EHR selected genes plus Acr-1)
0	Virulence, detoxification, adaption	6	5	5	3	Rv2031 (hspX)
1	Lipid metabolism	6	3	2	0	-
2	Information pathways	6	3	2	0	-
3	Cell wall and cell processes	19	10	10	17	Rv0188, Rv0847, Rv0849, Rv1986*, Rv2693c
5	Insertion sequence and phages*	4	3	3	7	Rv2658*, Rv2659c*
6	PE/PPE	4	6	3	0	-
7	Intermediary metabolism and respiration	22	23	21	10	Rv1471, Rv2780, Rv3406
9	Regulatory proteins	5	13	11	10	Rv1956, Rv3334, Rv2021
8 /10	Unknown/ conserved hypotheticals	29	35	43	53	Rv0767c, Rv0826c, Rv0967, Rv0990c, Rv0991c, Rv1284, Rv1954, Rv1955, Rv1957, Rv2022c, Rv2466c, Rv2517, Rv2660c, Rv2662, Rv2663, Rv2664

* These genes are described in Chapter 6.

Empirical evaluation of EHR proteins

Human Subjects

Patients with active or latent tuberculosis were recruited at the Ubuntu clinic at Khayelitsha site B, South Africa. All were of Xhosa ethnicity. Latent tuberculosis is described as a positive TST (≥15mm) or ≥20 SFC/10^6 PBMC IFN-γ response to ESAT-6 or CFP-10. Active tuberculosis is defined by a sputum smear or sputum culture positive for *M. tuberculosis*. The baseline characteristics of subjects enrolled to the study are shown in Table 17. There was a significant difference in the median age between the two clinical groups. However, this difference did not affect the magnitude or frequency of response.

Table 17 Characteristic of persons recruited for study

	Active tuberculosis	Latent tuberculosis	Significance, p
Number	37	40	
Median age (Range)	36	20	<0.0001
Sex (Male: Female)	26M:11F	11M:32F	
Sputum smear positive	34 Positive; 2 negative; 1 unknown	NA	NA
Sputum culture positive	33 Positive; 2 negative 2 unknown/Unavailable	NA	NA
BCG vaccinated	11 vaccinated; 11 Not vaccinated; 15 Don't know/ data unavailable	20 vaccinated 16 Not vaccinated 4 Don't know/ data unavailable	
Median Mantoux (IQR) and % positive	NA	11 mm (IQR 8-19)	NA

Antigens

Synthetic peptides of 20 amino acid length overlapping by 10 amino acid, dissolved in DMSO (<0.2%) were used in the study. Antigenic stimuli were in the form of pools (each of the peptide pool contained minimum of 7 to maximum of 14 peptides per pool) for respective proteins with each peptide used at a final concentration of 10µg/ml. Control stimuli for ELISpot included anti-CD3 mAb CD3-2 (Mabtech) at 100 ng/ml final concentration and unstimulated wells. In addition 3 antigens were included as references: RD-1 encoded ESAT-6 and CFP-10; and DosR encoded Acr1 (which was also present amongst the 100 highly induced EHR genes).

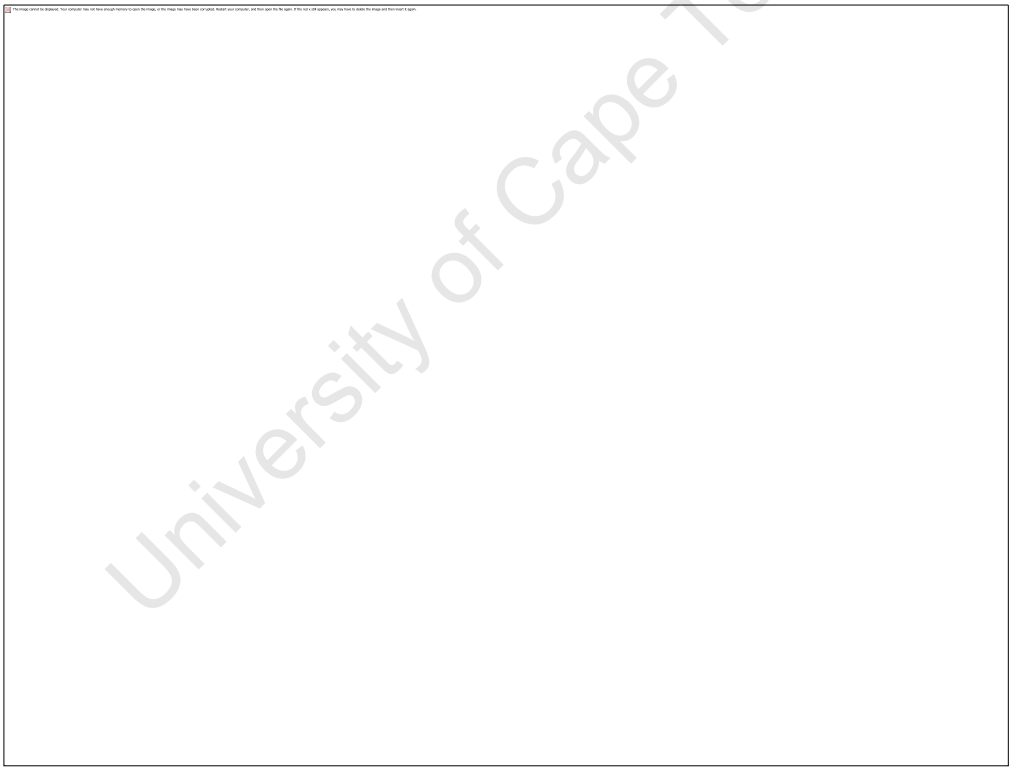


Figure 35 Comparison of two rounds of evaluation

Shows the frequency of response (described as response ≥ 30 SFC/10^6 PBMC above background) of Acr-1, CFP-10 and ESAT-6. There was no significant difference across two rounds of evaluation for all three control antigens.

Due to the limitation of availability of cells, the assays were done in two rounds. Three reference antigens: ESAT-6, CFP-10 and Acr-1 were used in both rounds. Since, there was no significant difference in the response to these 3 internal controls between the two rounds of evaluation (Figure 35), the results are described together.

Immunodominance was assessed both quantitatively (median SFC/10^6 PBMC) and as frequency of positive response. A positive response was defined by >20 SFC/10^6 PBMC for IL-2 and since IFN-γ response was primarily used to define the immunodominance, it was arbitrarily increased to > 30 SFC/10^6 PBMC of IFN-γ. The background response for IFN-γ and IL-2 were 6 SFC/10^6 PBMC and 8 SFC/10^6 PBMC respectively. Results presented are corrected for background.

The SFC responses were analyzed both as peptide pools and as proteins (summed response of respective peptide pools per protein) and described as response to proteins unless specified otherwise.

IFN-γ analysis of active and latent tuberculosis

IFN-γ ELISpot was performed in a total of 77 persons with active (37) or latent (40) tuberculosis. Each antigen was tested in at least 20 latent and 17-20 active tuberculosis persons.

When analyzed quantitatively, the highest IFN-γ SFC response in persons with latent disease was to CFP-10 (median 84 SFC/10^6 PBMC, IQR 36-234), which was higher than most (not all) EHR proteins and Acr-1 ($p \leq 0.034$). The exception were 7 EHR proteins: Rv0826, Rv0849, Rv1471, Rv1954c, Rv1955, Rv1957 and Rv2022c; and ESAT-6 ($p \geq 0.06$).

In persons with active tuberculosis, the highest response was to Rv1957 (136 SFC/10^6 PBMC, IQR 45-328), with the exception of Rv1954c, CFP-10 and ESAT-6 ($p \geq 0.205$), the response was higher than all other EHR proteins and Acr-1 ($p \leq 0.04$).

Amongst EHR proteins, Rv1955, Rv1471, Rv2022, Rv1954, Rv1284 and Rv0990 had dominant IFN-γ SFC response in persons with either latent or active tuberculosis; Rv0826c, Rv0849 and Rv2780 were recognized better by those with latent disease. All the other proteins evoked moderate levels of IFN-γ SFC response similar to that of Acr-1 or lower (Table 18; Figure 36, 41A).

When analyzed by the frequency of response, the highest response in those with latent and active disease was to CFP-10 (78%) and Rv1957 (88%) respectively. Rv1954, Rv1955 and Rv2022 were recognized at a similar frequency (>70%) as CFP-10 and ESAT-6 in both latent and active disease, while Rv0849 (70%) only in those with latent infection.

Rv0826, Rv0990, Rv1284 and Rv1471 were recognized >50% of those tested in both latent and active disease, while Rv2693 and Rv2780; and Rv2466 recognized only in latent and active disease respectively. All other antigens were recognized similar to that of Acr-1 or lower (Table 18; Figure 37, 41B).

Figure 36 Interferon-gamma analysis of EHR antigens in active and latent tuberculosis

Shows the spot forming cells (SFC) in patients with latent (blue) and active (red) tuberculosis represented as Box and whiskers (minimum to maximum, with median and IQR). The strongest response in latent disease was to CFP-10. The strongest response in active disease was to Rv1957. Peptide pool Rv2021c was preferentially recognized by latently infected persons ($p = 0.02$).

Figure 37 Interferon-gamma analysis of EHR antigens by frequency of response in active and latent tuberculosis

Shows the frequency of responders (defined by a response of ≥ 30 SFC/106 PBMC above background). No pool was preferentially recognized by either clinical group. The most frequent response in active disease was to Rv1957 (88%) and in latent disease was to CFP-10 (78%).

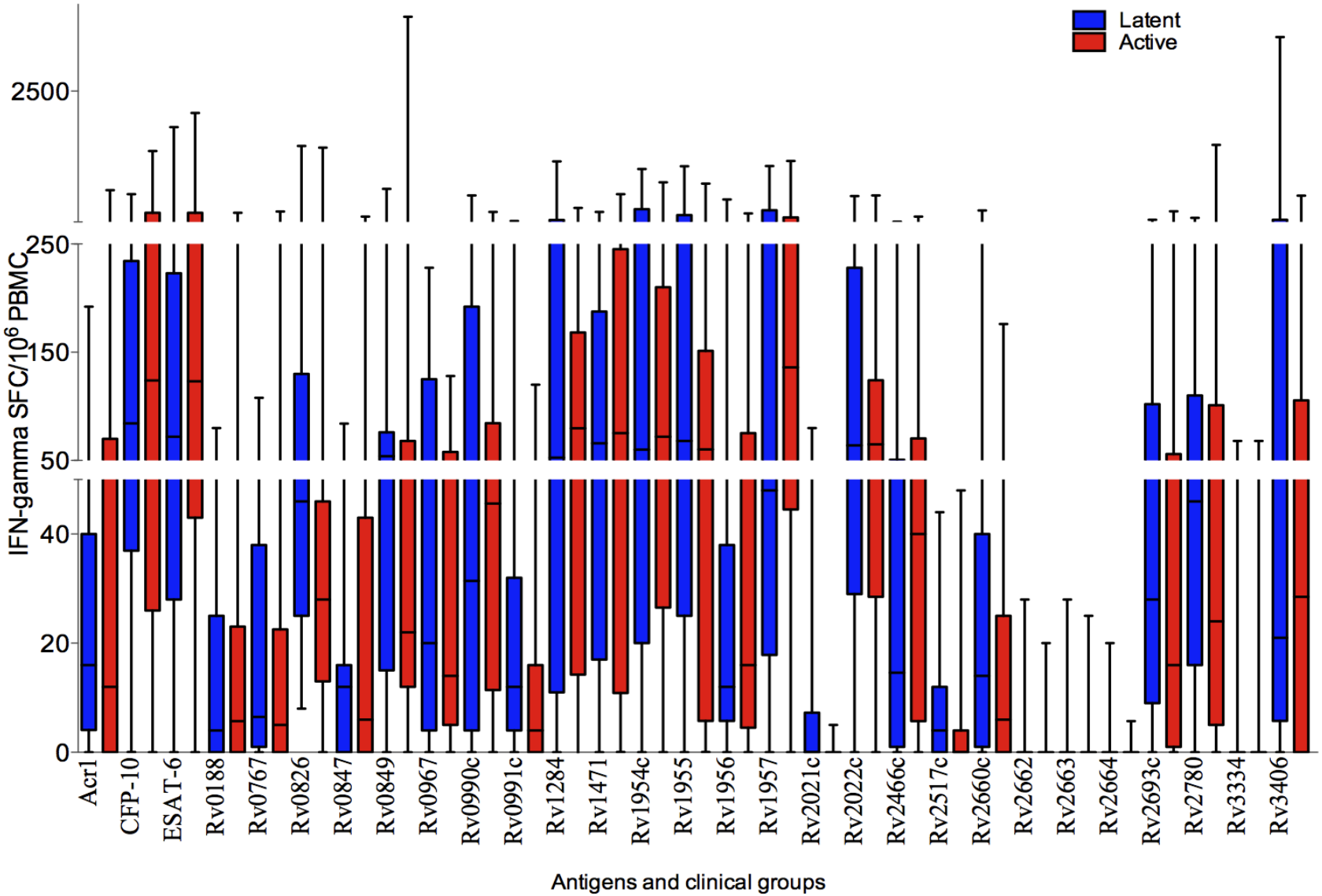


Figure 36 IFN -γ analysis of EHR antigens in active and latent tuberculosis

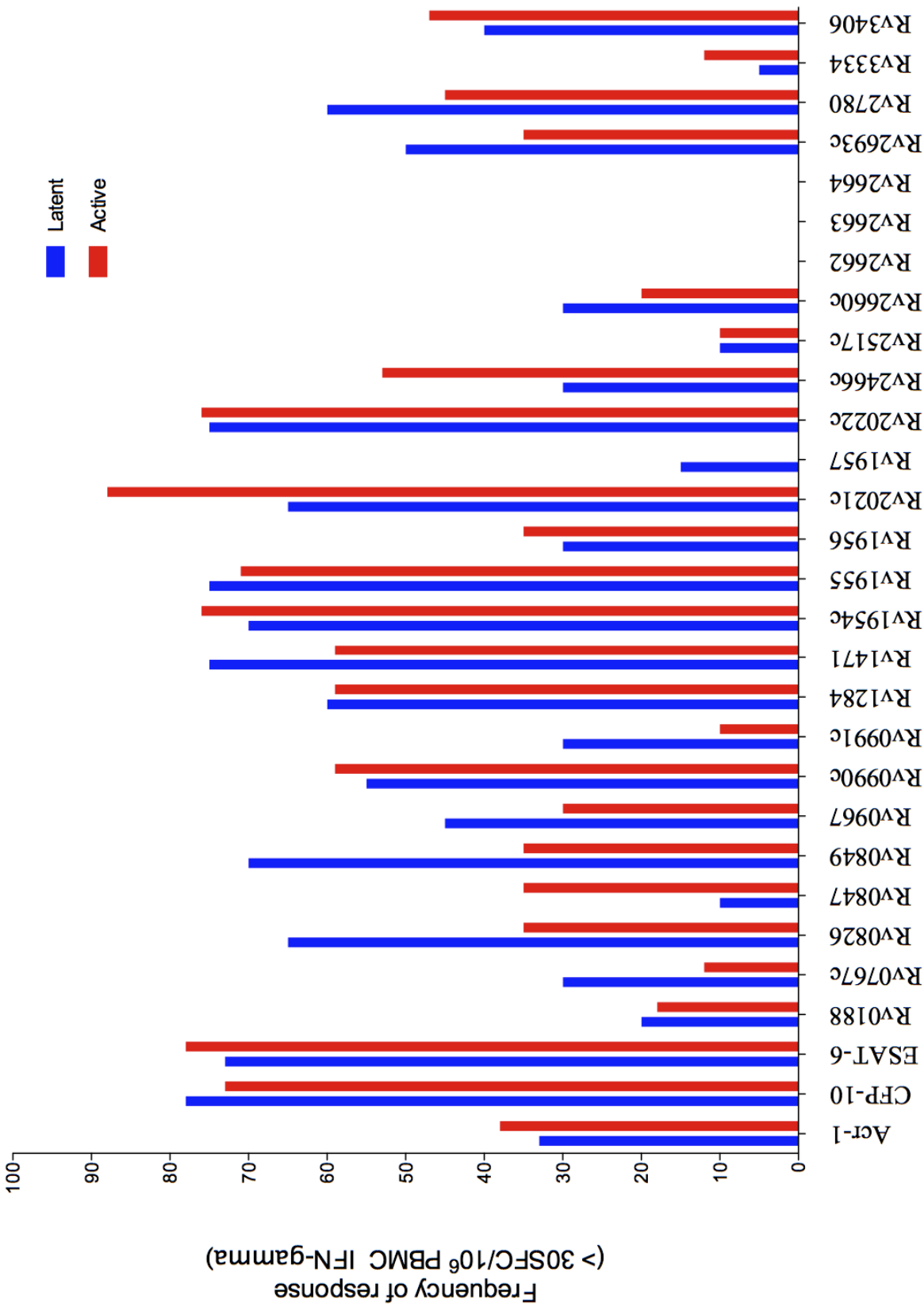


Figure 37 IFN-γ analysis of EHR antigens by frequency of response in active and latent tuberculosis

Preferential recognition in latent tuberculosis

When analyzed as proteins, Rv2021c, evoked a significantly higher response in those with latent tuberculosis (median 0 SFC/106 PBMC, IQR 0-7 in latent disease *vs.* 0 SFC/106 PBMC, IQR 0-0; p=0.02). However this response was very modest. None of the other antigens had significantly higher IFN-γ SFC in either of clinical groups.

However, when analyzed as peptide pools, peptide pools Rv0849-2 and Rv2693-1 were preferentially recognized by persons with latent infection (14 SFC/10^6 PBMC, IQR 1-23 in latent disease *vs.* 0 SFC/10^6 PBMC, IQR 0-11 in active disease; p=0.04) and 16 SFC/10^6 PBMC, IQR 5-65 *vs.* 6 SFC/10^6 PBMC, IQR 0-12; (p=0.04) respectively (Figure 38). However, when analyzed by the frequency of response, none of the antigens were preferentially recognized by latent or active disease groups.

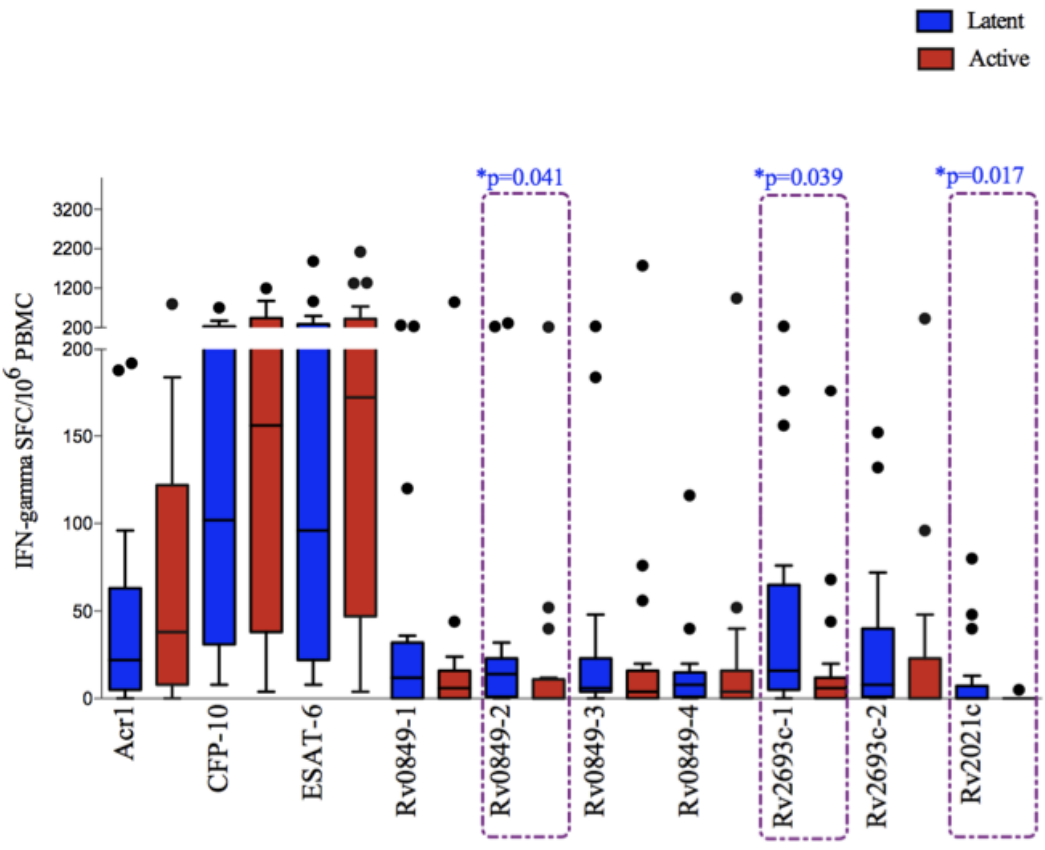


Figure 38 Interferon-gamma analysis of peptide pools that are preferentially recognized by latent disease

Spot forming cells (SFC) in patients with latent (L) and active (A) tuberculosis are shown as Box and whiskers (minimum to maximum, with median and IQR). The strongest response in latent disease was to CFP-10. With the exception of ESAT-6 all other responses were significantly lower ($p \leq 0.007$). The strongest response in active disease was to ESAT-6. With the exception of CFP-10 all other responses were significantly lower ($p \leq 0.002$). Peptide pool Rv0849-2, Rv2693-1 and Rv2021c were preferentially recognised by latently infected persons (p = 0.04, p=0.4 and p=0.02 respectively). However, the magnitude of the response was low and when analysed by the frequency of response, there was no significant difference.

IL-2 analysis of active and latent tuberculosis

IL-2 ELISpot was performed in the same persons as IFN-γ ELISpot, however complete data were not available for all subjects and all antigens/peptide pools, due to the limitation in cell numbers, and non specific binding of some peptides to the membrane and activation of the ALP-substrate complex (data not shown) which were excluded from the analysis.

Whilst the IFN-γ responses were spread amongst the EHR proteins tested, IL-2 responses were generally of lower magnitude and more focused, by both quantitative and frequency of response analysis (Figure 41).

Rv0826c dominated the IL-2 responses in both latent and active disease (median 276 SFC/10^6 PBMC, IQR 44-669 and 153 SFC/10^6 PBMC, IQR 5-575) and Rv2780 in latent disease (114 SFC/10^6 PBMC, IQR 9-761), both of which were dominant over CFP-10 and ESAT-6. All other antigens induced modest levels of IL-2, which were similar to that of Acr-1 or lower (Table 18).

All antigens induced similar numbers of IL-2 SFC amongst the two clinical groups with the exception of Rv2780, which induced higher numbers of IL-2 SFC in latent compared to active disease, however this difference did not attain statistical significance (Table 10; Figure 39, 41C).

When analyzed by the frequency of response, Rv0826 was the most frequently recognized antigen amongst those with latent infection (85%) with, Rv1284 greatest amongst those with active disease (73%). However, none of the antigens were preferentially recognized by latent or active disease groups (Table 18; Figure 40, 41D).

Figure 39 IL-2 analysis of EHR antigens in active and latent tuberculosis

Shows IL-2 spot forming cells (SFC) in patients with latent (L) and active (A) tuberculosis represented as Box and whiskers (minimum to maximum, with median and IQR). The strongest response in both active and latent disease was to Rv0826. All other responses with the exception of CFP-10, ESAT-6, Rv2780 and Rv1284 were lower. No protein was preferentially recognized by latently infected persons.

Figure 40 IL-2 analysis of EHR antigens by frequency of response in active and latent tuberculosis

Shows the frequency of responders (defined by a response ≥ 20 SFC/106 PBMC above background). No pool was preferentially recognized by either clinical group. The most frequent response in the latent group was to Rv0826 (86%) and in the active group to Rv1284 (73%).

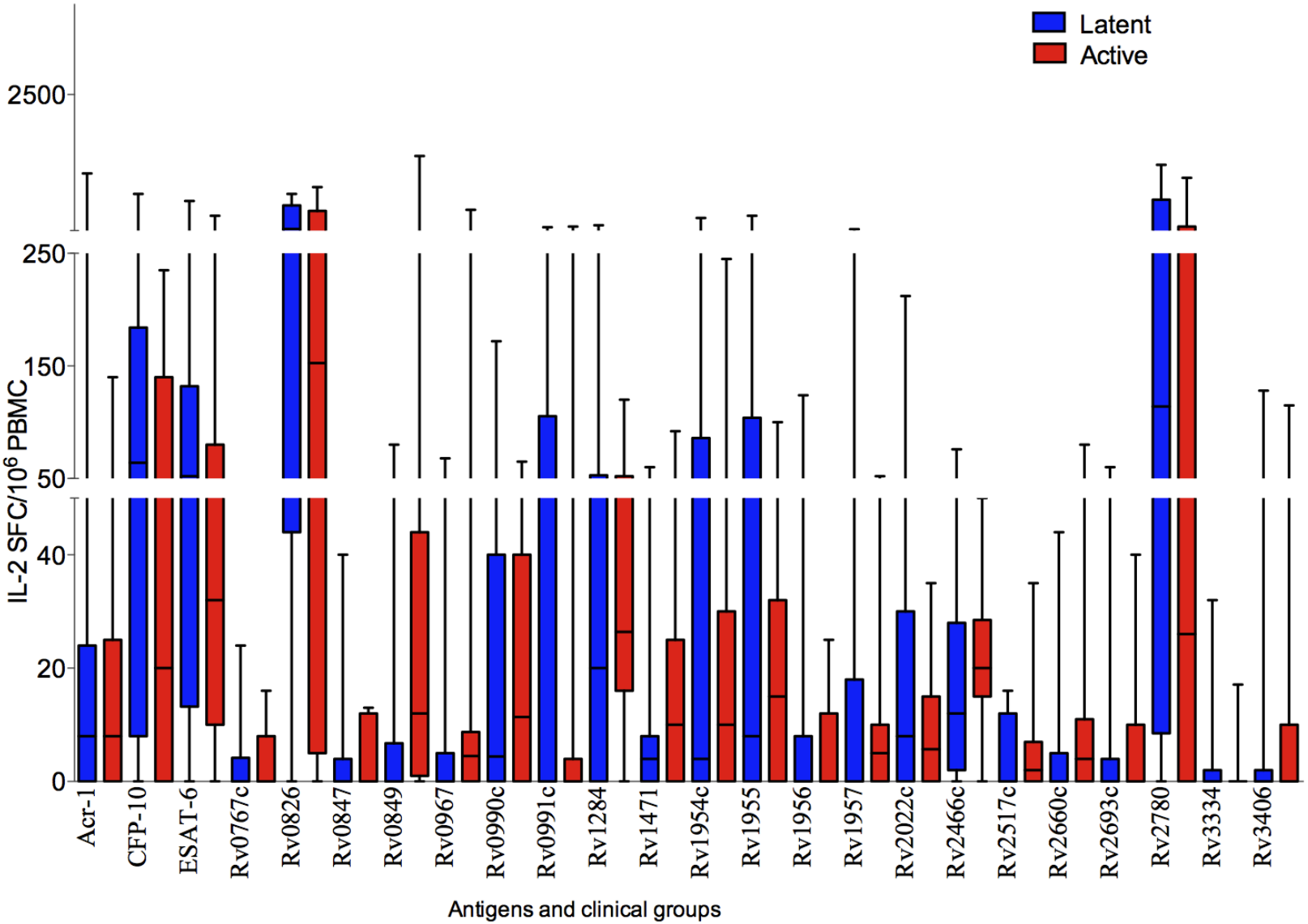


Figure 39 IL-2 analysis of EHR antigens in active and latent tuberculosis

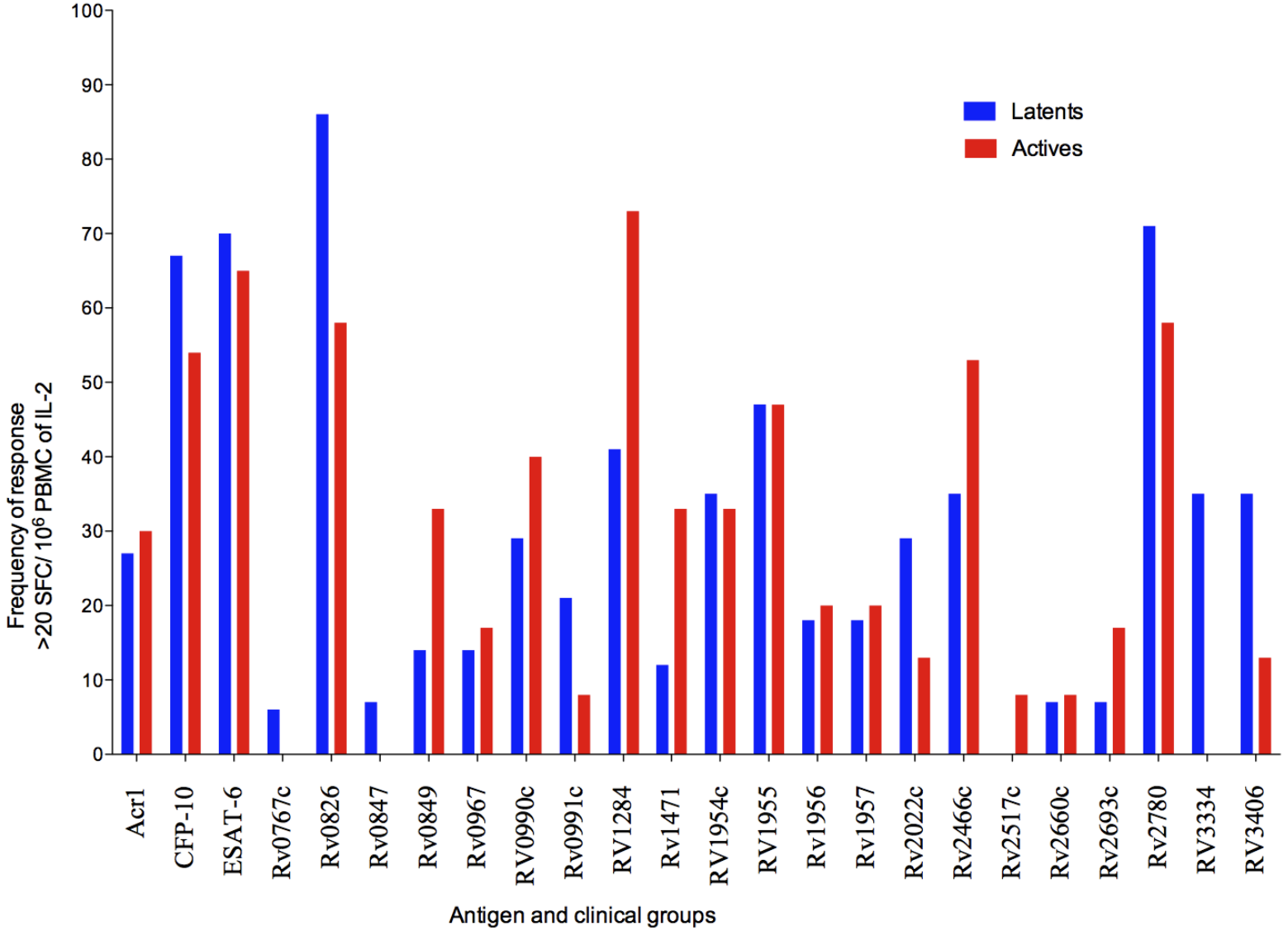


Figure 40 IL-2 analysis of EHR antigens by frequency of response in active and latent tuberculosis

Figure 41 Summary of interferon-gamma and IL-2 analysis of active and latent tuberculosis

Panel A shows the spot forming cells (SFC) in patients with latent (L) and active (A) tuberculosis represented as Box and whiskers (minimum to maximum, with median and IQR). The strongest response in latent disease was to CFP-10. The strongest response in active disease was to Rv1957. Peptide pool Rv2021c was preferentially recognized by latently infected persons ($p = 0.02$).

Panel B shows the frequency of responders (defined by a response of $\geq$ 30 SFC/10^6 PBMC above background). No pool was preferentially recognized by either clinical group. The most frequent response in active disease was to Rv1957 (88%) and in latent disease was to CFP-10 (78%).

Panel C shows IL-2 spot forming cells (SFC) in patients with latent (L) and active (A) tuberculosis represented as Box and whiskers (minimum to maximum, with median and IQR). The strongest response in both active and latent disease was to Rv0826. All other responses with the exception of CFP-10, ESAT-6, Rv2780 and Rv1284 were lower. No protein was preferentially recognized by latently infected persons.

Panel D shows the frequency of responders (defined by a response $\geq$ 20 SFC/10^6 PBMC above background). No pool was preferentially recognized by either clinical group. The most frequent response in the latent group was to Rv0826 (86%) and in the active group to Rv1284 (73%).

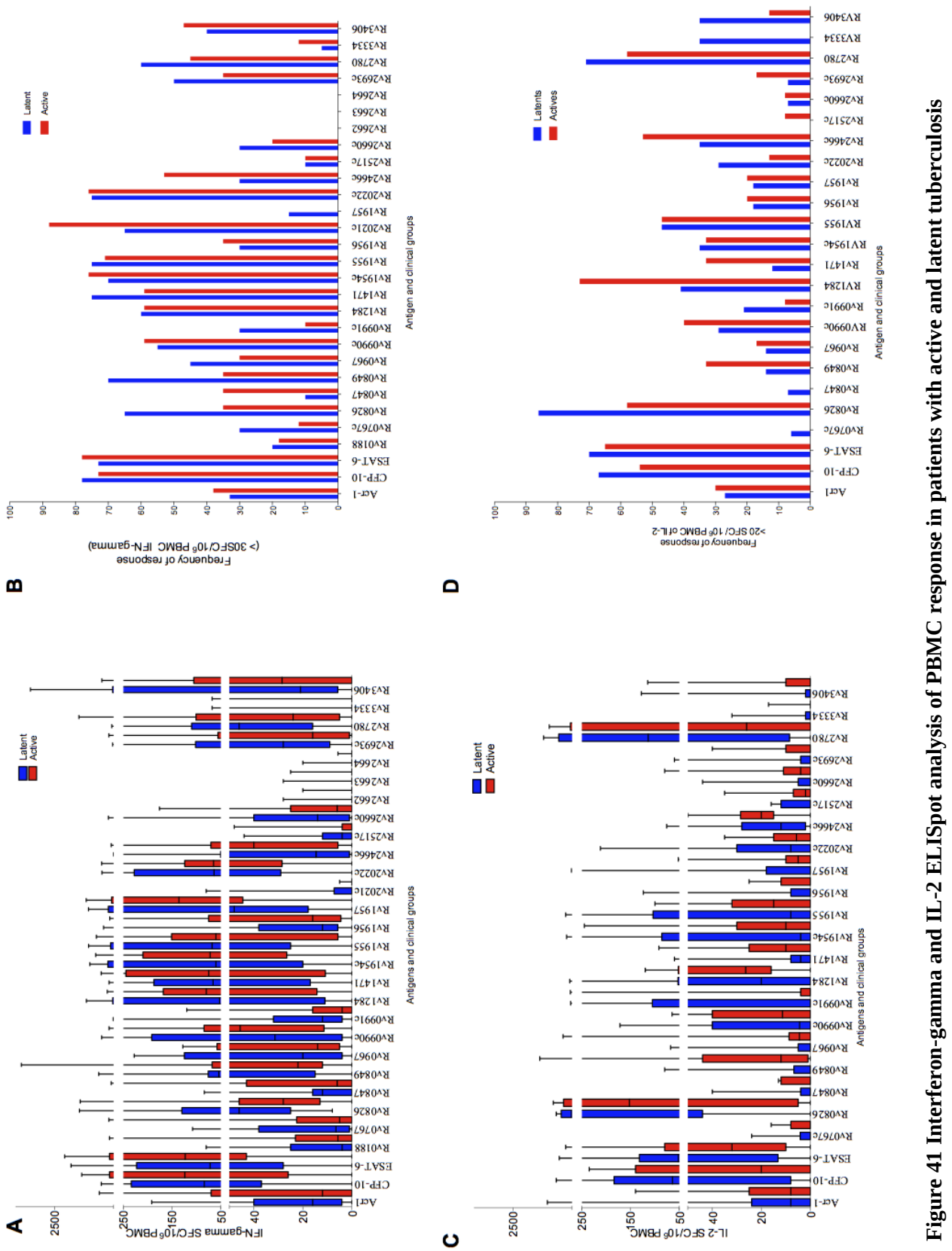


Figure 41 Interferon-gamma and IL-2 ELISpot analysis of PBMC response in patients with active and latent tuberculosis

Table 18 Summary of quantitative and Frequency of response of EHR antigens including Acr1, CFP-10 and ESAT-6

	Medians (SFC/10^6 PBMC) (IQR)						Proportion of responders					
	IFN-γ			IL-2			IFN-γ			IL-2		
	Latent	Active	p value	Latent	Active	p value	Latent	Active	p value	Latent	Active	p value
Acr-1	16 (4-40)	12 (0-70)	0.59	8(0-24)	8 (0-25)	0.68	33	38	0.64	26	30	0.78
CFP-10	84 (37-234)	124 (26-408)	0.87	64 (8-184)	20 (0-140)	0.06	78	73	0.79	68	52	0.28
ESAT-6	72 (28-223)	123 (43-410)	0.20	52 (13-132)	32 (10-80)	0.25	73	78	0.60	71	63	0.58
Rv0188	4 (0-25)	6 (0-23)	0.55	NA	NA	NA	20	18	1.00	NA	NA	NA
Rv0767 c	7 (1-38)	5 (0-23)	0.44	0 (0-4)	0 (0-8)	0.87	30	12	0.25	12	0	0.49
Rv0826	46 (25-130)	28 (13-46)	0.15	276 (44-669)	153 (5-575)	0.42	65	35	0.11	86	58	0.19
Rv0847	12 (0-16)	6 (0-43)	0.80	0 (0-4)	0 (0-12)	0.49	10	35	0.13	8	0	1.00
Rv0849	54 (15-76)	22 (12-68)	0.19	0 (0-7)	12 (1-44)	0.05	70	35	0.06	14	33	0.37
Rv0967	20 (4-125)	14 (5-58)	0.41	0 (0-5)	5 (0-9)	0.15	45	30	0.51	14	17	1.00
Rv0990c	32 (4-192	46 (11-85)	0.77	5 (0-40)	11 (0-40)	0.74	55	59	1.00	29	40	0.71
Rv0991c	12 (4-32)	4 (0-16)	0.09	0 (0-105)	0 (0-4)	0.78	30	10	0.24	21	8	0.60
Rv1284	57 (11-286)	80 (14-168)	0.89	20 (0-53)	27 (16-52)	0.34	60	59	1.00	53	73	0.29
Rv1471	66 (17-188)	75 (11-245)	0.84	4 (0-8)	10 (0-25)	0.18	75	59	0.48	12	33	0.21
Rv1954c	60 (20-470)	72 (27-210)	0.89	4 (0-86)	10 (0-30)	0.89	70	76	0.72	41	33	0.73
Rv1955	68 (25-367)	60 (6-151)	0.29	8 (0-104)	15 (0-32)	0.63	75	71	1.00	47	47	1.00
Rv1956	12 (6-38)	16 (5-75)	0.98	0 (0-8)	0 (0-12)	0.85	30	35	1.00	18	20	1.00
Rv1957	48 (18-454)	136 (45-328)	0.41	0 (0-18)	5 (0-10)	0.54	65	88	0.14	24	20	1.00
Rv2021c	0 (0-7)	0 (0-0)	**0.02**	Not tested	Not tested	NA	15	0	0.23	NA	NA	NA
Rv2022c	64 (29-228)	65 (29-124)	0.66	8 (0-30)	6 (0-15)	0.58	75	76	1.00	29	13	0.40
Rv2466c	15 (1-51)	40 (6-71)	0.18	12 (2-28)	20 (15-29)	0.16	30	53	0.19	35	53	0.48
Rv2517c	4 (0-12)	0 (0-4)	0.12	0 (0-12)	2 (0-7)	0.77	10	10	1.00	0	8	0.46
Rv2660c	14 (1-40)	6 (0-25)	0.34	0 (0-5)	4 (0-11)	0.30	30	20	0.72	7	8	1.00
Rv2662	0 (0-0)	0 (0-0)	0.91	Not tested	Not tested	NA	0	0	NA	NA	NA	NA
Rv2663	0 (0-0)	0 (0-0)	0.84	Not tested	Not tested	NA	0	0	NA	NA	NA	NA
Rv2664	0 (0-0)	0 (0-0)	0.35	Not tested	Not tested	NA	0	0	NA	NA	NA	NA
Rv2693c	28 (9-102)	16 (1-56)	0.16	0 (0-4)	0 (0-10)	0.72	50	35	0.52	7	17	0.58
Rv2780	46 (16-110)	24 (5-101)	0.87	114 (9-761)	26 (0-319)	0.20	60	45	0.53	71	58	0.68
Rv3334	0 (0-0)	0 (0-0)	0.85	0 (0-2)	0 (0-0)	0.79	5	12	0.58	6	0	1.00
Rv3406	21 (6-290)	29 (0-106)	0.59	0 (0-2)	0 (0-10)	0.36	40	47	0.75	12	13	1.00

Responses of adjacent gene products: Rv1954c-Rv1957

Four adjacent genes: Rv1954-Rv1957 covering 2 predicted operons (Rv1954c-Rv1955 and Rv1956-Rv1957) were amongst the 100 highly induced EHR genes. IFN-γ responses to these antigens, with the exception of Rv1956 were similar in magnitude (Rv1954c: $p \geq 0.579$; Rv1955: $p \geq 0.242$; Rv1957: $p \geq 0.781$) and frequency of recognition ($p \geq 0.787$) as that of CFP-10 and ESAT-6 (Figure 42 A and B). However, responses to Rv1956 were moderate and at similar levels of Acr-1 (p=1.000) and significantly lower than CFP-10 and ESAT-6 ($p \leq 0.0001$). Rv1957 (77%) was recognized with greater frequency (not significantly different) than either CFP-10 (75%) or ESAT-6 (75%) (Figure 42 B). The IL-2 responses to these antigens were however very modest.

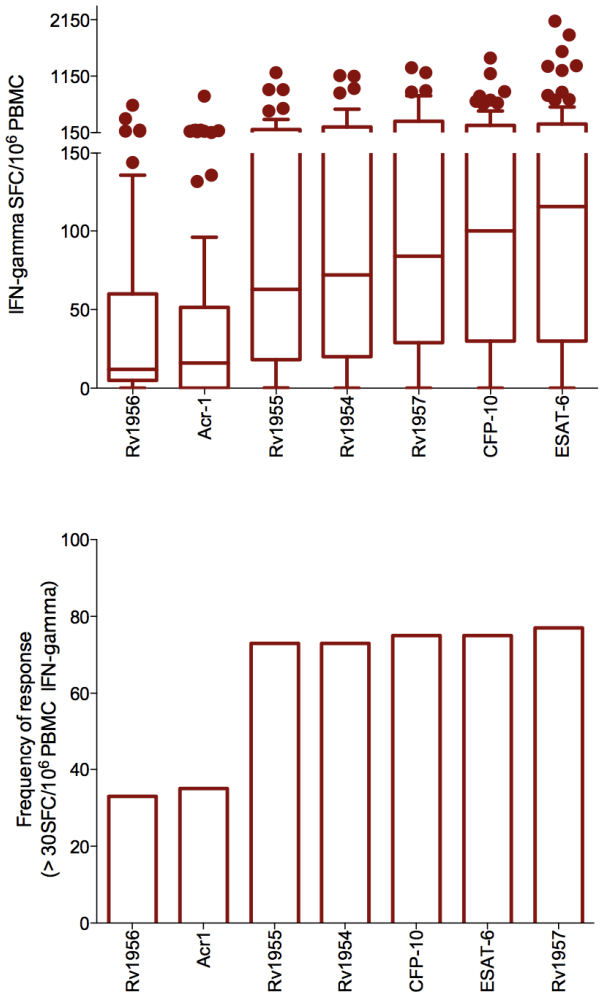


Figure 42 IFN-γ analysis of Rv1954c, Rv1955, Rv1956 and Rv1957

Panel A shows IFN-γ spot forming cells (SFC) in combined patient group (latent and active disease) are represented as Box and whiskers (minimum to maximum, with median and IQR).
Panel B shows frequency of recognition by IFN-γ response, and arranged according to the increase in frequency response to the protein.

Intramolecular influence on the phenotype of T cell response

Rv0826c, a relatively large molecule of 351 amino-acids, had to be divided into 3 peptide pools for empirical testing. Pools 1 and 2 (covering the αα region 1-120 (Pool-1) and 111-240 (pool 2) were moderate stimulants of an IFN-γ response that was of similar order of magnitude (not significantly higher) to the response to Acr1 (p=0.055). However, pool 3 (αα 231-350) induced significantly lower IFN-γ when compared to Pool 1 and 2 (p=0.008). By contrast, peptide pools 1 and 2 induced lower levels IL-2 response, while the C-terminal peptides (peptide pool 3) induced a strong IL-2 response that was significantly greater than pool 1 and 2 and Acr-1 ($p \leq 0.007$). The levels of IL-2 tended to be higher than the response to CFP-10 and ESAT-6 (p=0.07) (Figure 43).

Figure 43 Intramolecular influence on the phenotype of T cell response

The IFN-γ (Panel A) (n=40) and IL-2 (Panel B) (n=27) response (median SFC response with IQR) to 3 pools of Rv0826 is shown. Since there was no significant difference between active and latent disease groups, the results are represented for combined groups. Pools 1 and 2 of Rv0826 were moderate stimulants of the Interferon-gamma response that was of similar order of magnitude to Acr1. By contrast peptide pools 1 and 2 poorly stimulated an IL-2 response but the C-terminal peptides (pool 3) induced significantly stronger IL-2 response.

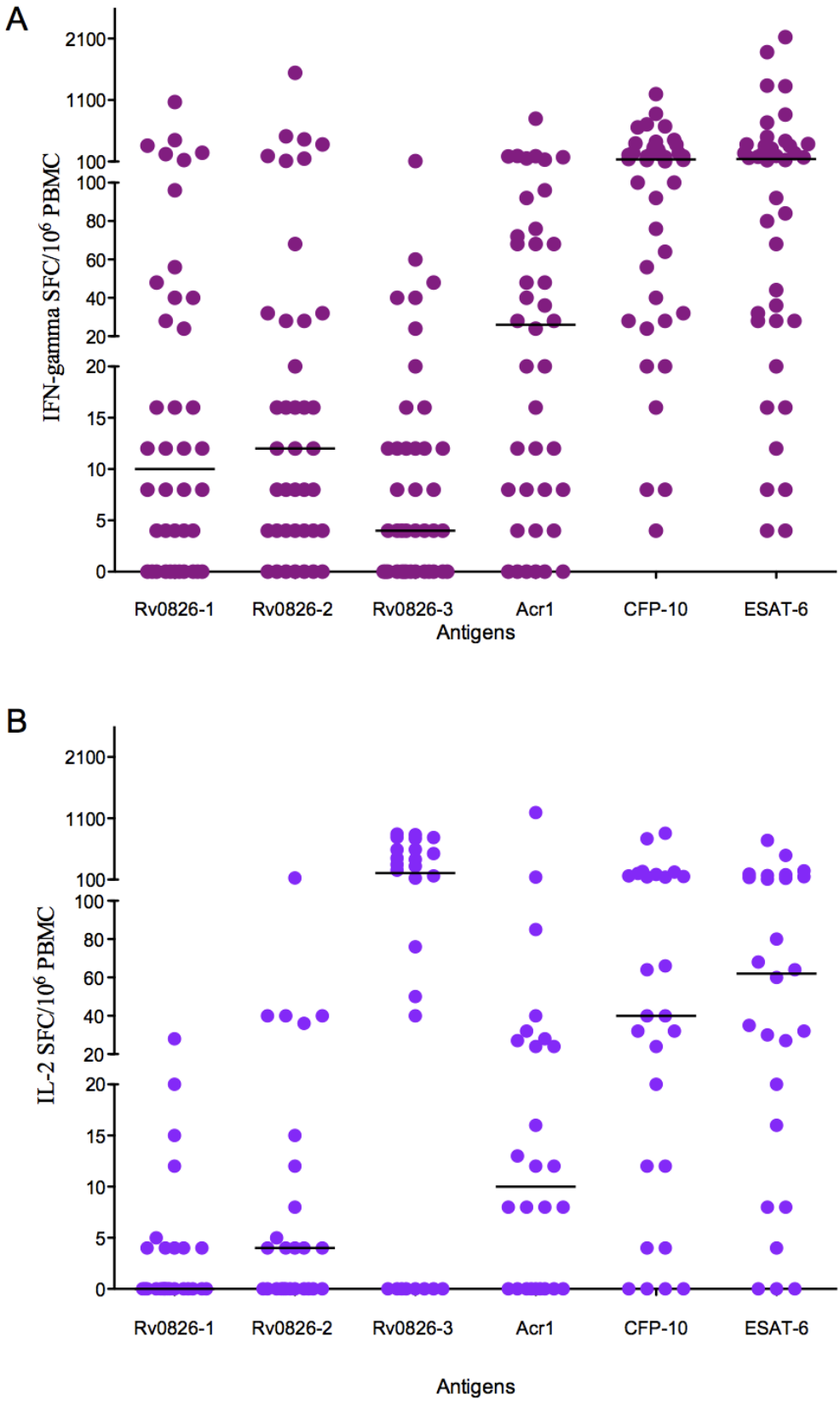


Figure 43 Intramolecular influence on the phenotype of T cells

Reconciliation of bioinformatic prediction with empirical findings

I next evaluated hierarchies of immunodominance (empirical testing) by both quantitative (Median SFC) and Frequency of response (>30 SFC/10^6 PBMC) measures and reconciled with hierarchies of informatic prediction. Only IFN-γ ELISpot results were used for this analysis, as it formed the most complete data set. The median SFC and frequency of response to EHR proteins were independently ranked from 1-26 by clinical groups (persons with latent and active disease), and since there was no significant difference in the responses between the 2 clinical groups, the average rank of clinical groups were calculated and represented as average rank for each protein (Table 19; Figure 44).

For hierarchies of informatic prediction, existing ranks were extracted from rank list: R(FI), R(NI) and R(cPP); and re-assigned (adjusted) ranks (iR(FI), iR(NI) and iR(cPP)) between 1-26 for every category. Adjusted ranks, with combinations (as described in Table 19) were used to test the correlation between informatic prediction and empirical findings. Non parametric Spearman's correlation was used to test the correlation between empirical ranks (median and proportion of responders) against informatic ranks (individual and combination ranks) of iR(FI) ,iR(NI) and iR(cPP) (Table19 and Table 20).

When empirical ranking was correlated with bioinformatic prediction, there was no correlation of ranks based on fold induction (iR(FI); r=0.1; p=0.5); normalized intensity (iR(NI); r=-0.1;p=0.7); combined ranks of normalized intensity or population specific MHC-peptide binding prediction (iR(NI) + iR(cPP); r=0.2; p=0.23). However, significant correlation was found with the combined fold induction and population specific MHC-binding peptide (iR(FI) + iR(cPP); r=0.5; p=0.02) (Figure 45-46 and Table 20).

Table 19 Ranks of Empirical and Informatic analysis

* There was no significant difference observed in the IFN-γ response between active and latent tuberculosis. Therefore, those with latent and active tuberculosis were combined for this analysis.

(a) Hierarchy defined by the Median IFN-γ ELISpot response. IFN-γ median SFC/ 10^6 PMBC responses of all the proteins were ranked. The ranks ranged from 1 to 26 (highest response to lowest response).

(b) Hierarchy defined by Frequency of response. IFN-γ Frequency of response (defined as >30 SFC/10^6 PBMC of IFN-γ) of all the proteins were ranked. The ranks ranged from 1 to 26 (highest response to lowest response).

(c) Hierarchy defined by Fold induction of EHR proteins (iR(FI)). The ranks from R(FI) were extracted and re-assigned as per hierarchy of informatics as referred as iR(FI).

(d) Hierarchy defined by *sig*A normalized transcript levels of EHR proteins (iR(NI)). The ranks from R(NI) were extracted and re-assigned as per hierarchy of informatics as referred as iR(NI).

(e) Hierarchy defined by population specific MHCII binding peptide prediction (iR(cPP)). The ranks from R(cPP) were extracted and re-assigned as per hierarchy of informatics as referred as iR(cPP).

(f) Hierarchy defined by combined score: iR(FI)+ iR(cPP). Hierarchy defined by Fold induction of EHR proteins (iR(FI)) was complemented with hierarchy of population specific MHCII binding peptide prediction(iR(cPP)).

(g) Hierarchy defined by combined score: iR(NI)+ iR(cPP). Hierarchy defined by *sig*A normalized transcript levels of EHR proteins (iR(NI)) was complemented with hierarchy of population specific MHCII binding peptide prediction (iR(cPP)).

Table 19 Ranks of Empirical and Informatic analysis

	[a]Median*	[b]Frequency of response *	[c]$iR_{(FI)}$	[d]$iR_{(NI)}$	[e]$iR_{(cPP)}$	[f]$iR_{(FI)}$+ $iR_{(cPP)}$	[g]$iR_{(NI)}$+ $iR_{(cPP)}$
Rv0188	20	20	20	10	7	15	5
Rv0767c	19	18	8	8	12	8	10
Rv0826	10	10	11	6	4	2	2
Rv0847	17	17	1	2	16	4	6
Rv0849	9	9	22	24	3	12	14
Rv0967	13	14	3	1	18	9	8
Rv0990c	8	7	9	16	10	7	12
Rv0991c	18	19	2	3	23	13	13
Rv1284	6	6	15	22	8	10	17
Rv1471	3	5	4	13	19	11	18
Rv1954c	2	3	19	25	14	19	23
Rv1955	5	4	5	5	13	6	7
Rv1956	15	15	14	11	22	21	19
Rv1957	1	2	10	18	6	3	11
Rv2021c	23	23	23	19	9	18	16
Rv2022c	4	1	16	21	17	20	21
Rv2466c	14	13	7	4	11	5	3
Rv2517c	21	21	21	9	25	25	20
Rv2660c	16	16	24	26	26	26	26
Rv2662	24	24	13	23	24	22	25
Rv2663	25	25	17	17	21	24	22
Rv2664	26	26	18	20	20	23	24
Rv2693c	12	12	26	14	2	17	4
Rv2780	7	8	25	7	1	14	1
Rv3334	22	22	12	12	15	16	15
Rv3406	11	11	6	15	5	1	9

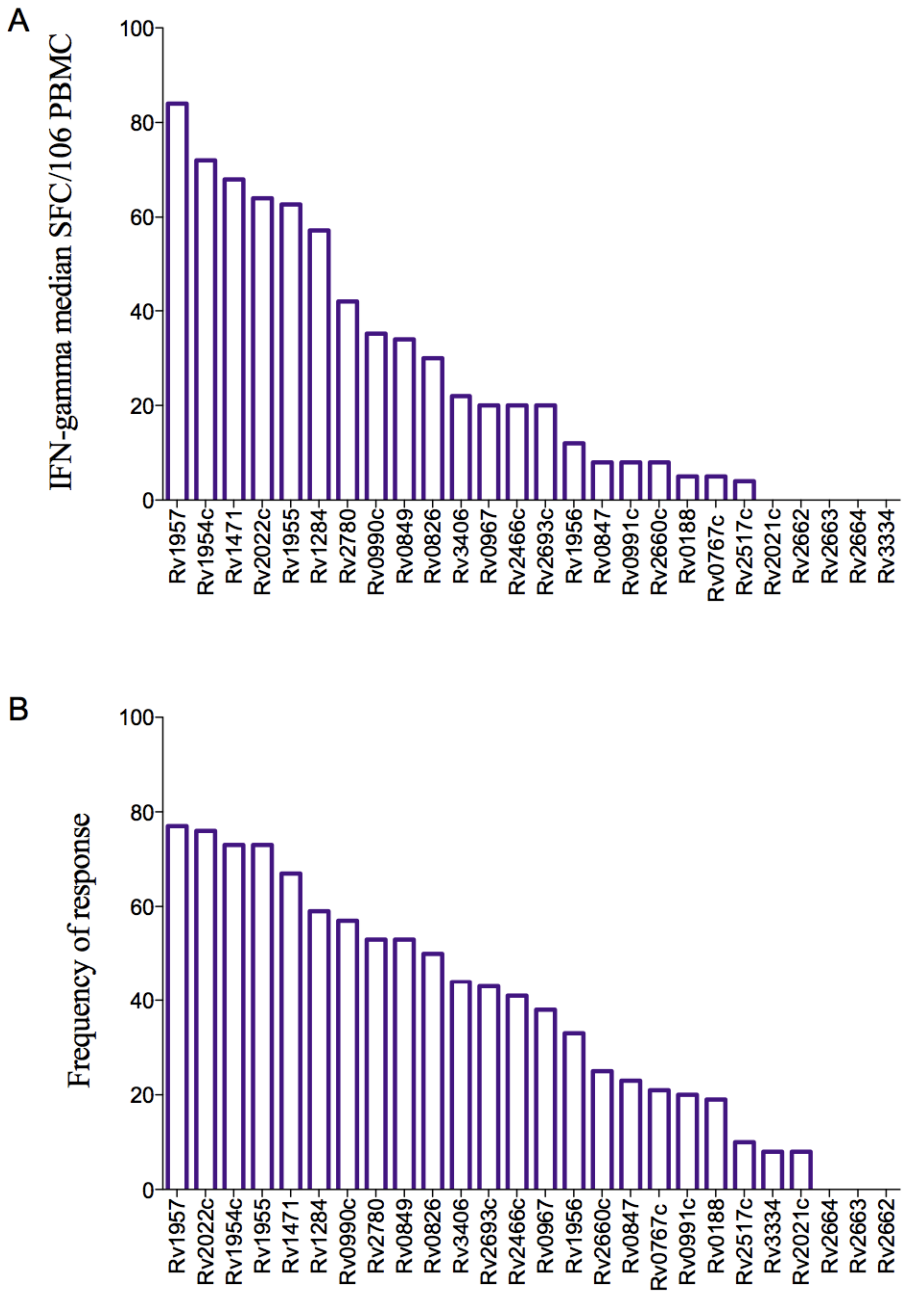


Figure 44 Hierarchy of immunodominance amongst EHR antigens

Panel A shows hierarchies of immunodominance by quantitative response (Median SFC/10^6 PBMC of IFN-γ). Bars indicate median response.

Panel B shows hierarchies of immunodominance by frequency of recognition (>30 SFC/10^6 PBMC of IFN-γ). Bars indicate frequency of recognition.

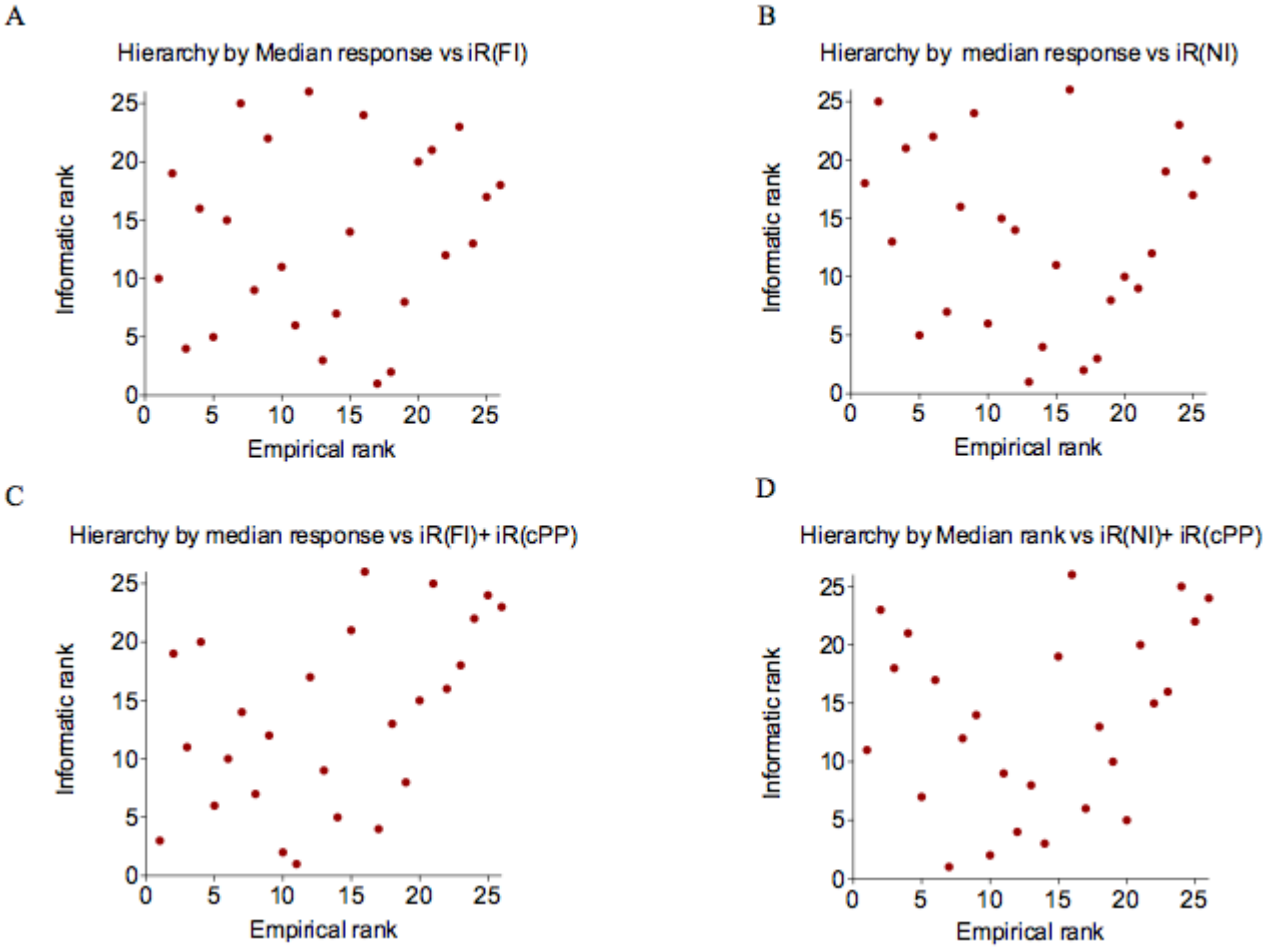


Figure 45 Reconciliation of bioinformatic prediction and empirical evaluation (Median response)

A-D shows correlation of empirical ranking by median response against informatic ranks: $iR_{(FI)}$ (A) ,$iR_{(NI)}$ (B) , $iR_{(FI)}$ +$iR_{(cPP)}$ (C) and $iR_{(NI)}$+$iR_{(cPP)}$ (D).

Panel A shows the correlation between the hierarchies defined by the median response and fold induction $iR_{(FI)}$. The spearman's rho = 0.14 (p=0.49).

Panel B shows the correlation between the hierarchies defined by the median response and normalised transcript levels $iR_{(FI)}$.The spearman's rho = -0.09 (p=0.67).

Panel C shows the correlation between the hierarchies defined by the median response and Fold induction, complemented with the population specific MHCII binding peptide prediction ($iR_{(NI)}$+$iR_{(cPP)}$). There was significant correlation: spearman's rho = 0.47 (p=0.01).

Panel D shows the correlation between the hierarchies defined by the median response and Normalised transcript levels, complemented with the population specific MHCII binding peptide prediction ($iR_{(NI)}$+$iR_{(cPP)}$). There was no correlation. The spearman's rho = 0.24 (p=0.23).

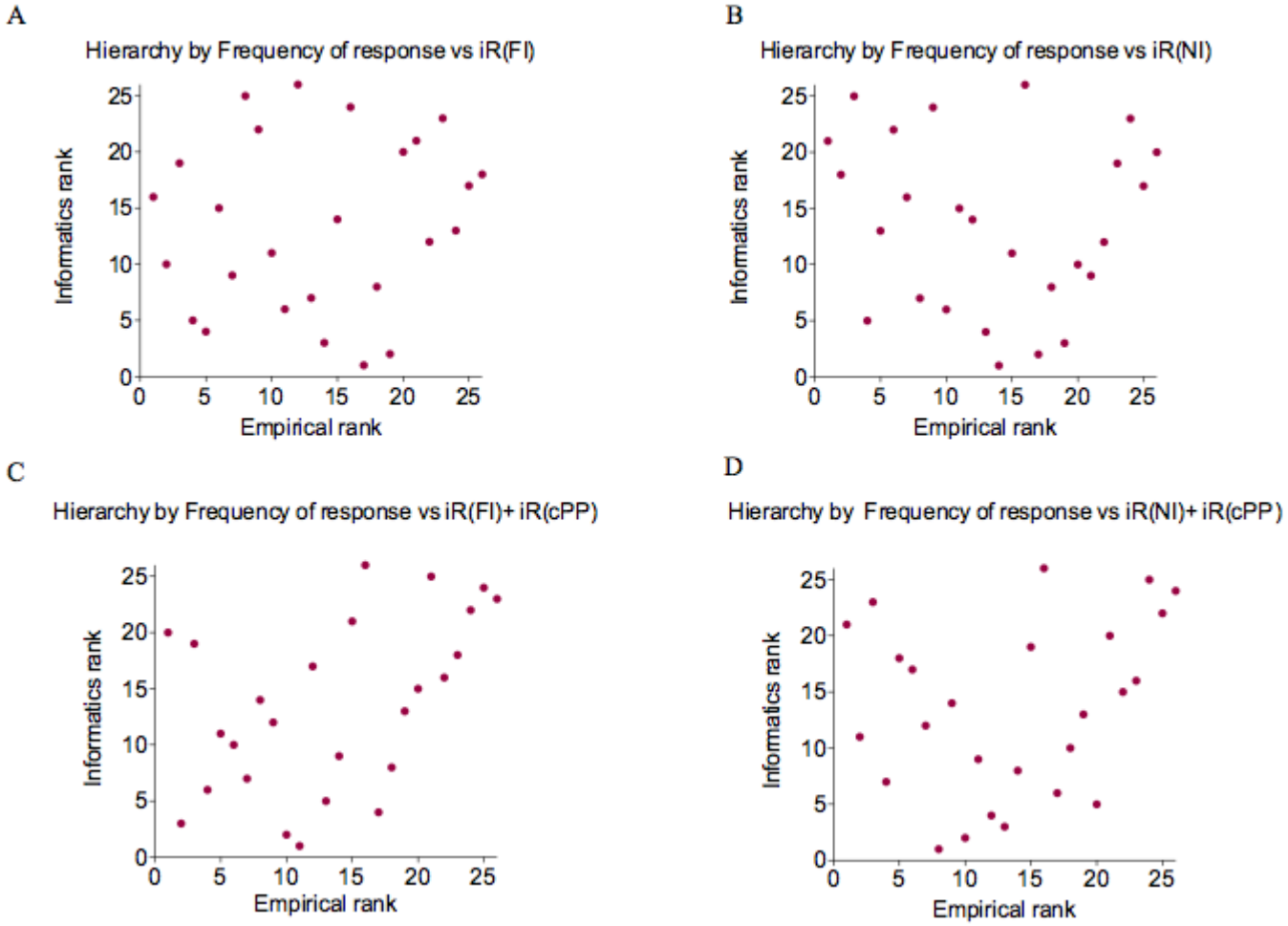


Figure 46 Reconciliation of Bioinformatics prediction and empirical evaluation (Frequency of response)

A-D shows correlation of empirical ranking by frequency of response against informatics ranks: $iR_{(FI)}$ (A) ,$iR_{(NI)}$ (B) , $iR_{(FI)}$ +$iR_{(cPP)}$ (C) and $iR_{(NI)}$+$iR_{(cPP)}$ (D).

Panel A shows the correlation between the hierarchies defined by the frequency of response and Fold induction $iR_{(FI)}$. The spearman's rho = 0.14 (p=0.51).

Panel B shows the correlation between the hierarchies defined by the frequency of response and normalised transcript levels $iR_{(NI)}$. The spearman's rho = -0.10 (p=0.63).

Panel C shows the correlation between the hierarchies defined by the frequency of response and Fold induction, complemented with the population specific MHCII binding peptide prediction ($iR_{(NI)}$+$iR_{(cPP)}$). There was significant correlation: spearman's rho = 0.47 (p=0.02).

Panel D shows the correlation between the hierarchies defined by the median response and Normalised transcript levels, complemented with the population specific MHCII binding peptide prediction ($iR_{(NI)}$+$iR_{(cPP)}$). There was no correlation. The spearman's rho = 0.24 (p=0.24).

Table 20 Correlation of Empirical and Informatic prediction

	Rank by Median			**Rank by Proportion of response**		
Parameter	Spearman r	P value	Summary	Spearman r	P value	Summary
iR(FI)	0.14	0.49	ns	0.14	0.51	ns
iR(NI)	-0.09	0.67	ns	-0.10	0.63	ns
iR(FI)+ iR(cPP)	0.47	0.01	*	0.47	0.02	*
iR(NI)+ iR(cPP)	0.24	0.23	ns	0.24	0.24	ns

Discussion

Several, novel immunodominant T cell targets induced during enduring hypoxic response of *Mycobacterium tuberculosis* have been discovered by a combined approach, utilizing bioinformatic prediction algorithms and empirical evaluation.

Whole genome based transcriptional profiles of *M. tuberculosis* subject to prolonged hypoxia to guide the discovery of novel and potential antigens were analyzed. Based on the discovery of ~230 genes, significantly induced at four and seven days of hypoxia, but not at the initial time points, these sets of genes were described as the Enduring hypoxic response (EHR) of *Mycobacterium tuberculosis* (Rustad et al., 2008).

The major hypothesis was that these molecules may be expressed *in vivo* and there by could be targets of the immune response. To further understand the biology of latent infection and to consider what T cells may preferentially recognize during a latent infection, this systematic evaluation of one hundred EHR genes was conducted. Another hypothesis that was of interest to be tested was: when faced with the genomic data from a pathogen, can antigens be predicted?

Therefore, an Immuno informatic approach was designed, whereby the available genomic data was coupled with the use of *in silico* tools to search for promising proteins with epitopes that are highly predicted to bind to MHC class II molecules, and *in vitro* screening of shortlisted candidate antigens. This rational approach, was not confined only to the predicted binding region of proteins, but developed a protocol for scoring and ranking the prediction and prioritization for the gene products (Tables 11 and 12). The prediction algorithm was complemented with correction for MHC allelic polymorphism that exists between human populations (du Toit et al., 1988). As the screening of 100 novel candidates empirically would

have been a considerable undertaking, antigens were selected for *in vitro* screening not only based on the predicted MHC-class II–peptide binding, but combined by hierarchies of genome based transcriptomic data and absolute abundance of each of the transcripts.

This study shows that genomic knowledge can aid antigen discovery to an extent when it is complemented with population specific MHC-class II-peptide prediction analysis. These findings revealed that a number of genes induced by hypoxia were in fact potent T-cell targets and therefore offer general support to the hypoxia hypothesis and its relationship to the natural infection of MTB. In addition, also have described interesting, immunologically potent T cell targets, which are potential candidates for further evaluation. However, these findings did not provide substantial support the hypothesis of preferential recognition of these antigens in those with latent disease and rather provides support for a strong overlapping immunological spectrum between those with latent and active TB disease as suggested (Barry et al., 2009; Young et al., 2009)

In silico tools offer an attractive alternative to the cumbersome empirical evaluation to identify T-cell epitopes that has hitherto been necessary for the discovery of vaccine candidates. However, studies have confirmed that *in silico* methods cannot be wholly relied on for selecting crucial peptides for development of vaccines (Gowthaman and Agrewala, 2008; Gowthaman and Agrewala, 2009; Lin et al., 2008).

The ProPred prediction algorithm was selected because of its easy user interface, availability of all 51 HLA DR alleles (including allelic variants) (Singh and Raghava, 2001), previous experimental evaluation (Gowthaman and Agrewala, 2008; Lin et al., 2008) and in antigen discovery in both MTB (Al-Attiyah and Mustafa, 2004; Mustafa and Shaban, 2006; Vordermeier et al., 2003) and in *Plasmodium vivax* (Lima-Junior et al., 2010).

Therefore a combination of algorithms were used to aid our antigen discovery and also reconciled with empirical analysis to test our prediction algorithm. These findings suggest, that these *in silico* tools when complemented with the correction for allelic polymorphism found within the different populations can be useful to an extent for the prediction of T cell antigens.

Another hypothesis that was of interest was to test whether, by virtue of up regulation during prolonged hypoxia, proteins encoded by such genes would be preferentially recognized by latently infected persons. Reconciliation of informatic and empirical evaluation shows that, neither the fold induction nor the normalized intensity of the genes/proteins correlates with the IFN-γ response. With the exception of modest IFN-γ response to the peptide pool of Rv0849-2 and Rv2693-1, and very small levels for Rv2021c (Figure 36 & 38), which were significant towards preferential recognition in latent disease, this was not the case. Although, there was a trend towards preferential recognition, especially with those with moderate IFN-γ response, the latently infected group had a slightly higher magnitude and frequency of recognition. Overall, the response between the two clinical groups overlapped. This suggests a strong overlap between the immunological spectrum of active and latent disease. Hypoxia does characterize granulomas in tuberculosis infection, but it is increasingly re-appreciated that even those with active TB disease will have a spectrum of lesions, similar to those of the latently infected. Therefore, it is likely that the hypoxic lesions are present in both clinical states (Barry et al., 2009).

This study is the first evaluation of EHR genes for human T cell response. IFN-γ ELISpot analysis revealed 5 of 10 immunodominant candidates: Rv1957, Rv2022c, Rv1954, Rv1955 and Rv1471 with a frequency of response >70%, similar to that of ESAT-10 and CFP-10 irrespective of clinical group. These are potential candidates for further evaluation as potent T cell targets. It is also worth of note, of these 5 proteins, with an exception of Rv1471, do not

share regions of similarity other than *M. tuberculosis*, *M. bovis* and *M. bovis* BCG as determined by BLASTn and BLASTp analysis against all nucleotide and protein database (Tables 14 and 15). Therefore a response to these proteins is less likely to arise because of exposure or infection by environmental or non-tuberculous mycobacteria.

The immunodominant EHR T cell targets for IFN-γ response are classified as being involved in intermediately metabolism and respiration (Rv1471) and unknown function or conserved hypotheticals (Rv1954c, Rv1955, Rv1957, Rv2022).

Rv1955-Rv1956 encode a Toxin-antitoxin (TA) module (Ramage et al., 2009) (Gupta, 2009), one of 38 such loci in MTB (Gerdes et al., 2005; Pandey and Gerdes, 2005; Ramage et al., 2009). Studies have shown that Rv1955, Rv1956 and Rv1957 are co-regulated (Smollett et al., 2009). Studies on TA module suggests that these loci provide a control mechanism that helps to cope with nutritional stress (Gerdes et al., 2005) such as hypoxia. Recently Schuck and colleagues described another protein Rv3407, which forms part of the TA module, to be preferentially recognized by T cells of those with latent tuberculosis infection (Schuck et al., 2009). Taken together, these findings further provide support and indicate a possible role for these genes in the maintenance of *M. tuberculosis* under stressful conditions *in vivo*.

Rv1284, a beta-carbonic anhydrase, was the only essential gene that was included in the empirical evaluation: It induced both IFN-γ (>50% response) and IL-2 responses.

Since IFN-γ is a poor correlate of protection (Hanekom et al., 2008), IL-2 ELISpot was included in the analysis as it might better mark immunological memory (Hawkridge et al., 2008; Millington et al., 2007; Soares et al., 2008). While a range of IFN-γ response to the antigens tested by in vitro screening was observed, the IL-2 response was more focused, as observed by others (Millington et al., 2007). The magnitude of IL-2 recall response to RD-1 encoded antigens was lower compared to the IFN-γ response. Apart from 3 EHR antigens

which induced an IL-2 response similar in magnitude and frequency to ESAT-6 and CFP-10, others were of lower magnitude and similar to that of Acr-1.

These findings support the hypothesis that it is possible, to an extent, to predict antigens using genomic data, provided combinations are taken into account. This approach may not only hold true for *Mycobacterium tuberculosis*, but for other pathogens as well. This study identified novel immunodominant molecules from the enduring hypoxic response, that could be potentially further characterized, to better understand the mechanism and importance of hypoxia *in vivo*. However, this data do not provide strong support, the hypothesis that a differential pattern of immune response is associated with latency as antigens that are targets of the immune response in latent infection were most often targets in active disease and *vice versa*. This is highly consistent with the evolving view of tuberculosis as a spectrum of infection rather than an arbitrary division of active and latent disease forms (Barry et al., 2009; Young et al., 2009). General support for the importance of hypoxia *in vivo* does however come from the fact that so many of the EHR genes encode potent T cell antigens.

Chapter 6: Evaluation of hypoxia inducible *Mycobacterium tuberculosis* specific proteins absent from BCG vaccine strains

Abstract

M. tuberculosis (MTB) species-specific antigenic determinants of the human T cell response are important for immunodiagnosis and vaccination. As hypoxia is a stimulus in chronic tuberculosis infection, I analyzed transcriptional profiles of MTB subject to 168 hours of hypoxia *in vitro*. I identified two region of difference (RD) 11, and one RD2, encoded proteins absent from all *M. bovis* strains including BCG (Rv2658c and Rv2659c) and from commonly used BCG strains (Rv1986) respectively.

In MTB infected persons the IL-2 ELISpot response to Rv1986 peptides was several times greater than the corresponding IFN-γ response to the recognized immunodominant ESAT-6 or CFP-10 antigens. This IL-2 response was directed narrowly at two epitopic regions containing residues 61-80 and 161-180. The predominant T cells responsible for IL-2 production were CD4+CD27+CD45RA- single cytokine positive central memory T cells. The IL-2 ELISpot response to live MTB bacilli lacking Rv1986 was significantly lower than the response to wild type or mutant complemented with Rv1986. The IL-2 ELISpot response to Rv1986 was significantly lower in HIV-TB co-infected persons than in HIV uninfected persons and significantly increased during antiretroviral therapy. These findings illuminate the partial efficacy of currently used BCG vaccines and provide evidence for a clinical trial comparing BCG strains.

Introduction

One third of the world's population is considered latently infected: a reservoir from which active TB disease will continue to develop for the foreseeable future and thus forming a major obstacle to achieve global control. The existence of so many with latent infection highlights that the partially effective vaccine: *Mycobacterium bovis* Bacille Calmette Guérin (BCG) does not prevent such infection (Maartens and Wilkinson, 2007).

Species-specific genes of *Mycobacterium tuberculosis* are targets of great interest to improve vaccination *via* the selection of protective antigens, and to define immunodiagnostic candidates that enhance the specificity and sensitivity of the widely used purified protein derivative (PPD) based tuberculin skin test (TST).

A significant landmark in this respect was the discovery that an *M. tuberculosis* genomic region designated region of difference (RD) 1 is deleted from all strains of BCG and thereby partially accounts for the avirulence of the vaccine (Mahairas et al., 1996; Pym et al., 2002). RD1 encodes a pair of co-regulated secreted proteins (ESAT-6 and CFP-10) that are highly immunogenic (Berthet et al., 1998; Sorensen et al., 1995): restoration of these genes into BCG improves its vaccine efficacy (Pym et al., 2003). Assay of the T cell interferon (IFN)-γ secretion in response to the combination of ESAT-6 and CFP-10 (interferon-γ release assays, IGRA) have been developed that have operational advantages and improve the specificity and possibly sensitivity of tuberculosis immunodiagnosis (Pai et al., 2008).

In addition, *M. tuberculosis* specific genes such as the insertion sequences, phage inserts may be an excellent source of candidate antigens for highly specific tuberculosis diagnostic test, in particular those which are also lacking in *M. avium* and other non-tuberculous mycobacteria. One such candidate RD11-encoded Rv2654c has been evaluated previously and is included in a commercial IGRA assay and also described in Chapter 4. Few other RD11 encoded antigens

have been described to be highly induced when MTB was subject to hypoxia, nutritional starvation and high temperature.

The availability of the complete sequence of *M. tuberculosis* also permitted further genomic characterization of various BCG strains (Behr et al., 1999; Brosch et al., 2007; Cole et al., 1998). It became apparent that, against a background of accumulating single nucleotide polymorphisms, BCG underwent sequential genomic deletions that thereby characterize various strains. The strains most commonly in use such as BCG Glaxo, Danish and Pasteur have most deletions. This led to the proposal that one of the reasons behind the partial vaccine efficacy of BCG was that it had become too attenuated to successfully mimic natural MTB infection (Behr and Small, 1997). Some empirical evidence in humans favoring this hypothesis is provided by the finding that BCG Japan induced greater cytotoxic and T helper 1 responses in infants than Danish BCG (Davids et al., 2006). The largest difference between BCG Japan and BCG Danish is the presence of RD2 in the former but not the latter.

The discovery of immunodominant antigens in *M. tuberculosis* has hitherto largely been based on dominance in antibody responses that are neither the basis of protection against tuberculosis nor of IGRA. A more rational approach might be to relate what is highly expressed by bacilli *in vivo* or *in vitro* (and thereby potentially available as an antigen) as recently investigated in bovines (Sidders et al., 2008). In humans there has been investigation of proteins encoded by genes of the dosR regulon that is induced in axenic culture by hypoxia and by nitric oxide (Sherman et al., 2001), stresses that are considered relevant to bacilli in nature (MacMicking et al., 1997; Nicholson et al., 1996; Via et al., 2008). Analysis of selected dosR regulated proteins confirmed the immunodominance of α-crystallin 1 (Acr1) encoded by Rv2031 (Leyten et al., 2006; Lin et al., 2007; Wilkinson et al., 1998b), as well as potentially infection stage specific antigens (Agrewala and Wilkinson, 1998; Black et al., 2009; Schuck et al., 2009; Wilkinson et al., 2005a).

Informatic evaluation of 100 highly induced enduring hypoxic response genes as described in Chapter 5, revealed two species specific RD11-encoded (Rv2658c and Rv2659c) and one partially species specific RD-2 encoded (Rv1986) genes. Therefore, these genes were considered to be excellent candidates for the study of immunogenicity along with other well-characterized antigens (ESAT-6, CFP-10, Acr1) in humans with active and latent tuberculosis and their analysis will be discussed separately in this Chapter.

Results

Antigen selection

Cross-reference of genes with the greatest fold induction in hypoxic culture over 7 days (Rustad et al., 2008) described as the enduring hypoxic response of *Mycobacterium tuberculosis* with sequence databases revealed three RD encoded genes. Two were species-specific (Rv2658c and Rv2659c, both RD11 encoded), which are absent from all *M. bovis* and *M. bovis* BCG vaccine strains and one partially species-specific gene (Rv1986, RD-2 encoded), which is absent from commonly used *M. bovis* BCG vaccine strains. In addition, 3 well described antigens were included as internal control: species specific RD1-encoded CFP-10 and ESAT-6; and dosR encoded Acr-1, which was also present amongst the 100 highly induced genes on day 7 of hypoxic culture.

Analysis of transcripts levels in of bacilli exposed to 7 days of hypoxia

The fold induction and *sig*A normalized transcript intensity over a time course of 168 hrs hypoxia for these genes (and of Acr1, CFP-10 and ESAT-6) are shown in Figure 47 and Appendix B.5. Interestingly whilst the fold induction for the RD1 encoded genes fell, the normalized intensity remained at a similar absolute level to that of both EHR and the dosR regulated Acr1 gene.

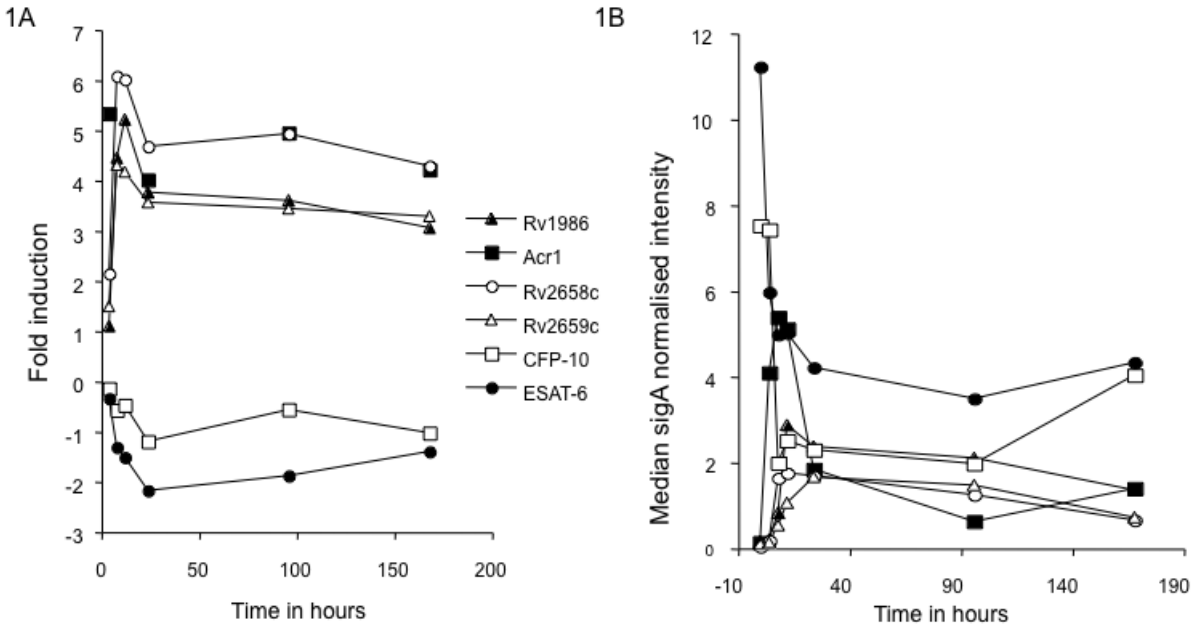


Figure 47 *M. tuberculosis transcript levels from bacilli exposed to 7 days hypoxia in vitro*

Panel A shows the fold induction with reference to the phase of aerobic cultures. There was a relative increase in the transcripts of Rv1986, Rv2658c, Rv2659c and Acr1 (Rv2031c) and a fall for those of Rv3874 and Rv3875 (CFP-10 and ESAT-6 respectively).

Panel B shows the median transcript intensity normalized to *sig*A. There is a very rapid peak in Acr1 levels followed by a decline; a sustained increase in Rv1986, Rv2658c and Rv2659c; and a fall in Rv3874 and Rv3875. However overall abundance is similar at 168 hours with Rv3874 and Rv3875 actually being highest.

Derived from the data obtained from (Rustad et al., 2008), Appendix B.5

Bioinformatic evaluation of Rv1986, Rv2658c and Rv2659c

Rv1986 is encoded by the RD2 region, which is deleted during the attenuation of *M. bovis* BCG strains. Rv1986 is a probable conserved integral membrane protein, of unknown function, although thought to be involved in transport of lysine across the membrane and belongs to the LYSE/YGGA family. It has 5 predicted trans-membrane segments as predicted by transMembrane prediction using Hidden Markov Models (TMHMM) as represented in Figure 48. TMHMM predicts transmembrane helices and location of the intervening loop regions.

Gene ontology analysis of Rv1986 suggests its involvement in:

- Biological process: aminoacid transport (Gene Ontology (GO):0006865), which is the directed movement of aminoacids, organic acids containing one or more amino substituents, into, out of, within or between cells;
- Cellular component: membrane (GO:0016020) and integral to membrane (GO:00162021);
- Molecular function: lysine permease activity (GO:0005293), which is the catalysis of the stereo-specific transfer of lysine, 2,6-diaminohexanonic, across a biological membrane.
- ♦ Rv1986 belongs to the functional classes of carrier activity (GO:0005386) and biological process (GO:0008150).

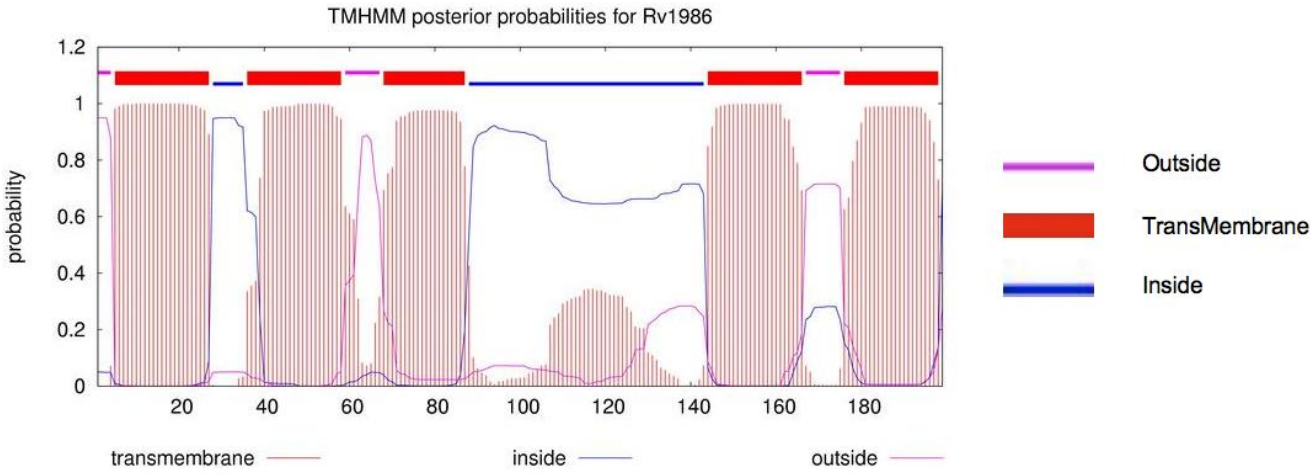


Figure 48 Predicted TransMembrane structure of Rv1986

Rv2658c and Rv2659c are encoded by phiRv2 phage insert in *M. tuberculosis* and therefore absent from other members of the MTB complex and other *Mycobacteria*. Rv2658c is of unknown function and in addition to hypoxia, is also highly induced after 4h, 24h and 96hours of starvation (Betts et al., 2002). There are no predicted trans-membrane segments and no gene ontology or functional classification predicted for this gene.

Rv2659c is a probable integrase, phiRv2 prophage protein, and putative member of the phage integrase family of tyrosine recombinases (Bibb and Hatfull, 2002; Hendrix et al., 1999). Functionally it is predicted to be involved in sequence integration. Integrases are necessary for the integration of a phage in to the host genome by site-specific recombination. In conjunction with excision with excisionase, the integrase is also necessary for the excision of the prophage from the host genome. Rv2659c in addition to hypoxia, is also highly induced under high temperature (Stewart et al., 2002) and up-regulated after 4h, 24h and 96hours of starvation (Betts et al., 2002).

Gene Ontology analysis of Rv2659c suggests its involvement in:

- Biological process: DNA recombination (GO:0006310) and DNA integration (GO:0015074)
- Molecular Function: DNA binding (GO:0003677).
- It belongs to the functional classes of nucleic acid binding (GO:0003676) and cellular process (GO:0009987). An ortholog of Rv2659c is present in *M. marinum* (MMAR_4848).

MHC class II binding peptide prediction

Analysis of MHC class II binding peptide prediction using ProPred reveled higher number of peptides binding to alleles DR1, DR3, DR4, DR7, DR8, DR11 and DR13. (Table 21)

Table 21 MHC class II binding peptides

	HLA alleles	DR1	DR3	DR4	DR7	DR8	DR11 (DR5)	DR13 (DR6)	DR15 (DR2)
(a) Allelic frequency in Xhosa Population	Allelic frequency (*af*)	0.0839	0.4193	0.1335	0.1429	0.0124	0.3137	0.3292	0.2174
(b)Number of peptides with >30% of the highest binding score *	Rv1986	6	29	32	4	28	25	39	8
	Rv2658c	1	11	11	2	3	2	9	2
	Rv2659c	3	14	19	10	21	25	28	12

Allelic frequency in Xhosa population (du Toit et al., 1988)

Number of peptides with >30% of the highest binding score any peptide can obtain for the particular allele.

*Not corrected for the number of allelic variants.

Empirical Evaluation

Human subjects

Patients with active or latent tuberculosis were recruited at the Ubuntu clinic at Khayelitsha site B, South Africa. All were of Xhosa ethnicity. Latent tuberculosis is described as a positive TST (≥15mm) or ≥20 SFC/10^6 PBMC IFN-γ response to ESAT-6 or CFP-10. Active tuberculosis is defined by a sputum smear or sputum culture positive for *M. tuberculosis*. The baseline characteristics of subjects enrolled to the study are shown in Table 22. There was a significant difference in the median age between the two clinical groups. However, this difference did not affect the magnitude or frequency of response.

Table 22 Demographic details of persons included in studies described

	Active tuberculosis	**Latent tuberculosis**	**Significance, p**	**HIV-1 Infected**	
Number	20	48		Number	19
Median age	33	21	<0.0001	Median age	35
Sex (M:F)	11M:9F	12M:36F		Sex (M:F)	10M:9F
Sputum smear positive	19 Positive 1 unknown	NA	NA	Median CD4 at recruitment (IQR)	90 (22-118)
Sputum culture positive	18 Positive 2 not available	NA	NA	Median CD4 after 6 months of cART (IQR)	209 (126-312)
BCG vaccinated	5 vaccinated 6 Not vaccinated 9 Unknown	23 vaccinated 20 Not vaccinated 5 Unknown		Median Viral Load at recruitment (IQR)	130,000 copies/ml (49,722-347,500)
Median Mantoux (IQR)	NA	11mm (IQR 8-20)	NA	Median Viral Load after 6 months of cART (IQR)	50 copies/ml (50-158)

Interferon-γ ELISpot analysis of active and latent tuberculosis

Interferon-γ (IFN-γ) ELISpot was performed using PBMC from 40 persons with active (20) or latent (20) tuberculosis, and IL-2 ELISpot on 13 and 14 persons in each category. Immunodominance was assessed both quantitatively (median SFC/10^6 PBMC) and by frequency of response (> 20 SFC/10^6 PBMC). The median background response for IFN-γ and IL-2 ELISpot were 6 SFC/10^6 PBMC and 8 SFC/10^6 PBMC respectively. Data presented after correction for background.

CFP-10 and ESAT-6 were co-dominant for the IFN-γ response when assessed by both methods (Figure 49 A and B). The largest SFC response in latent disease was to CFP-10 (102 SFC/10^6 PBMC, IQR 38-444). With the exception of ESAT-6 all other responses were significantly lower ($p \leq 0.007$). The largest response in active disease was to ESAT-6 (172 SFC/10^6 PBMC, IQR 47-423). With the exception of CFP-10 all other responses were significantly lower ($p \leq 0.002$).

Although peptide pool, Rv2659c-2 was preferentially recognized by latently infected persons (6 SFC/10^6 PBMC, IQR 1-28 versus 0 SFC/10^6 PBMC, IQR 0-7, $p = 0.028$) these responses were very modest.

When analyzed by the frequency of response, no pool was preferentially recognized by either clinical group (Figure 49 B). The most frequent response in the combined group (latent plus active) was to CFP-10 (36/40, 90%): with the exception of ESAT-6 the proportion of persons responding to the other antigens was lower in every case ($p \leq 0.002$).

IL-2 ELISpot analysis of active and latent tuberculosis

Patients with active tuberculosis preferentially recognized pooled peptides from Rv2658c (7 IL-2 SFC/10^6 PBMC, IQR 1-23 versus 0 SFC/10^6 PBMC, IQR 0-6 p = 0.042). However, when analyzed by proportion of response, no pool was preferentially recognized by either clinical group (Figure 49 C and D).

There was however a striking IL-2 response in both active and latent disease to Rv1986 pool 2 (795 SFC/10^6 PBMC, IQR 51-1428 in active infection; 1194 SFC/10^6 PBMC, IQR 862-1650 in latent infection, Figure 49 C and D). All other antigen specific IL-2 responses were significantly lower in both latent ($p \leq 0.0007$) and active infection ($p \leq 0.02$) when compared to Rv1986 (Figure 49 C and D).

In persons with latent tuberculosis, peptide pools covering the amino acid region 1-130 (peptide pool 1) and covering region 121-250 (peptide pool 2) of Rv2659c induced similar levels of IL-2 (median of 76 SFC/10^6 PBMC, IQR 1-206 and 76 SFC/10^6 PBMC, IQR 0-358). These median levels were higher than those of CFP-10 (3 SFC/10^6 PBMC, IQR 3-194) or ESAT-6 (62 SFC/10^6 PBMC, IQR 7-15). However, in those with active tuberculosis, median levels of IL-2 induced by peptide pools 1 (21 SFC/10^6 PBMC, IQR 1-186) and 2 (14 SFC/10^6 PBMC, IQR 0-409) of Rv2659c were lower than those of CFP-10 (40 SFC/10^6 PBMC, IQR 14-158) and ESAT-6 (58 SFC/10^6 PBMC, IQR 22-163) (Figure 49 C and D).

The most frequent response in the combined group (latent plus active) was also to Rv1986-2 (24/26, 92%): with the exception of ESAT-6 and CFP-10, the proportion of subjects responding to the other antigens was lower in every case ($p \leq 0.009$). (Figure 49 D).

Figure 49

Interferon-γ and IL-2 ELISpot response to peptide pools

Panel A shows the spot forming cells (SFC) in patients with latent (L) and active (A) tuberculosis. The strongest response in latent disease was to CFP-10. With the exception of ESAT-6 all other responses were significantly lower ($p \leq 0.007$). The strongest response in active disease was to ESAT-6. With the exception of CFP-10 all other responses were significantly lower ($p \leq 0.002$). Peptide pool Rv2659c-2 was preferentially recognized by latently infected persons ($p = 0.03$).

Panel B shows the proportion of responders (defined by a response of ≥ 20 SFC/106 PBMC above background). No pool was preferentially recognized by either clinical group. The most frequent response in the combined group (latent plus active) was to CFP-10: with the exception of ESAT-6 the proportion subjects responding to the other antigens was lower in every case ($p \leq 0.002$).

Panel C shows IL-2 spot forming cells (SFC). The strongest response in both active and latent disease was to Rv1986-2. All other responses in both latent ($p \leq 0.0007$) and active infection ($p \leq 0.02$) were significantly lower. Peptide pool Rv2658c was preferentially recognized by actively infected persons ($p = 0.042$).

Panel D shows the proportion of responders. No pool was preferentially recognized by either clinical group. The most frequent response in the combined group (latent plus active) was to Rv1986-2: with the exception of ESAT-6 and CFP-10 the proportion subjects responding to the other antigens was lower in every case ($p \leq 0.009$).

Lines indicate the median response.

Peptide mapping of the IL-2 response to Rv1986

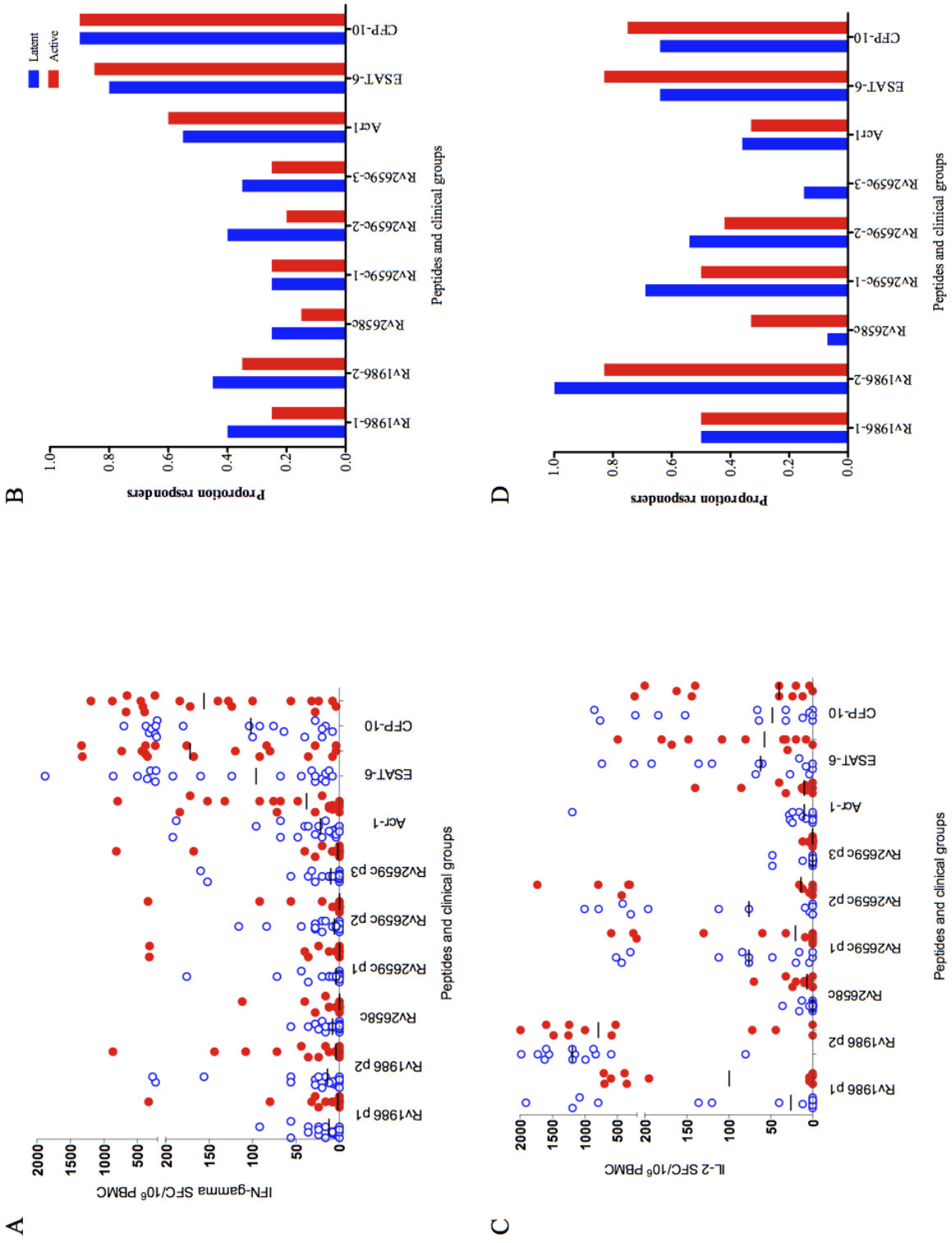


Figure 49 Interferon-γ and IL-2 ELISpot response to peptide pools

single peptide giving rise to a response in > 50% subjects (Lalvani et al., 2001a; Pathan et al., 2001); similar findings are reported for other antigens of *M. tuberculosis* (Wilkinson et al., 2005a), I were therefore interested to determine whether a similarly 'degenerate' pattern of multiple IL-2 inducing epitopes occurred in Rv1986.

PBMC from 20 persons with latent tuberculosis were assayed in the presence of 10 µg/ml of each peptide or no stimulus. A highly focused pattern of dominance was observed with peptides p61-80 (84 SFC/10^6 PBMC, IQR 56-134) and p161-180 (68 SFC/10^6 PBMC, IQR 49-104) being clearly the best recognized (Figure 50). 90% of subjects had a response > 20 SFC/10^6 PBMC to p61-80 and 95% to p161-180 with no other peptide being recognized by > 45% subjects. There were less frequent and lower magnitude responses to p151-170 perhaps suggesting the epitope core for some MHC Class II molecules may include residues common to both peptides (161-170).

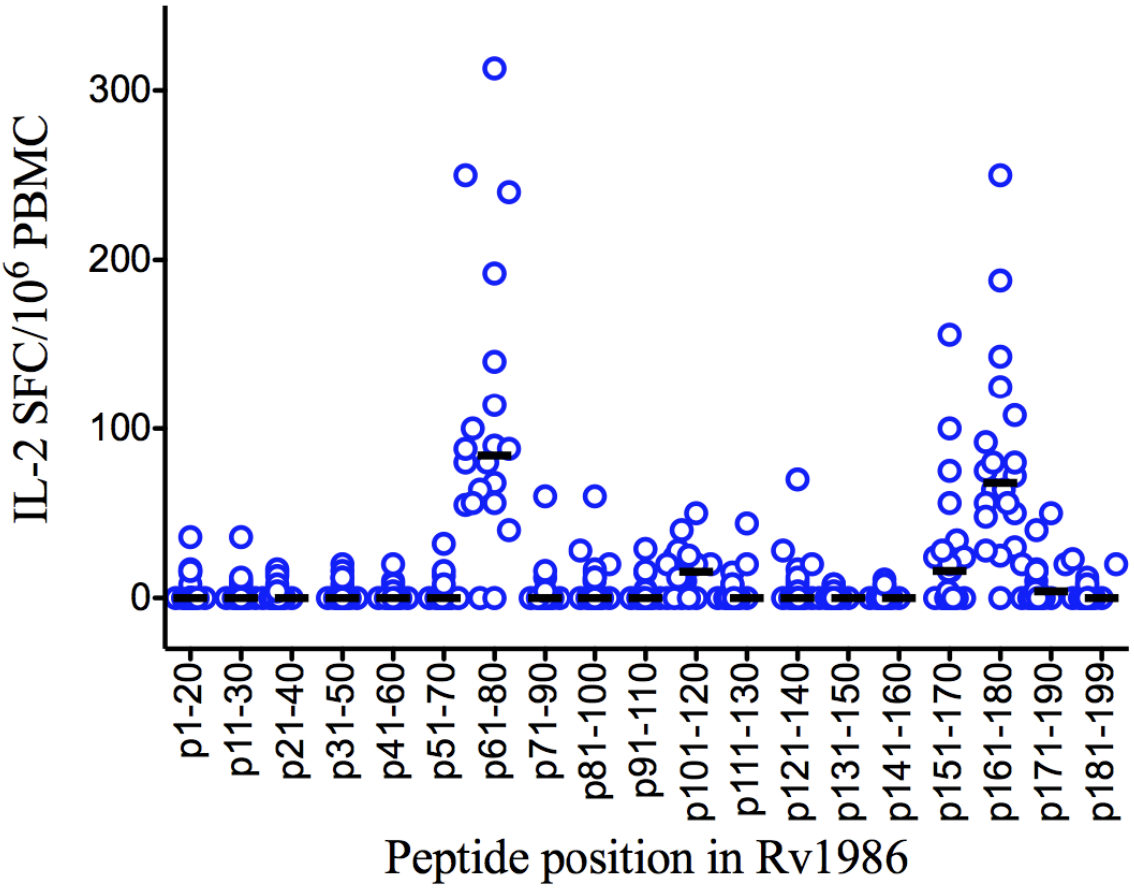


Figure 50 Peptide map of IL-2 response to Rv1986

PBMC from 20 subjects with latent tuberculosis were set up in an IL-2 ELISpot assay with individual peptides. p61-80 and p161-180 were the dominant determinants of the response. Lines indicate median responses.

Phenotype of CD4+ T cells responding to Rv1986

Analysis of the T cells responsible for type 1 cytokine responses is critical to understand protective immunity against TB (Soares et al., 2008). In PBMC from 5 persons with latent tuberculosis, the phenotype of CD4+T cells responsible for type 1 cytokine (IFN-γ, IL-2 and TNF) production when stimulated with the peptides of Rv1986 (61-80 and 161-180) or the combination of peptides from CFP-10 and ESAT-6 as a comparison were determined. T cell phenotypes were defined based on the surface markers CD45RA and CD27: Central memory cells (TCM) positive for CD27 and negative for CD45RA; effector memory (TEM) negative for both CD27 and CD45RA and Terminally differentiated T cells (Tdiff) negative for CD27 and positive for CD45RA. Single cytokine positive cells predominated overall (Figure 51). Most IL-2 derived from TCM irrespective of stimulus. The two Rv1986 peptides restimulated nearly ten times the percentage of IL-2 producing TCM cells than CFP-10 and ESAT-6 (median: 0.226 % CD3+CD4+ TCM vs. 0.024% CD3+CD4+TCM, p=0.055, Figure 51 panel A and B).

Figure 51

Phenotype of CD4+ T cells responding to Rv1986

PBMC from 5 donors were stimulated either with CFP-10+ESAT-6 (A) or with Rv1986 p61-80+p161-180 (B), and phenotype of T cells producing single and multiple cytokines were analyzed using surface and intracellular cytokine staining by FACS. T cell phenotype were defined based on the surface markers CD45RA and CD27: Central memory cells (TCM) positive for CD27 and negative for CD45RA; effector memory (TEM) negative for both CD27 and CD45RA and Terminally differentiated T cells (Tdiff) negative for CD27 and positive for CD45RA. The results are expressed as the percentage of CD3+ CD4+ T cells. The stimuli induced different populations of T cells. The strongest response to Rv1986 p61-80+p161-180 stimuli was the IL-2 producing central memory cells, median 0.22% (Panel B).

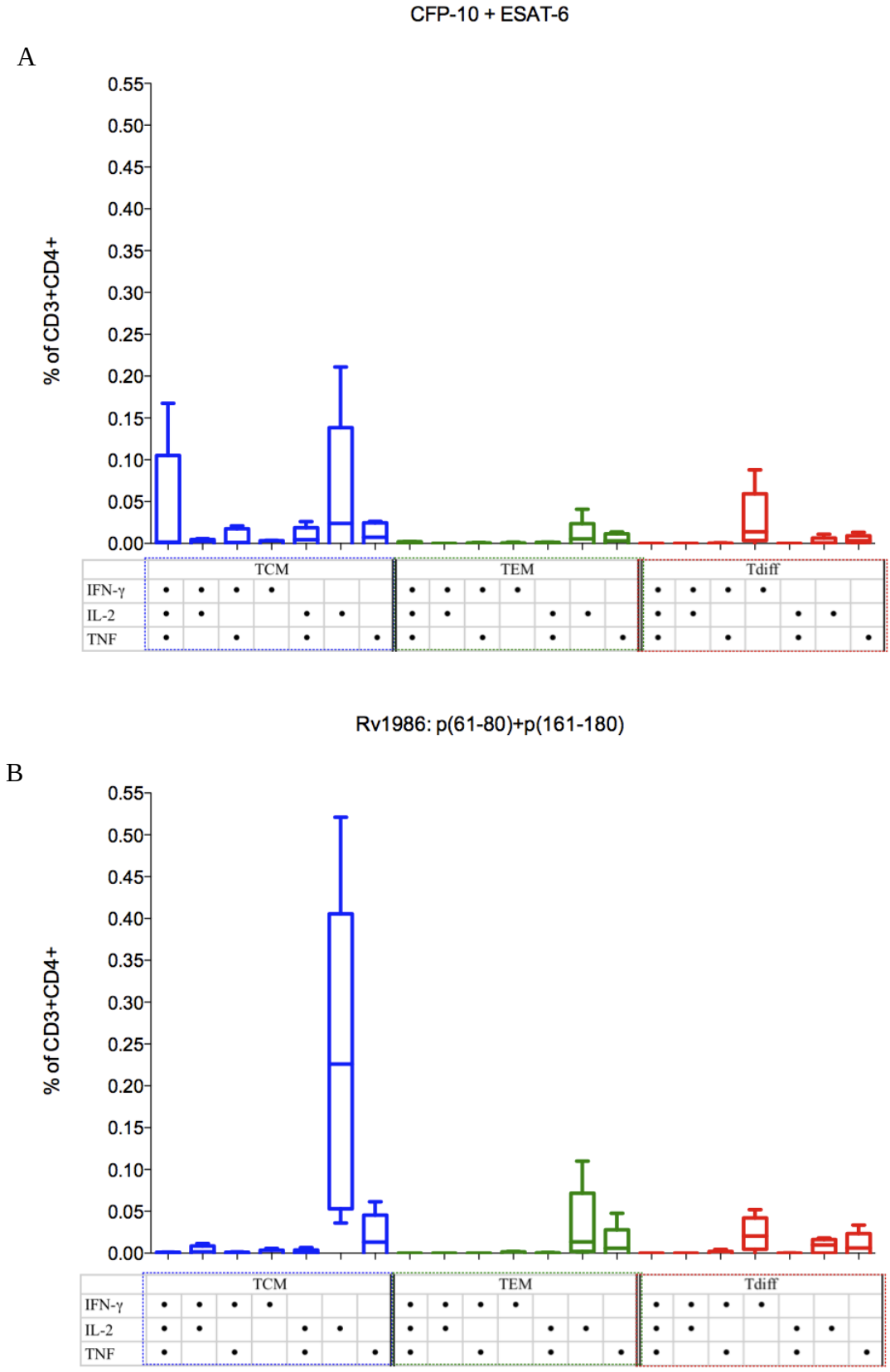


Figure 51 CD4+T cell phenotype and type-1 cytokine response to CFP-10 plus ESAT-6 and Rv1986

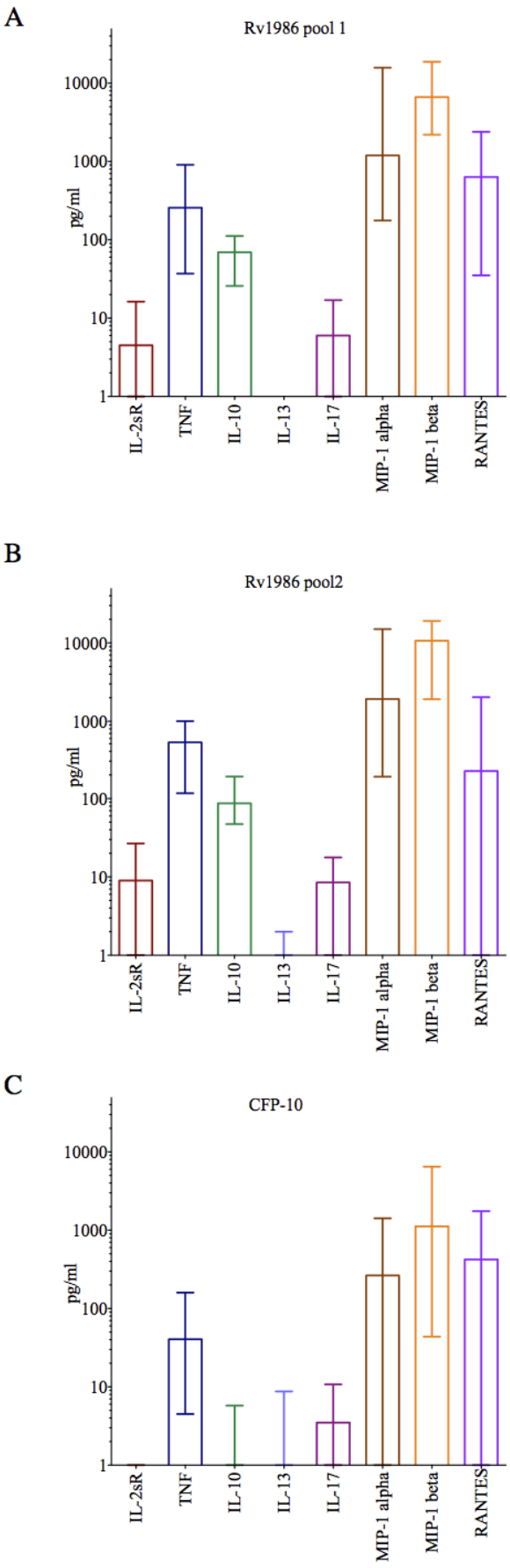


Figure 52 Luminex multiplex cytokine analysis for Rv1986 and CFP-10 peptide pools

Panels A, B and C shows the production of different cytokines and chemokines in 39 persons with latent or active tuberculosis after 16 hrs stimulation of PBMC with pooled peptides of Rv1986 (1 and 2) or CFP-10 in cell culture supernatants respectively. The results are represented as pg/ml; bars indicate median response with IQR. The cytokine responses to Rv1986 (pool 1 and 2) were significantly higher than that of CFP-10 for IL2sR ($p<0.0001$), TNF ($p<0.0001$), IL-10 ($p<0.0001$), IL-13 ($p\leq0.01$), IL-17($p=0.01$), MIP-1alpha ($p \leq 0.005$) and MIP-1beta ($p<0.004$), but not RANTES ($p>0.5$).

Rv1986 induces a distinct pattern of cytokine secretion in addition to IL-2

The ability of Rv1986 to induce the secretion of cytokines other than IFN-γ and IL-2 when compared to CFP-10 was also investigated. 16-hour cell culture supernatants from 39 persons, with either active (19) or latent (20) tuberculosis were used for this analysis. Multiple cytokine secretion was assessed both quantitatively (pg/ml, after background correction) and by frequency of response (> 2 fold above background). Similar levels of cytokine responses were observed in both analyses in persons with active and latent tuberculosis, (data not shown), therefore the clinical groups were combined for further analysis. When analyzed quantitatively, Rv1986 pool 1 and 2 stimulated significantly higher levels of IL2sR, TNF, IL-10, IL-13, IL-17, MIP-1 alpha and MIP-1 beta than CFP-10, and similar levels of RANTES. Levels of IL-13 were very modest (Figure 52 A, B and C). Thus the Rv1986 peptides were associated with a distinct pattern of cytokine production other than IL-2 when compared to CFP-10.

T cell recognition of M. tuberculosis H37Rv with and without Rv1986

The recognition of Rv1986 in the context of the whole bacilli was determined using IL-2 and IFN-γ responses to live strains of MTB in which Rv1986 was intact (MTB H37Rv) or deleted (H37RvΔRD-2) or complemented (H37RvΔRD-2::Rv1986). 13 persons with latent tuberculosis were tested (only 9 patients for IFN-γ due to limitation in cell numbers). Although the overall IFN-γ SFC responses to these MTB strains were much stronger than the IL-2 response, there was no significant difference in IFN-γ response between these strains (Figure 53 B). The IL-2 SFC response to MTB H37Rv was significantly higher than to the H37RvΔRD-2 mutant (median 228 SFC/10^6 PBMC, IQR 142-325 *vs.* 130 SFC/10^6 PBMC, IQR 53-268; p= 0.002) and complementation by Rv1986 alone substantially restored the SFC response (183 SFC/10^6 PBMC, IQR 86-285; p= 0.002, when compared to H37RvΔRD-2. Figure 53 A).

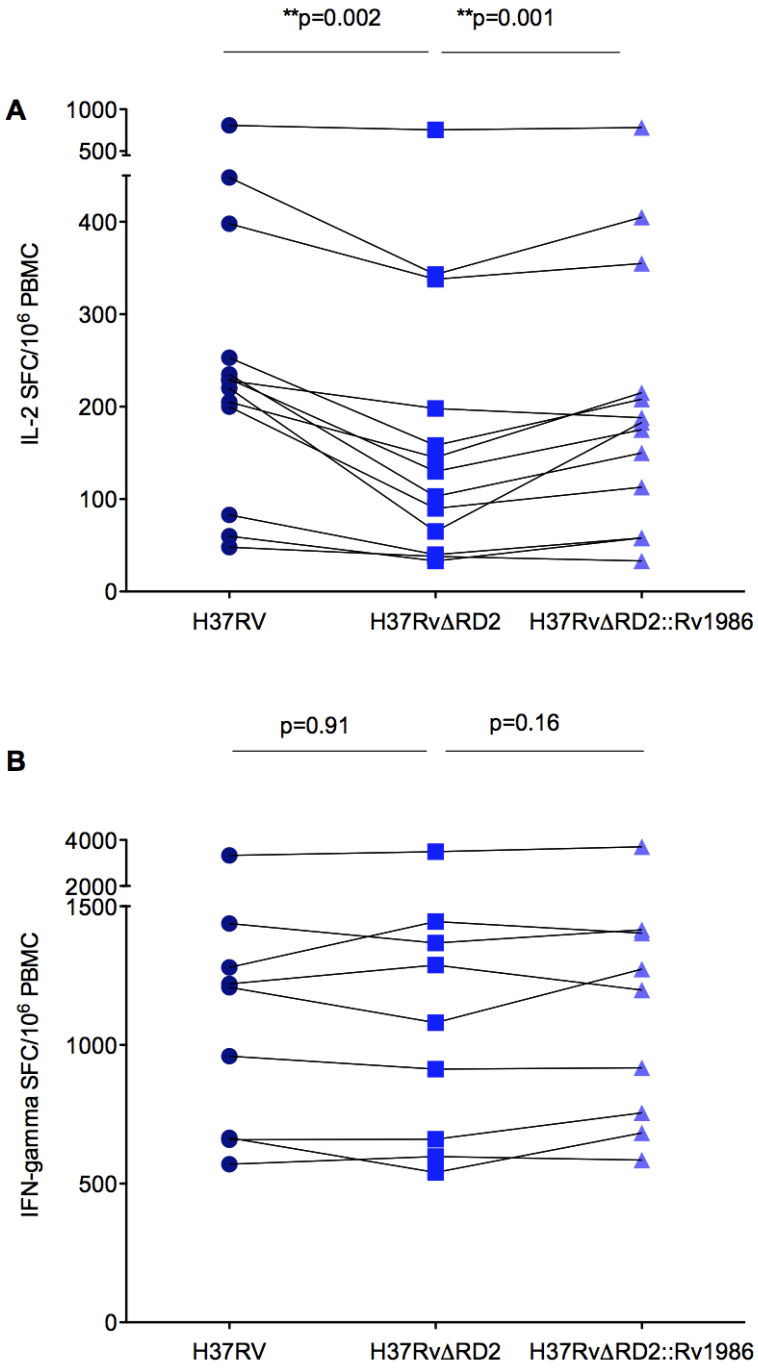


Figure 53 *Live recognition of M.tuberculosis H37Rv,* H37RvΔRD2 *and* H37RvΔRD2::Rv1986

PBMC from persons with latent tuberculosis were co-cultured with *M.tuberculosis* H37Rv (λ), H37RvΔRD2 (ν) and H37RvΔRD2::Rv1986 (σ) (complemented with Rv1986) for IL-2 and IFN-γ ELISpot. Results are represented as SFC/10^6 PBMC.

Panel A shows IL-2 SFC response. All 3 strains induced substantial IL-2 response (>30 SFC/10^6 PBMC) in all donors. The median IL-2 SFC to MTB H37Rv was highest, followed by H37RvΔRD2::Rv1986 and the H37RvΔRD2 mutant.

Panel B shows IFN-γ SFC response, which was much stronger than the corresponding IL-2 response. All 3 strains induced similar levels of IFN-γ response.

Previous studies have described an increase in the IFN-γ response when the MTB strains were subject to treatment *in vitro* with INH (Wilkinson et al., 2006). In order to investigate whether the same effect is observed with respect to the IL-2 response, the panel was extended to study strains of MTB (in which Rv1986 was intact or deleted) that were subject to different conditions namely: INH treatment, heat kill and heat kill with INH.

	Live median (IQR)	Live + INH median (IQR)	Heat Killed median (IQR)	Heat Killed + INH median (IQR)
H37Rv	222 (142-325)	295 (190-378)	195 (95-246)	218 (102-276)
H37RvΔRD2	130 (53-268)	230 (125-313)	158 (46-246)	170 (74-222)
H37RvΔRD2::Rv1986	183 (85-285)	240 (162-353)	183 (65-254)	157 (53-257)
p values (Wilcoxon signed rank paired t test)				
H37Rv vs. H37RvΔRD2	**0.002	**0.002	**0.001	*0.033
H37RvΔRD2 vs. H37RvΔRD2::Rv1986	**0.001	*0.021	*0.033	0.846

Table 23 IL-2 response to different strains of MTB with and without treatment

MTB H37Rv strain induced higher levels of IL-2 in all conditions (Live+INH, HK and HK+INH) than MTB H37RvΔRD2 (without Rv1986) and MTB H37RvΔRD2::Rv1986 (with Rv1986) and followed similar trend as that of the live MTB strains. With the exception of Heat Killed strains with INH treatment, both Live+INH and Heat Killed H37RvΔRD2::Rv1986 stains induced significantly higher level of IL-2 when compared to H37RvΔRD2 subject to same treatment (Table 24).

The effects of INH treatment and Heat killing of MTB stains on the resultant IL-2 response were also compared in MTB H37Rv. INH treatment significantly increased the IL-2 response ($p \leq 0.002$) irrespective of the strain used. However there was no significant difference observed in the response between live and heat killed or heat killed with INH treatment ($p \geq 0.17$) and also between Heat killed and heat killed with INH treatment. Overall, comparable levels of IL-2 were induced by MTB strains subject to heat killing with or without INH as that those without any treatment, with the exception of INH treatment to live strains (Figure 54). These analysis suggest that INH, which acts by disturbing the cell wall biosynthesis of actively dividing bacilli, contributes to the increased availability of cell wall related antigenic targets. These antigenic targets thus would be more accessible to the circulating antigen specific T cells in the peripheral blood and therefore results in the increased T cell response. Similar pattern of response were observed in those strains with and without Rv1986.

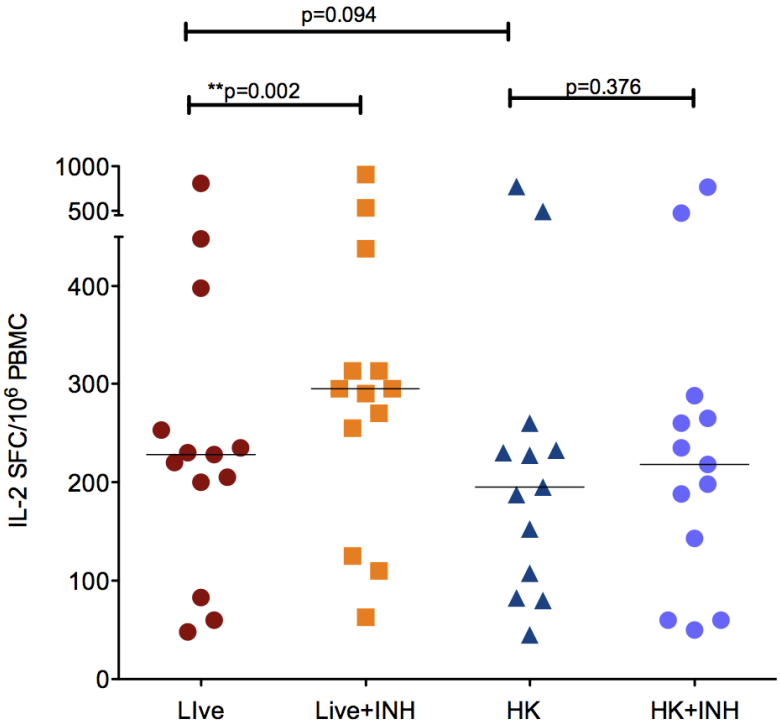


Figure 54 Effect of treatment on MTB H37Rv

PBMC from persons with latent tuberculosis were co-cultured with *M. tuberculosis* H37Rv either as live with (ν) or without (λ) INH treatment or heat killed (HK) with (λ) or without (σ) INH treatment.

Significant increase in the median response was observed between live strain without treatment and with treatment (p=0.002). However, there was no increase in the IL-2 response between Live and Heat killed (p=0.094) or heat killed with INH treatment (p=0.376).

Bars indicate median response.

Response of HIV-1 infected persons to Rv1986, ESAT-6 and CFP-10

The CD4 deficiency caused by HIV-1 infection is the greatest recognized predisposing factor to tuberculosis and conversely combined antiretroviral therapy (cART) reduces susceptibility by suppressing viral replication and allowing CD4 recovery (Wilkinson et al., 2009). I reasoned it would therefore be of interest to compare the IL-2 response to Rv1986 and IFN-γ and IL-2 response to CFP-10 and ESAT-6 before and during the course of antiretroviral therapy. As the IFN-γ response to Rv1986 had not been prominent in HIV-1 uninfected persons this was not assayed. The ELISpot response of 19 HIV-1 infected persons without evidence of active tuberculosis was therefore tracked longitudinally over the first 36 weeks of therapy. All patients experienced CD4 increases and suppression of HIV-1 replication during cART. I could not sample all time points and patients for both cytokines due to limitation in the number of cells. Figure 55 shows results of patients whose IL-2 and IFN-γ response to CFP-10 and ESAT-6 was assayed at least twice and 9 patients in whom the corresponding IL-2 response to peptides p61-80 and p161-180 could be determined.

Peptide responses were summed for analysis and compared to the values obtained from 20 HIV-1 uninfected persons of similar background, age and sex (i.e. those shown in Figure 50). The IL-2 response to the peptides of Rv1986 was significantly lower in HIV-1 infected persons prior to cART (median 24, IQR 11-43) than in HIV-1 uninfected persons (median 160, IQR 114-256, p = 0.009, Figure 55A). A significant increase in response occurred during cART therapy such that the median at 36 weeks increased to 106 (IQR 79-157, p = 0.005). By contrast the IFN-γ response to ESAT-6 and CFP-10 was not significantly lower in HIV-1 infected persons prior to cART (median 147, IQR 50-965) than in HIV-1 uninfected persons (232, IQR 56-563, p = 0.84). Whilst the median response did increase during cART therapy, the overall trend was not significant p = 0.22, Figure 50 B). The IL-2 response to the peptides of ESAT-6 and CFP-10 was significantly lower in

HIV-1 infected persons prior to cART (median 2, IQR 0-31) than in HIV-1 uninfected persons (median 148, IQR 44-323, p = 0.02, Figure 55 C). No significant increase in response occurred during cART therapy. Taken together these findings indicate the decreased IL-2 to Rv1986 response prior to therapy correlates with increased susceptibility better than the IFN-γ response to CFP-10/ESAT-6; and that the partial but significant recovery of IL-2 to Rv1986, but unchanged IFN-γ response to CFP-10/ESAT-6 also correlates with the recognized decrease in tuberculosis susceptibility that is conferred by cART.

Figure 55 Response of HIV-1 infected persons to Rv1986, ESAT-6 and CFP-10

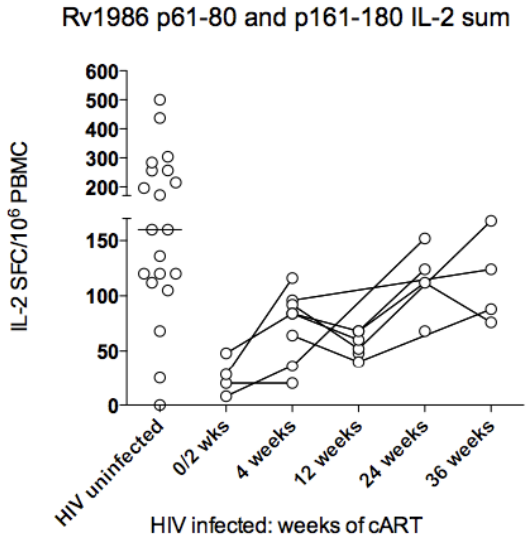
The IL-2 response to Rv1986 and IFN-γ response to CFP-10 and ESAT-6 before and during the course of antiretroviral therapy was compared. The respective responses were summed for analysis and compared to the values previously obtained for HIV-1 uninfected persons of similar background, age and sex.

A The IL-2 response to the peptides of Rv1986 was significantly lower in HIV-1 infected persons prior to cART than in HIV-1 uninfected persons. A significant increase in response occurred during cART therapy (p = 0.005 Kruskal-Wallis test).

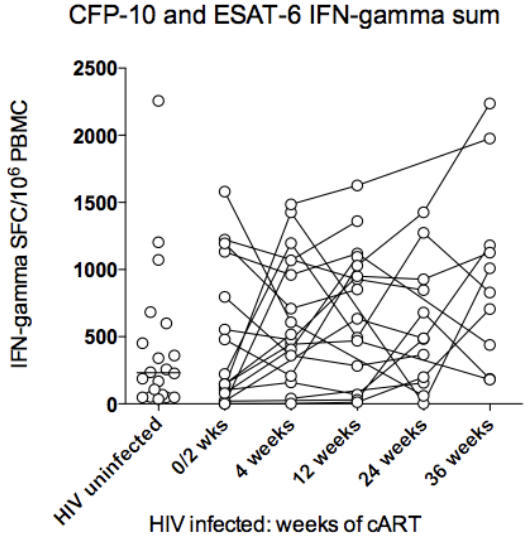
B The IFN-γ response to ESAT-6 and CFP-10 was not significantly lower in HIV-1 -1 infected persons prior to cART than in HIV-1 uninfected persons. The trend in median response during cART therapy was not significant (p = 0.22).

C The IL-2 response to the peptides of ESAT-6 and CFP-10 was significantly lower in HIV-1 infected persons prior to cART than in HIV-1 uninfected persons (p = 0.02). No significant increase in response occurred during cART therapy.

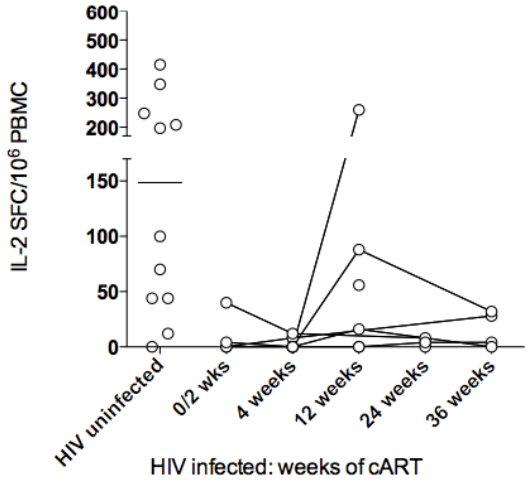
A


Rv1986 p61-80 and p161-180 IL-2 sum
IL-2 SFC/10^6 PBMC
600
500
400
300
200
150
100
50
0
HIV uninfected
0/2 wks
4 weeks
12 weeks
24 weeks
36 weeks
HIV infected: weeks of cART

B


CFP-10 and ESAT-6 IFN-gamma sum
IFN-gamma SFC/10^6 PBMC
2500
2000
1500
1000
500
0
HIV uninfected
0/2 wks
4 weeks
12 weeks
24 weeks
36 weeks
HIV infected: weeks of cART

C


CFP-10 and ESAT-6 IL-2 sum
IL-2 SFC/10^6 PBMC
600
500
400
300
200
150
100
50
0
HIV uninfected
0/2 wks
4 weeks
12 weeks
24 weeks
36 weeks
HIV infected: weeks of cART

Discussion

Whole genome-based transcriptional profiles of *M. tuberculosis* subject to prolonged hypoxia were analyzed to guide the discovery of potential antigens. Because the diagnostic potential of species-specific proteins is greatest, two genes upregulated during hypoxia that are absent from all *M. bovis* strains including BCG by virtue of being RD11-encoded (Rv2658c and Rv2659c) (Cole, 2002; Gordon et al., 2001) were considered for further investigation. The RD2-encoded Rv1986 was also investigated because it is absent from most commonly used BCG strains.

When compared to the well-characterized immunodominant and species-specific molecules ESAT-6 and CFP-10, RD11 proteins had inferior ability to restimulate IFN-γ from T cells of persons with either latent or active tuberculosis. However a striking finding was the immunodominance of Rv1986 for the IL-2 recall response, directed narrowly at two epitopic regions.

The quantitative IL-2 response to Rv1986 was several times greater than the corresponding IFN-γ response to either ESAT-6 or CFP-10 (Figure 43 C and D). FACS analysis suggested Rv1986 to be a major target of long lived CD4+ central memory T cells. Further, Rv1986 peptides are associated with a distinct pattern of cytokine production when compared to CFP-10. There was significant recovery of IL-2 response to the peptides of Rv1986 than of IFN-γ response to ESAT-6 or CFP-10 during the course of cART in HIV-1 infected persons. Moreover, deletion of Rv1986 from the genome of *M. tuberculosis* substantially decreased its ability to restimulate IL-2 secretion. These interesting findings are potentially important when considering vaccine-induced and natural immunity to tuberculosis and how immunodiagnosis may be improved. In addition the persons included in this study would have received BCG Japan, (vaccine strain used before 1998), which does contain RD2.

One hypothesis to be tested was whether, by virtue of upregulation during hypoxia, proteins encoded by such genes would be preferentially recognized by latently infected persons. Rv2658c and Rv2659c in addition to being hypoxia inducible are also highly induced in nutritional starvation (Betts et al., 2002) and Rv2659c under high temperature (Stewart et al., 2002). With the exception of the weak IFN-γ response to Rv2659c pool 2 (Figure 43A), preferential recognition proved not to be the case. In contrast, Rv2658c was preferentially recognized in those with active tuberculosis by a very weak IL-2 response (Figure 43C). Overall, when compared to CFP-10 and ESAT-6, Rv2658c had inferior ability to restimulate either IFN-γ or IL-2 response. Rv2659c although stimulated only moderate IFN-γ response, had the ability to induce higher levels of IL-2 than CFP-10 and ESAT-6 in those with latent tuberculosis.

Hypoxia does characterize tuberculous granulomas *in vivo* (Via et al., 2008) but it is increasingly re-appreciated that both active and latent tuberculous lesions exhibit a dynamic spectrum of overlapping morphologies (Barry et al., 2009; Capuano et al., 2003; Kaplan et al., 2003; Lin et al., 2006; Timm et al., 2003; Ulrichs et al., 2005; Young et al., 2009) and that hypoxic lesions likely occur in both clinical circumstances. A dominant antigenic target that is dosR regulated is Acr1 encoded by Rv2031c, and documented preferential T cell recognition of Acr1 by latently infected people (Geluk et al., 2007; Wilkinson et al., 1998b). Preferential recognition of Acr1 by latently infected persons was not observed in this study (Figure 43) nor in previous IFN-γ ELISpot analysis in this laboratory (Wilkinson et al., 2005a), which is in fact consistent with expression of Acr1 throughout experimental infection (Shi et al., 2003; Stewart et al., 2005).

Although IFN-γ is essential to human defense against mycobacteria, it is increasingly recognized that assay of PBMC secretion of IFN-γ is a poor correlate of protection in field studies of tuberculosis (Hanekom et al., 2008). Greater attention to markers, such as IL-2,

that might better reflect immunological memory is now being paid and formed the basis for this investigation by ELISpot assay of this cytokine (Hawkridge et al., 2008; Millington et al., 2007; Soares et al., 2008). Nevertheless, the finding that Rv1986 was so dominant for the IL-2 response yet elicited moderate IFN-γ secretion was striking.

The cytokine phenotype of antigen specific T cells is greatly influenced by co-stimulation and the cytokine milieu (Hsieh et al., 1993; McKnight et al., 1994). However it has also been suggested that the overall affinity of the TcR-peptide-MHC interaction may play a role as well (Agrewala and Wilkinson, 1999; Murray et al., 1992). Interestingly an epitope in Rv1986 with an anchor at position 167 (corresponding with p161-180) is predicted for several DRB1*03, *04, *08, *11 and *13 alleles (Singh and Raghava, 2001): a promiscuous binding ability that has been noted for other *M. tuberculosis* epitopes (Agrewala and Wilkinson, 1999) and which might contribute to the almost universal response, I observed to this peptide. TMHMM prediction (Figure 42) for Rv1986, predicted two regions to be present outside of the membrane, which also corresponds to the dominant epitope regions.

Rv1986 is a putative membrane protein that is recognised by T cells from *M. bovis* infected cows (Cockle et al., 2002). Although the responses to other RD2- encoded antigens has been previously evaluated in humans (Chen et al., 2009; Fu et al., 2009; Grover et al., 2006; Kalra et al., 2009), the human T cell response to Rv1986 has not been reported. Finding that Rv1986 is a dominant target of IL-2 secreting memory T cells suggests that this recall response could contribute to protective immunity as suggested by the preliminary findings in the HIV-1 infected people on ART (Figure 50).

These findings also bring a novel twist to an old story: the partial and globally variable efficacy of BCG vaccine (Bloom and Fine, 1994; Trunz et al., 2006). A recent study noted no experimental difference in protective efficacy in Guinea Pigs between RD2- negative

(e.g. BCG Pasteur) and RD2-positive (e.g. BCG Japan) strains (Horwitz et al., 2009) but the discovery that a major target of the human IL-2 response is absent from the most commonly used strains is intriguing. Whilst the *in vitro* diagnostic potential of the two dominant peptides presented here is considerable, the most important consequence of this work may be to re-evaluate by clinical trials whether BCG strains with and without RD2 vary in clinical efficacy.

Chapter 7: Conclusions

The overall aim of the studies presented in this thesis was evaluation of novel *Mycobacterium tuberculosis* antigens by a whole genome based strategy, whereby genes highly expressed by the bacilli *in vitro* or *in vivo* were related to what is recognized by human T cells as an antigen. This was achieved by the analysis of whole genome based, transcriptional profiles of *Mycobacterium tuberculosis* subject to 7 days of hypoxia (enduring hypoxic response) using bioinformatic analysis and empirical testing to guide antigen discovery.

The hypotheses tested in this study include: (i) the enduring hypoxic response genes are expressed *in vivo* and thereby available as antigenic targets for the T cell mediated immune response, (ii) by virtue of upregulation under hypoxia, these antigens are preferentially recognized by the T cells during latent tuberculosis, (iii) genomic information from a pathogen might aid in antigen prediction, and (iv) species specific genes of *Mycobacterium tuberculosis* are potential diagnostic targets.

In addition, within this context, a preliminary comparison of various techniques for evaluating the human T cell response (Chapter 3), and T cell epitope mapping using synthetic overlapping peptides of the previously described RD11-encoded Rv2654c and dosR encoded Rv1733c (Chapter 4) were also performed.

A systematic Immuno informatic approach was designed to address the over all aim (Chapter 5 and 6). The approach included analysis of transcriptional profiles of MTB subject to 7 days of hypoxia, coupled with the use of *in silico* tools to search for promising proteins with epitopes that are highly predicted to bind to MHC II molecules, and *in vitro* screening of selected shortlisted candidates. In addition, the MHC II prediction algorithm was complemented with a correction for the HLA allelic

frequency (*af*) present in the South African Xhosa population, to account for the HLA polymorphism that could influence the T cell response. Selection of candidates was also based on including those with high as well as low bioinformatic ranks, in order to relate the hierarchies of bioinformatic and empirical testing of these proteins (Chapter 5).

Overall, 29 antigen were selected and evaluated in two groups: non- MTB species specific (n=26) and MTB species specific (n=3) by IFN-γ and IL-2 ELISpot (Chapter 5 and 6). Immunodominance was assessed by comparing magnitude and frequency of IFN-γ responses of the novel antigens to immunodominant RD1-encoded ESAT-6 and CFP-10. Since these candidate antigens were hypoxia-induced, I also included a well described dosR-encoded Acr-1 as a reference antigen. It is interesting to note that Acr-1 was also present amongst the 100 highly induced EHR genes.

Many of the EHR encoded MTB proteins were found to be potent T cell targets: 26 antigens induced an IFN-γ response, of which 17 antigens had a response greater than or similar to that of Acr-1 in both magnitude and frequency (Chapter 5 & 6: Figure 36, 37 & 49 and Table 18). This finding provides support to the over all strategy for antigen discovery and gives general support to the hypoxia hypothesis, and its involvement during the course of tuberculosis infection. Previous reports also suggest that, dosR encoded antigens induced by the initial transient hypoxic response of MTB also induce an immune response (Black et al., 2009; Leyten et al., 2006; Schuck et al., 2009).

In addition, these findings give support to the hypothesis that it is possible, to an extent, to predict antigens using genomic data, provided a combination of different

approaches are employed. This approach may not only hold true for *Mycobacterium tuberculosis,* but for other pathogens as well.

Ten EHR antigens (non species specific) were recognized by more than 50% of donors. Of these, 5 novel immunodominant antigens, Rv1957, Rv2022c, Rv1954c, Rv1955 and Rv1471 were recognized by more than 70% of donors, and at a similar magnitude and frequency to the response of ESAT-6 and CFP-10 (Chapter 5: Figure 36, 37, & 44 and Table 18). Rv1955 encodes a toxin molecule. The finding that it is a dominant target of the T cell response suggests a possible role for this gene in the maintenance of MTB under stressful conditions. In addition, these novel antigens, with the exception of Rv1471, are shared only amongst the MTB complex (Chapter 5: Table 14). This provides further support that the observed T cell responses are due to MTB infection. These antigens are potential diagnostic targets and warrant further investigations.

RD2-encoded Rv1986 is a major target of the central memory CD4+T cell response (Chapter 6). The IL-2 response induced by Rv1986 was several times greater than that of ESAT-6 or CFP-10 (Chapter 6 Figure 49 & 51), and was directed narrowly at two epitopic regions (Chapter 6 Figure 44). The significance of the Rv1986 IL-2 response was also studied in the context of the whole bacilli using MTB H37RvΔRD2 (RD2 knock out) and H37RvΔRD2::Rv1986 (RD2 knock out complemented with Rv1986) (Chapter 6 Figure 54). In addition, the IL-2 response to Rv1986 was found to correlate with decrease in susceptibility to TB in HIV-1 infected person on antiretroviral therapy (Chapter 6 Figure 56). This immunodominant target is absent from the commonly used BCG vaccine stains (Behr, 2002; Behr et al., 1999; Gordon et al., 2001; Mahairas et al., 1996). The most important consequence of this finding would be to re-evaluate the immune response of BCG vaccine strains with and

without RD2, which might provide further insight into the variations in vaccine efficacy.

By virtue of its upregulation during extended hypoxia, I investigated whether these antigens were preferentially recognized by individuals with latent tuberculosis. Reconciliation of informatic and empirical evaluation shows that, neither the fold induction nor the normalized intensity of the genes/proteins correlates with the IFN-γ response (Chapter 5 Figure 36-38). With the exception of a modest IFN-γ response to the peptide pool of Rv0849-2, Rv2693-1 (Chapter 5) and Rv2659c-2 (Chapter 6), and very small levels for Rv2021c (Figure 36, 38 & 49), which were significant towards preferential recognition in latent disease, this was not the case.

A link between transient increases in transcript abundance during hypoxia and the immunogenicity of dosR regulated proteins has earlier been attempted and the term 'latency antigen' has been introduced (Leyten et al., 2006). A dominant antigenic target that is dosR regulated is Acr1 encoded by Rv2031c and under some assay conditions, studies have documented preferential T cell recognition of Acr1 by latently infected people (Geluk et al., 2007; Wilkinson et al., 1998b). Preferential recognition of Acr1 by latently infected persons was not observed in this study (Chapter 5 and 6: Figure 34, 35 & 43) nor in a previous IFN-γ ELISpot analysis performed in the same population (Wilkinson et al., 2005a), which is in fact consistent with expression of Acr1 throughout experimental infection (Shi et al., 2003; Stewart et al., 2005).

In addition, another immunodominant dosR antigen Rv1733c, was shown to be preferentially recognized in latent tuberculosis (Leyten et al., 2006). Preferential recognition of this antigen was also not observed in this study (Chapter 4 Figure 31).

Conversely a quantitatively higher IFN-γ response to the RD1 encoded CFP-10 and ESAT-6 antigens has sometimes been associated with active disease (Millington et al., 2007; Rangaka et al., 2007a; Vordermeier et al., 2002), attributed to the secretion of these proteins by actively replicating bacilli. There was however no higher response observed in active tuberculosis compared to latently infected persons (Chapter 5 & 6: Figure 34, 35 and 43).

Differences in infection pressures between low and high incidence areas might feasibly contribute to these differences: the clinical environment in which this study was conducted suffers an extraordinarily high tuberculosis incidence of ~ 1500/100,000 with much ongoing transmission (Rangaka et al., 2007b).

It is also interesting to note that whilst the transcriptomic data showed a fold decrease in ESAT-6 and CFP-10 during hypoxia, the absolute abundance of these transcripts remained high (Chapters 5 & 6). Expression of ESAT-6 and CFP-10 under a variety of conditions is in agreement with other in vitro expression data (Boshoff et al., 2004; Sidders et al., 2008) and adds to data suggesting these molecules may play a role in bacillary persistence as well as active infection (Pym et al., 2002).

Hypoxia does characterize tuberculous granulomas *in vivo* (Wang et al., 2009b) but it is increasingly re-appreciated that both active and latent tuberculous lesions exhibit a dynamic spectrum of overlapping morphologies (Capuano et al., 2003; Kaplan et al., 2003; Lin et al., 2006; Timm et al., 2003; Ulrichs et al., 2005; Young et al., 2009) and that hypoxic lesions likely occur in both clinical circumstances. The findings presented in this thesis strongly suggest similar and/or overlapping pattern of IFN-γ responses between those with active and latent tuberculosis.

In summary,

Many of the EHR encoded MTB genes are potent T cell antigens. This finding gives general support to the hypoxia hypothesis, and its involvement during the course of tuberculosis infection.

This work does not provide strong support to the hypothesis that a differential pattern of immune response is associated with latency. However, these findings provide further support to the recent view of latent tuberculosis as a spectrum, as antigens that are targets of the immune response in latent infection were most often targets in active disease and *vice versa*.

A major target of memory T cells is absent from the commonly used BCG vaccine strains. Further evaluation of BCG vaccine strains with and without RD2 might give insight into the variation in the efficacy.

These data gives support to the hypothesis that it is possible, to an extent, to predict antigens using genomic data, provided combinations (of approaches) are employed.

The novel findings of this work include:

- Novel immunodominant EHR encoded antigens of *M. tuberculosis* (Chapter 5)
    - Rv1957, Rv1955, Rv1954, Rv1471 and Rv2022c, with the magnitude and frequency of response similar to that of RD1-encoded ESAT-6 and CFP-10
    - Rv1284, Rv0990c, Rv2780, Rv0849, Rv0826: ≥ 50% recognition in the population tested
- Rv1986 is a major target of central memory CD4+T cells (Chapter 6)
    - Absent from the commonly used BCG strains

    - IL-2 response directed narrowly to two peptides
    - IL-2 response to Rv1986 better correlate with the decrease in susceptibility to TB during antiretroviral therapy
- T cell epitope mapping of Rv2654c revealed, that the dominant epitope recognized in this population (p51-65) is different from what is previously described (and included in the commercialized assay QFT-GIT) (Chapter 4)
    - Preliminary experiments suggest addition of p51-65 to the QFT-GIT, increases the percentage positivity
- Synthetic overlapping peptides might degrade in prolonged incubation whole blood assay, and there by perform poorly (Chapter 3).

Reference

Aagaard, C., I. Brock, A. Olsen, T.H. Ottenhoff, K. Weldingh, and P. Andersen. 2004. Mapping immune reactivity toward Rv2653 and Rv2654: two novel low-molecular-mass antigens found specifically in the *Mycobacterium tuberculosis* complex. *J Infect Dis* 189:812-819.

Abel, B., M. Tameris, N. Mansoor, S. Gelderbloem, J. Hughes, D. Abrahams, L. Makhethe, M. Erasmus, M. de Kock, L. van der Merwe, A. Hawkridge, A. Veldsman, M. Hatherill, G. Schirru, M.G. Pau, J. Hendriks, G.J. Weverling, J. Goudsmit, D. Sizemore, J.B. McClain, M. Goetz, J. Gearhart, H. Mahomed, G.D. Hussey, J.C. Sadoff, and W.A. Hanekom. 2010. The Novel TB Vaccine, AERAS-402, Induces Robust and Polyfunctional CD4 and CD8 T Cells in Adults. *Am J Respir Crit Care Med*

Adetifa, I.M., M.D. Lugos, A. Hammond, D. Jeffries, S. Donkor, R.A. Adegbola, and P.C. Hill. 2007. Comparison of two interferon gamma release assays in the diagnosis of *Mycobacterium tuberculosis* infection and disease in The Gambia. *BMC Infect Dis* 7:122.

Agger, E.M., I. Brock, L.M. Okkels, S.M. Arend, C.S. Aagaard, K.N. Weldingh, and P. Andersen. 2003. Human T-cell responses to the RD1-encoded protein TB27.4 (Rv3878) from *Mycobacterium tuberculosis*. *Immunology* 110:507-512.

Agrewala, J.N., and R.J. Wilkinson. 1998. Differential regulation of Th1 and Th2 cells by p91-110 and p21-40 peptides of the 16-kD alpha-crystallin antigen of *Mycobacterium tuberculosis*. *Clin Exp Immunol* 114:392-397.

Agrewala, J.N., and R.J. Wilkinson. 1999. Influence of HLA-DR on the phenotype of CD4+ T lymphocytes specific for an epitope of the 16-kDa alpha-crystallin antigen of *Mycobacterium tuberculosis*. *Eur J Immunol* 29:1753-1761.

Aguirre-Blanco, A.M., P.T. Lukey, J.M. Cliff, and H.M. Dockrell. 2007. Strain-dependent variation in *Mycobacterium bovis* BCG-induced human T-cell activation and gamma interferon production in vitro. *Infect Immun* 75:3197-3201.

Al-Attiyah, R., and A.S. Mustafa. 2004. Computer-assisted prediction of HLA-DR binding and experimental analysis for human promiscuous Th1-cell peptides in the 24 kDa secreted lipoprotein (LppX) of *Mycobacterium tuberculosis*. *Scand J Immunol* 59:16-24.

Al-Attiyah, R., and A.S. Mustafa. 2008. Characterization of human cellular immune responses to novel *Mycobacterium tuberculosis* antigens encoded by genomic regions absent in *Mycobacterium bovis* BCG. *Infect Immun* 76:4190-4198.

Altman, J.D., P.A. Moss, P.J. Goulder, D.H. Barouch, M.G. McHeyzer-Williams, J.I. Bell, A.J. McMichael, and M.M. Davis. 1996. Phenotypic analysis of antigen-specific T lymphocytes. *Science* 274:94-96.

Andersen, P., A.B. Andersen, A.L. Sorensen, and S. Nagai. 1995. Recall of long-lived immunity to *Mycobacterium tuberculosis* infection in mice. *J Immunol* 154:3359-3372.

Arend, S.M., P. Andersen, K.E. van Meijgaarden, R.L. Skjot, Y.W. Subronto, J.T. van Dissel, and T.H. Ottenhoff. 2000a. Detection of active tuberculosis infection by T cell responses to early-secreted antigenic target 6-kDa protein and culture filtrate protein 10. *J Infect Dis* 181:1850-1854.

Arend, S.M., A. Geluk, K.E. van Meijgaarden, J.T. van Dissel, M. Theisen, P. Andersen, and T.H. Ottenhoff. 2000b. Antigenic equivalence of human T-cell responses to *Mycobacterium tuberculosis*-specific RD1-encoded protein antigens ESAT-6 and culture filtrate protein 10 and to mixtures of synthetic peptides. *Infect Immun* 68:3314-3321.

Bakir, M., D.P. Dosanjh, J.J. Deeks, A. Soysal, K.A. Millington, S. Efe, Y. Aslan, D. Polat, N. Kodalli, A. Yagci, I. Barlan, N. Bahceciler, E.E. Demiralp, and A. Lalvani. 2009. Use of T cell-based diagnosis of tuberculosis infection to optimize interpretation of tuberculin skin testing for child tuberculosis contacts. *Clin Infect Dis* 48:302-312.

Balasingham, S.V., T. Davidsen, I. Szpinda, S.A. Frye, and T. Tonjum. 2009. Molecular diagnostics in tuberculosis: basis and implications for therapy. *Mol Diagn Ther* 13:137-151.

Barry, C.E., 3rd, H.I. Boshoff, V. Dartois, T. Dick, S. Ehrt, J. Flynn, D. Schnappinger, R.J. Wilkinson, and D. Young. 2009. The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Microbiol* 7:845-855.

Behr, M.A. 2002. BCG--different strains, different vaccines? *Lancet Infect Dis* 2:86-92.

Behr, M.A., and P.M. Small. 1997. Has BCG attenuated to impotence? *Nature* 389:133-134.

Behr, M.A., M.A. Wilson, W.P. Gill, H. Salamon, G.K. Schoolnik, S. Rane, and P.M. Small. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 284:1520-1523.

Berthet, F.X., P.B. Rasmussen, I. Rosenkrands, P. Andersen, and B. Gicquel. 1998. A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). *Microbiology* 144 (Pt 11):3195-3203.

Bertholet, S., G.C. Ireton, M. Kahn, J. Guderian, R. Mohamath, N. Stride, E.M. Laughlin, S.L. Baldwin, T.S. Vedvick, R.N. Coler, and S.G. Reed. 2008. Identification of human T cell antigens for the development of vaccines against *Mycobacterium tuberculosis*. *J Immunol* 181:7948-7957.

Betts, J.C., P.T. Lukey, L.C. Robb, R.A. McAdam, and K. Duncan. 2002. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol Microbiol* 43:717-731.

Beveridge, N.E., H.A. Fletcher, J. Hughes, A.A. Pathan, T.J. Scriba, A. Minassian, C.R. Sander, K.T. Whelan, H.M. Dockrell, A.V. Hill, W.A. Hanekom, and H. McShane. 2008. A comparison of IFN gamma detection methods used in tuberculosis vaccine trials. *Tuberculosis (Edinb)* 88:631-640.

Beveridge, N.E., D.A. Price, J.P. Casazza, A.A. Pathan, C.R. Sander, T.E. Asher, D.R. Ambrozak, M.L. Precopio, P. Scheinberg, N.C. Alder, M. Roederer, R.A. Koup, D.C. Douek, A.V. Hill, and H. McShane. 2007. Immunisation with BCG and recombinant MVA85A induces long-lasting, polyfunctional *Mycobacterium tuberculosis*-specific CD4+ memory T lymphocyte populations. *Eur J Immunol* 37:3089-3100.

Bibb, L.A., and G.F. Hatfull. 2002. Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, phiRv1. *Mol Microbiol* 45:1515-1526.

Bivas-Benita, M., M.Y. Lin, S.M. Bal, K.E. van Meijgaarden, K.L. Franken, A.H. Friggen, H.E. Junginger, G. Borchard, M.R. Klein, and T.H. Ottenhoff. 2009. Pulmonary delivery of DNA encoding *Mycobacterium tuberculosis* latency antigen Rv1733c associated to PLGA-PEI nanoparticles enhances T cell responses in a DNA prime/protein boost vaccination regimen in mice. *Vaccine* 27:4010-4017.

Black, G.F., B.A. Thiel, M.O. Ota, S.K. Parida, R. Adegbola, W.H. Boom, H.M. Dockrell, K.L. Franken, A.H. Friggen, P.C. Hill, M.R. Klein, M.K. Lalor, H. Mayanja, G. Schoolnik, K. Stanley, K. Weldingh, S.H. Kaufmann, G. Walzl, and T.H. Ottenhoff. 2009. Immunogenicity of novel DosR regulon-encoded candidate antigens of *Mycobacterium tuberculosis* in three high-burden populations in Africa. *Clin Vaccine Immunol* 16:1203-1212.

Black, G.F., R.E. Weir, S. Floyd, L. Bliss, D.K. Warndorff, A.C. Crampin, B. Ngwira, L. Sichali, B. Nazareth, J.M. Blackwell, K. Branson, S.D. Chaguluka, L. Donovan, E. Jarman, E. King, P.E. Fine, and H.M. Dockrell. 2002. 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* 359:1393-1401.

Bloom, B.R., and P.E.M. Fine. 1994. The BCG experience: Implications for future vaccines against tuberculosis. In Tuberculosis: Pathogenesis, protection and control. B.R. Bloom, editor American Society for Microbiology, Washington DC. 531-557.

Boesen, H., B.N. Jensen, T. Wilcke, and P. Andersen. 1995. Human T-cell responses to secreted antigen fractions of *Mycobacterium tuberculosis*. *Infect Immun* 63:1491-1497.

Borsuk, S., J. Newcombe, T.A. Mendum, O.A. Dellagostin, and J. McFadden. 2009. Identification of proteins from tuberculin purified protein derivative (PPD) by LC-MS/MS. *Tuberculosis (Edinb)* 89:423-430.

Boshoff, H.I., and C.E. Barry, 3rd. 2005. Tuberculosis - metabolism and respiration in the absence of growth. *Nat Rev Microbiol* 3:70-80.

Boshoff, H.I., T.G. Myers, B.R. Copp, M.R. McNeil, M.A. Wilson, and C.E. Barry, 3rd. 2004. The transcriptional responses of *Mycobacterium tuberculosis* to inhibitors of metabolism: novel insights into drug mechanisms of action. *J Biol Chem* 279:40174-40184.

Botha, T., and B. Ryffel. 2003. Reactivation of latent tuberculosis infection in TNF-deficient mice. *J Immunol* 171:3110-3118.

Bothamley, G.H., R. Rudd, F. Festenstein, and J. Ivanyi. 1992. Clinical value of the measurement of *Mycobacterium tuberculosis* specific antibody in pulmonary tuberculosis. *Thorax* 47:270-275.

Brewer, T.F. 2000. Preventing tuberculosis with bacillus Calmette-Guerin vaccine: a meta-analysis of the literature. *Clin Infect Dis* 31 Suppl 3:S64-67.

Brock, I., K. Weldingh, E.M. Leyten, S.M. Arend, P. Ravn, and P. Andersen. 2004. Specific T-cell epitopes for immunoassay-based diagnosis of *Mycobacterium tuberculosis* infection. *J Clin Microbiol* 42:2379-2387.

Brosch, R., S.V. Gordon, T. Garnier, K. Eiglmeier, W. Frigui, P. Valenti, S. Dos Santos, S. Duthoy, C. Lacroix, C. Garcia-Pelayo, J.K. Inwald, P. Golby, J.N. Garcia, R.G. Hewinson, M.A. Behr, M.A. Quail, C. Churcher, B.G. Barrell, J. Parkhill, and S.T. Cole. 2007. Genome plasticity of BCG and impact on vaccine efficacy. *Proc Natl Acad Sci U S A* 104:5596-5601.

Brosch, R., S.V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L.M. Parsons, A.S. Pym, S. Samper, D. van Soolingen, and S.T. Cole. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* 99:3684-3689.

Buus, S. 1999. Description and prediction of peptide-MHC binding: the 'human MHC project'. *Curr Opin Immunol* 11:209-213.

Capuano, S.V., 3rd, D.A. Croix, S. Pawar, A. Zinovik, A. Myers, P.L. Lin, S. Bissel, C. Fuhrman, E. Klein, and J.L. Flynn. 2003. Experimental *Mycobacterium tuberculosis* infection of cynomolgus macaques closely resembles the various manifestations of human *M. tuberculosis* infection. *Infect Immun* 71:5831-5844.

Cardona, P.J. 2007. New insights on the nature of latent tuberculosis infection and its treatment. *Inflamm Allergy Drug Targets* 6:27-39.

Chen, J., X. Su, Y. Zhang, S. Wang, L. Shao, J. Wu, F. Wang, S. Zhang, J. Wang, X. Weng, H. Wang, and W. Zhang. 2009. Novel recombinant RD2- and RD11-encoded *Mycobacterium tuberculosis* antigens are potential candidates for

diagnosis of tuberculosis infections in BCG-vaccinated individuals. *Microbes Infect* 11:876-885.

Cockle, P.J., S.V. Gordon, A. Lalvani, B.M. Buddle, R.G. Hewinson, and H.M. Vordermeier. 2002. Identification of novel *Mycobacterium tuberculosis* antigens with potential as diagnostic reagents or subunit vaccine candidates by comparative genomics. *Infect Immun* 70:6996-7003.

Colditz, G.A., T.F. Brewer, C.S. Berkey, M.E. Wilson, E. Burdick, H.V. Fineberg, and F. Mosteller. 1994. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA* 271:698-702.

Cole, S., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. Gordon, K. Eiglmeier, S. Gas, C. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. Sulston, K. Taylor, S. Whitehead, and B. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393:537–544.

Cole, S.T. 2002. Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. *Microbiology* 148:2919-2928.

Connell, T.G., M.S. Shey, R. Seldon, M.X. Rangaka, G. van Cutsem, M. Simsova, Z. Marcekova, P. Sebo, N. Curtis, L. Diwakar, G.A. Meintjes, C. Leclerc, R.J. Wilkinson, and K.A. Wilkinson. 2007. Enhanced ex vivo stimulation of *Mycobacterium tuberculosis*-specific T cells in human immunodeficiency virus-infected persons via antigen delivery by the Bordetella pertussis adenylate cyclase vector. *Clin Vaccine Immunol* 14:847-854.

Daniel, T.M. 2006. The history of tuberculosis. *Respir Med* 100:1862-1870.

Daniel, T.M., and S.M. Debanne. 1987. The serodiagnosis of tuberculosis and other mycobacterial diseases by enzyme-linked immunosorbent assay. *Am Rev Respir Dis* 135:1137-1151.

Daniel, T.M., J.J. Ellner, L.S. Todd, D.W. McCoy, V.D. Payne, P.A. Anderson, and F.T. Bhe. 1979. Immunobiology and species distribution of *Mycobacterium tuberculosis* antigen 5. *Infect Immun* 24:77-82.

Davids, V., W.A. Hanekom, N. Mansoor, H. Gamieldien, S.J. Gelderbloem, A. Hawkridge, G.D. Hussey, E.J. Hughes, J. Soler, R.A. Murray, S.R. Ress, and G. Kaplan. 2006. The effect of bacille calmette-guerin vaccine strain and route of administration on induced immune responses in vaccinated infants. *J Infect Dis* 193:531-536.

Day, C.L., N. Mkhwanazi, S. Reddy, Z. Mncube, M. van der Stok, P. Klenerman, and B.D. Walker. 2008. Detection of polyfunctional *Mycobacterium tuberculosis*-specific T cells and association with viral load in HIV-1-infected persons. *J Infect Dis* 197:990-999.

de Jonge, M.I., R. Brosch, P. Brodin, C. Demangel, and S.T. Cole. 2005. Tuberculosis: from genome to vaccine. *Expert Rev Vaccines* 4:541-551.

Delogu, G., and G. Fadda. 2009. The quest for a new vaccine against tuberculosis. *J Infect Dev Ctries* 3:5-15.

Demissie, A., E.M. Leyten, M. Abebe, L. Wassie, A. Aseffa, G. Abate, H. Fletcher, P. Owiafe, P.C. Hill, R. Brookes, G. Rook, A. Zumla, S.M. Arend, M. Klein, T.H. Ottenhoff, P. Andersen, and T.M. Doherty. 2006. Recognition of stage-specific mycobacterial antigens differentiates between acute and latent infections with *Mycobacterium tuberculosis*. *Clin Vaccine Immunol* 13:179-186.

Dimitrov, I., P. Garnev, D.R. Flower, and I. Doytchinova. 2010. MHC Class II Binding Prediction-A Little Help from a Friend. *J Biomed Biotechnol* 2010:705821.

Disis, M.L., C. dela Rosa, V. Goodell, L.Y. Kuan, J.C. Chang, K. Kuus-Reichel, T.M. Clay, H. Kim Lyerly, S. Bhatia, S.A. Ghanekar, V.C. Maino, and H.T. Maecker. 2006. Maximizing the retention of antigen specific lymphocyte function after cryopreservation. *J Immunol Methods* 308:13-18.

du Toit, E.D., K.J. MacGregor, D.G. Taljaard, and M. Oudshoorn. 1988. HLA-A, B, C, DR and DQ polymorphisms in three South African population groups: South African Negroes, Cape Coloureds and South African Caucasoids. *Tissue Antigens* 31:109-125.

Ehlers, S. 2009. Lazy, dynamic or minimally recrudescent? On the elusive nature and location of the mycobacterium responsible for latent tuberculosis. *Infection* 37:87-95.

Ewer, K., P. Cockle, S. Gordon, H. Mansoor, M. Govaerts, K. Walravens, S. Marche, G. Hewinson, and M. Vordermeier. 2006. Antigen mining with iterative genome screens identifies novel diagnostics for the *Mycobacterium tuberculosis* complex. *Clin Vaccine Immunol* 13:90-97.

Flynn, J.L., J. Chan, K.J. Triebold, D.K. Dalton, T.A. Stewart, and B.R. Bloom. 1993. An essential role for interferon gamma in resistance to *Mycobacterium tuberculosis* infection. *J Exp Med* 178:2249-2254.

Flynn, J.L., M.M. Goldstein, J. Chan, K.J. Triebold, K. Pfeffer, C.J. Lowenstein, R. Schreiber, T.W. Mak, and B.R. Bloom. 1995. Tumor necrosis factor-alpha is required in the protective immune response against *Mycobacterium tuberculosis* in mice. *Immunity* 2:561-572.

Fu, R., C. Wang, C. Shi, M. Lu, Z. Fang, J. Lu, F. Wang, and X. Fan. 2009. An improved whole-blood gamma interferon assay based on the CFP21-MPT64 fusion protein. *Clin Vaccine Immunol* 16:686-691.

Gaseitsiwe, S., D. Valentini, S. Mahdavifar, I. Magalhaes, D.F. Hoft, J. Zerweck, M. Schutkowski, J. Andersson, M. Reilly, and M.J. Maeurer. 2008. Pattern recognition in pulmonary tuberculosis defined by high content peptide microarray chip analysis representing 61 proteins from M. tuberculosis. *PLoS One* 3:e3840.

Geluk, A., M.Y. Lin, K.E. van Meijgaarden, E.M. Leyten, K.L. Franken, T.H. Ottenhoff, and M.R. Klein. 2007. T cell recognition of the HspX protein of Mycobacterium tuberculosis correlates with latent *M. tuberculosis* infection but not BCG vaccination. *Infect Immun* 75:2914-2921.

Gerdes, K., S.K. Christensen, and A. Lobner-Olesen. 2005. Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol* 3:371-382.

Goletti, D., O. Butera, V. Vanini, F.N. Lauria, C. Lange, K.L. Franken, C. Angeletti, T.H. Ottenhoff, and E. Girardi. 2009. Response to Rv2628 latency antigen associates with cured tuberculosis and remote infection. *Eur Respir J*

Gomez, J.E., and J.D. McKinney. 2004. *M. tuberculosis* persistence, latency, and drug tolerance. *Tuberculosis (Edinb)* 84:29-44.

Gordon, S.V., R. Brosch, A. Billault, T. Garnier, K. Eiglmeier, and S.T. Cole. 1999a. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Mol Microbiol* 32:643-655.

Gordon, S.V., K. Eiglmeier, T. Garnier, R. Brosch, J. Parkhill, B. Barrell, S.T. Cole, and R.G. Hewinson. 2001. Genomics of *Mycobacterium bovis*. *Tuberculosis (Edinb)* 81:157-163.

Gordon, S.V., B. Heym, J. Parkhill, B. Barrell, and S.T. Cole. 1999b. New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 145 (Pt 4):881-892.

Gowthaman, U., and J.N. Agrewala. 2008. In silico tools for predicting peptides binding to HLA-class II molecules: more confusion than conclusion. *J Proteome Res* 7:154-163.

Gowthaman, U., and J.N. Agrewala. 2009. In silico methods for predicting T-cell epitopes: Dr Jekyll or Mr Hyde? *Expert Rev Proteomics* 6:527-537.

Grange, J.M. 1984. The humoral immune response in tuberculosis: its nature, biological role and diagnostic usefulness. *Adv Tuberc Res* 21:1-78.

Grover, A., M.F. Ahmed, I. Verma, P. Sharma, and G.K. Khuller. 2006. Expression and purification of the *Mycobacterium tuberculosis* complex-restricted antigen CFP21 to study its immunoprophylactic potential in mouse model. *Protein Expr Purif* 48:274-280.

Gupta, A. 2009. Killing activity and rescue function of genome-wide toxin-antitoxin loci of *Mycobacterium tuberculosis*. *FEMS Microbiol Lett* 290:45-53.

Hanekom, W.A., B. Abel, and T.J. Scriba. 2007. Immunological protection against tuberculosis. *S Afr Med J* 97:973-977.

Hanekom, W.A., H.M. Dockrell, T.H. Ottenhoff, T.M. Doherty, H. Fletcher, H. McShane, F.F. Weichold, D.F. Hoft, S.K. Parida, and U.J. Fruth. 2008. Immunological Outcomes of New Tuberculosis Vaccine Trials: WHO Panel Recommendations. *PLoS Med* 5:e145.

Hanekom, W.A., J. Hughes, M. Mavinkurve, M. Mendillo, M. Watkins, H. Gamieldien, S.J. Gelderbloem, M. Sidibana, N. Mansoor, V. Davids, R.A. Murray, A. Hawkridge, P.A. Haslett, S. Ress, G.D. Hussey, and G. Kaplan. 2004. Novel application of a whole blood intracellular cytokine detection

assay to quantitate specific T-cell frequency in field studies. *J Immunol Methods* 291:185-195.

Hawkridge, T., T.J. Scriba, S. Gelderbloem, E. Smit, M. Tameris, S. Moyo, T. Lang, A. Veldsman, M. Hatherill, L.V. Merwe, H.A. Fletcher, H. Mahomed, A.V. Hill, W.A. Hanekom, G.D. Hussey, and H. McShane. 2008. Safety and Immunogenicity of a New Tuberculosis Vaccine, MVA85A, in Healthy Adults in South Africa. *J Infect Dis* 198:544-552.

Hendrix, R.W., M.C. Smith, R.N. Burns, M.E. Ford, and G.F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A* 96:2192-2197.

Herzog, H. 1998. History of tuberculosis. *Respiration* 65:5-15.

Hesseling, A.C., B.J. Marais, R.P. Gie, H.S. Schaaf, P.E. Fine, P. Godfrey-Faussett, and N. Beyers. 2007. The risk of disseminated Bacille Calmette-Guerin (BCG) disease in HIV-infected children. *Vaccine* 25:14-18.

Hesseling, A.C., H.S. Schaaf, W.A. Hanekom, N. Beyers, M.F. Cotton, R.P. Gie, B.J. Marais, P. van Helden, and R.M. Warren. 2003. Danish bacille Calmette-Guerin vaccine-induced disease in human immunodeficiency virus-infected children. *Clin Infect Dis* 37:1226-1233.

Hewinson, R.G., H.M. Vordermeier, N.H. Smith, and S.V. Gordon. 2006. Recent advances in our knowledge of *Mycobacterium bovis*: a feeling for the organism. *Vet Microbiol* 112:127-139.

Hewitt, J., A.R. Coates, D.A. Mitchison, and J. Ivanyi. 1982. The use of murine monoclonal antibodies without purification of antigen in the serodiagnosis of tuberculosis. *J Immunol Methods* 55:205-211.

Horwitz, M.A., G. Harth, B.J. Dillon, and S. Maslesa-Galic. 2009. Commonly administered BCG strains including an evolutionarily early strain and

evolutionarily late strains of disparate genealogy induce comparable protective immunity against tuberculosis. *Vaccine* 27:441-445.

Hsieh, C.S., S.E. Macatonia, C.S. Tripp, S.F. Wolf, A. O'Garra, and K.M. Murphy. 1993. Development of TH1 CD4+ T cells through IL-12 produced by Listeria-induced macrophages [see comments]. *Science* 260:547-549.

Huebner, R.E., M.F. Schein, and J.B. Bass, Jr. 1993. The tuberculin skin test. *Clin Infect Dis* 17:968-975.

Hussey, G.D., M.L. Watkins, E.A. Goddard, S. Gottschalk, E.J. Hughes, K. Iloni, M.A. Kibel, and S.R. Ress. 2002. Neonatal mycobacterial specific cytotoxic T-lymphocyte and cytokine profiles in response to distinct BCG vaccination strategies. *Immunology* 105:314-324.

Jackett, P.S., G.H. Bothamley, H.V. Batra, A. Mistry, D.B. Young, and J. Ivanyi. 1988. Specificity of antibodies to immunodominant mycobacterial antigens in pulmonary tuberculosis. *J Clin Microbiol* 26:2313-2318.

Jurcevic, S., A. Hills, G. Pasvol, R.N. Davidson, J. Ivanyi, and R.J. Wilkinson. 1996. T cell responses to a mixture of *Mycobacterium tuberculosis* peptides with complementary HLA-DR binding profiles. *Clin Exp Immunol* 105:416-421.

Kalra, M., G.K. Khuller, A. Grover, D. Behera, A. Wanchu, and I. Verma. 2009. Utility of a combination of RD1 and RD2 antigens as a diagnostic marker for tuberculosis. *Diagn Microbiol Infect Dis*

Kaplan, G., F.A. Post, A.L. Moreira, H. Wainwright, B.N. Kreiswirth, M. Tanverdi, B. Mathema, S.V. Ramaswamy, G. Walther, L.M. Steyn, C.E. Barry, 3rd, and L.G. Bekker. 2003. *Mycobacterium tuberculosis* growth at the cavity surface: a microenvironment with failed immunity. *Infect Immun* 71:7099-7108.

Kaufmann, S.H. 2001. How can immunology contribute to the control of tuberculosis? *Nat Rev Immunol* 1:20-30.

Keane, J., S. Gershon, R.P. Wise, E. Mirabile-Levens, J. Kasznica, W.D. Schwieterman, J.N. Siegel, and M.M. Braun. 2001. Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. *N Engl J Med* 345:1098-1104.

Knezevic, I., and M.J. Corbel. 2006. WHO discussion on the improvement of the quality control of BCG vaccines. *Vaccine* 24:3874-3877.

Kreher, C.R., M.T. Dittrich, R. Guerkov, B.O. Boehm, and M. Tary-Lehmann. 2003. CD4+ and CD8+ cells in cryopreserved human PBMC maintain full functionality in cytokine ELISPOT assays. *J Immunol Methods* 278:79-93.

Lafuente, E.M., and P.A. Reche. 2009. Prediction of MHC-peptide binding: a systematic and comprehensive overview. *Curr Pharm Des* 15:3209-3220.

Lagrange, P.H., and J.L. Herrmann. 2008. Diagnosing Latent Tuberculosis Infection in the HIV Era. *Open Respir Med J* 2:52-59.

Lalvani, A. 2007. Diagnosing tuberculosis infection in the 21st century: new tools to tackle an old enemy. *Chest* 131:1898-1906.

Lalvani, A., R. Brookes, R.J. Wilkinson, A.S. Malin, A.A. Pathan, P. Andersen, H. Dockrell, G. Pasvol, and A.V. Hill. 1998. Human cytolytic and interferon gamma-secreting CD8+ T lymphocytes specific for *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 95:270-275.

Lalvani, A., P. Nagvenkar, Z. Udwadia, A.A. Pathan, K.A. Wilkinson, J.S. Shastri, K. Ewer, A.V. Hill, A. Mehta, and C. Rodrigues. 2001a. Enumeration of T cells specific for RD1-encoded antigens suggests a high prevalence of latent *Mycobacterium tuberculosis* infection in healthy urban Indians. *J Infect Dis* 183:469-477.

Lalvani, A., and M. Pareek. 2009a. A 100 year update on diagnosis of tuberculosis infection. *Br Med Bull* 93:69-84.

Lalvani, A., and M. Pareek. 2009b. Interferon gamma release assays: principles and practice. *Enferm Infecc Microbiol Clin*

Lalvani, A., A.A. Pathan, H. McShane, R.J. Wilkinson, M. Latif, C.P. Conlon, G. Pasvol, and A.V. Hill. 2001b. Rapid detection of *Mycobacterium tuberculosis* infection by enumeration of antigen-specific T cells. *Am J Respir Crit Care Med* 163:824-828.

Lange, C., and T. Mori. 2010. Advances in the diagnosis of tuberculosis. *Respirology* 15:220-240.

Lange, C., M. Pai, F. Drobniewski, and G.B. Migliori. 2009. Interferon-gamma release assays for the diagnosis of active tuberculosis: sensible or silly? *Eur Respir J* 33:1250-1253.

Lazarevic, V., D. Nolt, and J.L. Flynn. 2005. Long-term control of *Mycobacterium tuberculosis* infection is mediated by dynamic immune responses. *J Immunol* 175:1107-1117.

Lee, J.Y., H.J. Choi, I.N. Park, S.B. Hong, Y.M. Oh, C.M. Lim, S.D. Lee, Y. Koh, W.S. Kim, D.S. Kim, W.D. Kim, and T.S. Shim. 2006. Comparison of two commercial interferon-gamma assays for diagnosing *Mycobacterium tuberculosis* infection. *Eur Respir J* 28:24-30.

Lein, A.D., C.F. von Reyn, P. Ravn, C.R. Horsburgh, Jr., L.N. Alexander, and P. Andersen. 1999. Cellular immune responses to ESAT-6 discriminate between patients with pulmonary disease due to *Mycobacterium avium* complex and those with pulmonary disease due to *Mycobacterium tuberculosis*. *Clin Diagn Lab Immunol* 6:606-609.

Lewis, K.N., R. Liao, K.M. Guinn, M.J. Hickey, S. Smith, M.A. Behr, and D.R. Sherman. 2003. Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guerin attenuation. *J Infect Dis* 187:117-123.

Leyten, E.M., S.M. Arend, C. Prins, F.G. Cobelens, T.H. Ottenhoff, and J.T. van Dissel. 2007. Discrepancy between *Mycobacterium tuberculosis*-specific gamma interferon release assays using short and prolonged in vitro incubation. *Clin Vaccine Immunol* 14:880-885.

Leyten, E.M., M.Y. Lin, K.L. Franken, A.H. Friggen, C. Prins, K.E. van Meijgaarden, M.I. Voskuil, K. Weldingh, P. Andersen, G.K. Schoolnik, S.M. Arend, T.H. Ottenhoff, and M.R. Klein. 2006. Human T-cell responses to 25 novel antigens encoded by genes of the dormancy regulon of *Mycobacterium tuberculosis*. *Microbes Infect* 8:2052-2060.

Li, H., J.C. Ulstrup, T.O. Jonassen, K. Melby, S. Nagai, and M. Harboe. 1993. Evidence for absence of the MPB64 gene in some substrains of *Mycobacterium bovis* BCG. *Infect Immun* 61:1730-1734.

Lima-Junior, J.C., D.M. Banic, T.M. Tran, V.S. Meyer, S.G. De-Simone, F. Santos, L.C. Porto, M.T. Marques, A. Moreno, J.W. Barnwell, M.R. Galinski, and J. Oliveira-Ferreira. 2010. Promiscuous T-cell epitopes of Plasmodium merozoite surface protein 9 (PvMSP9) induces IFN-gamma and IL-4 responses in individuals naturally exposed to malaria in the Brazilian Amazon. *Vaccine* 28:3185-3191.

Lin, H.H., G.L. Zhang, S. Tongchusak, E.L. Reinherz, and V. Brusic. 2008. Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. *BMC Bioinformatics* 9 Suppl 12:S22.

Lin, M.Y., A. Geluk, S.G. Smith, A.L. Stewart, A.H. Friggen, K.L. Franken, M.J. Verduyn, K.E. van Meijgaarden, M.I. Voskuil, H.M. Dockrell, K. Huygen, T.H. Ottenhoff, and M.R. Klein. 2007. Lack of immune responses to Mycobacterium tuberculosis DosR regulon proteins following *Mycobacterium bovis* BCG vaccination. *Infect Immun* 75:3523-3530.

Lin, M.Y., and T.H. Ottenhoff. 2008a. Host-pathogen interactions in latent *Mycobacterium tuberculosis* infection: identification of new targets for

tuberculosis intervention. *Endocr Metab Immune Disord Drug Targets* 8:15-29.

Lin, M.Y., and T.H. Ottenhoff. 2008b. Not to wake a sleeping giant: new insights into host-pathogen interactions identify new targets for vaccination against latent *Mycobacterium tuberculosis* infection. *Biol Chem* 389:497-511.

Lin, M.Y., T.B. Reddy, S.M. Arend, A.H. Friggen, K.L. Franken, K.E. van Meijgaarden, M.J. Verduyn, G.K. Schoolnik, M.R. Klein, and T.H. Ottenhoff. 2009a. Cross-reactive immunity to *Mycobacterium tuberculosis* DosR regulon-encoded antigens in individuals infected with environmental, nontuberculous mycobacteria. *Infect Immun* 77:5071-5079.

Lin, P.L., S. Pawar, A. Myers, A. Pegu, C. Fuhrman, T.A. Reinhart, S.V. Capuano, E. Klein, and J.L. Flynn. 2006. Early events in *Mycobacterium tuberculosis* infection in cynomolgus macaques. *Infect Immun* 74:3790-3803.

Lin, P.L., M. Rodgers, L. Smith, M. Bigbee, A. Myers, C. Bigbee, I. Chiosea, S.V. Capuano, C. Fuhrman, E. Klein, and J.L. Flynn. 2009b. Quantitative comparison of active and latent tuberculosis in the cynomolgus macaque model. *Infect Immun* 77:4631-4642.

Maartens, G., and R.J. Wilkinson. 2007. Tuberculosis. *Lancet* 370:2030-2043.

MacMicking, J.D., R.J. North, R. LaCourse, J.S. Mudgett, S.K. Shah, and C.F. Nathan. 1997. Identification of nitric oxide synthase as a protective locus against tuberculosis. *Proc. Natl. Acad. Sci. USA* 94:5243-5248.

Maecker, H.T., J. Moon, S. Bhatia, S.A. Ghanekar, V.C. Maino, J.K. Payne, K. Kuus-Reichel, J.C. Chang, A. Summers, T.M. Clay, M.A. Morse, H.K. Lyerly, C. DeLaRosa, D.P. Ankerst, and M.L. Disis. 2005. Impact of cryopreservation on tetramer, cytokine flow cytometry, and ELISPOT. *BMC Immunol* 6:17.

Mahairas, G.G., P.J. Sabo, M.J. Hickey, D.C. Singh, and C.K. Stover. 1996. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent M. bovis. *J Bacteriol* 178:1274-1282.

Malen, H., T. Softeland, and H.G. Wiker. 2008. Antigen analysis of Mycobacterium tuberculosis H37Rv culture filtrate proteins. *Scand J Immunol* 67:245-252.

McKnight, A.J., V.L. Perez, C.M. Shea, G.S. Gray, and A.K. Abbas. 1994. Costimulator dependence of lymphokine secretion by naive and activated CD4+ T lymphocytes from TCR transgenic mice. *J Immunol* 152:5220-5225.

Menzies, D., M. Pai, and G. Comstock. 2007. Meta-analysis: new tests for the diagnosis of latent tuberculosis infection: areas of uncertainty and recommendations for research. *Ann Intern Med* 146:340-354.

Millington, K.A., J.A. Innes, S. Hackforth, T.S. Hinks, J.J. Deeks, D.P. Dosanjh, V. Guyot-Revol, R. Gunatheesan, P. Klenerman, and A. Lalvani. 2007. Dynamic relationship between IFN-gamma and IL-2 profile of *Mycobacterium tuberculosis*-specific T cells and antigen load. *J Immunol* 178:5217-5226.

Mitchison, D.A. 1972. Bacteriological aspects of mycobacterial infections. *Br Med J* 1:424-425.

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

Mostowy, S., and M.A. Behr. 2002. Comparative genomics in the fight against tuberculosis: diagnostics, epidemiology, and BCG vaccination. *Am J Pharmacogenomics* 2:189-196.

Murphy, D.J., and J.R. Brown. 2008. Novel drug target strategies against *Mycobacterium tuberculosis*. *Curr Opin Microbiol* 11:422-427.

Murray, J.S., C. Pfeiffer, J. Madri, and K. Bottomly. 1992. Major histocompatibility complex (MHC) control of CD4 T cell subset activation. II. A single peptide induces either humoral or cell-mediated responses in mice of distinct MHC genotype. *Eur J Immunol* 22:559-565.

Mustafa, A.S. 2009a. HLA-promiscuous Th1-cell reactivity of MPT64 (Rv1980c), a major secreted antigen of *Mycobacterium tuberculosis*, in healthy subjects. *Med Princ Pract* 18:385-392.

Mustafa, A.S. 2009b. Th1 cell reactivity and HLA-DR binding prediction for promiscuous recognition of MPT63 (Rv1926c), a major secreted protein of *Mycobacterium tuberculosis*. *Scand J Immunol* 69:213-222.

Mustafa, A.S., F. Oftung, H.A. Amoudy, N.M. Madi, A.T. Abal, F. Shaban, I. Rosen Krands, and P. Andersen. 2000. Multiple epitopes from the *Mycobacterium tuberculosis* ESAT-6 antigen are recognized by antigen-specific human T cell lines. *Clin Infect Dis* 30 Suppl 3:S201-205.

Mustafa, A.S., and F.A. Shaban. 2006. ProPred analysis and experimental evaluation of promiscuous T-cell epitopes of three major secreted antigens of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* 86:115-124.

Narayanan, P.R. 2006. Influence of sex, age & nontuberculous infection at intake on the efficacy of BCG: re-analysis of 15-year data from a double-blind randomized control trial in South India. *Indian J Med Res* 123:119-124.

Newport, M.J., C.M. Huxley, S. Huston, C.M. Hawrylowicz, B.A. Oostra, R. Williamson, and M. Levin. 1996. A mutation in the interferon-gamma-receptor gene and susceptibility to mycobacterial infection. *N Engl J Med* 335:1941-1949.

Nicholson, S., M.d.G. Bonecini-Almeida, J.R. Lapa e Silva, C. Nathan, Q.-W. Xie, R. Mumford, J.R. Weidner, J. Calaycay, J. Geng, N. Boechat, C. Linhares, W. Rom, and J.L. Ho. 1996. Inducible Nitric Oxide Synthase in pulmonary

alveolar macrophages from patients with tuberculosis. *J Exp Med* 183:2293-2302.

Nielsen, M., O. Lund, S. Buus, and C. Lundegaard. 2010. MHC Class II epitope predictive algorithms. *Immunology*

Nusslein, K., L. Arnt, J. Rennie, C. Owens, and G.N. Tew. 2006. Broad-spectrum antibacterial activity by a novel abiogenic peptide mimic. *Microbiology* 152:1913-1918.

O'Brien, C., D.R. Flower, and C. Feighery. 2008. Peptide length significantly influences in vitro affinity for MHC class II molecules. *Immunome Res* 4:6.

Oni, T., J. Patel, H.P. Gideon, R. Seldon, K. Wood, Y. Hlombe, K.A. Wilkinson, M.X. Rangaka, M. Mendelson, and R.J. Wilkinson. 2010. Enhanced diagnosis of HIV-1 associated tuberculosis by relating T-SPOT.TB and CD4 counts. *Eur Respir J*

Ovsyannikova, I.G., R.M. Jacobson, N. Dhiman, R.A. Vierkant, V.S. Pankratz, and G.A. Poland. 2008. Human leukocyte antigen and cytokine receptor gene polymorphisms associated with heterogeneous immune responses to mumps viral vaccine. *Pediatrics* 121:e1091-1099.

Pai, M., J. Minion, H. Sohn, A. Zwerling, and M.D. Perkins. 2009. Novel and improved technologies for tuberculosis diagnosis: progress and challenges. *Clin Chest Med* 30:701-716, viii.

Pai, M., A. Zwerling, and D. Menzies. 2008. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. *Ann Intern Med* 149:177-184.

Palomino, J.C. 2005. Nonconventional and new methods in the diagnosis of tuberculosis: feasibility and applicability in the field. *Eur Respir J* 26:339-350.

Pandey, D.P., and K. Gerdes. 2005. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res* 33:966-976.

Parkash, O., B.P. Singh, and M. Pai. 2009. Regions of differences encoded antigens as targets for immunodiagnosis of tuberculosis in humans. *Scand J Immunol* 70:345-357.

Pathan, A.A., K.A. Wilkinson, P. Klenerman, H. McShane, R.N. Davidson, G. Pasvol, A.V. Hill, and A. Lalvani. 2001. Direct *Ex Vivo* Analysis of Antigen-Specific IFN-gamma-Secreting CD4 T Cells in *Mycobacterium tuberculosis*-Infected Individuals: Associations with Clinical Disease State and Effect of Treatment. *J Immunol* 167:5217-5225.

Pym, A.S., P. Brodin, R. Brosch, M. Huerre, and S.T. Cole. 2002. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Mol Microbiol* 46:709-717.

Pym, A.S., P. Brodin, L. Majlessi, R. Brosch, C. Demangel, A. Williams, K.E. Griffiths, G. Marchal, C. Leclerc, and S.T. Cole. 2003. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat Med* 9:533-539.

Ramage, H.R., L.E. Connolly, and J.S. Cox. 2009. Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. *PLoS Genet* 5:e1000767.

Rangaka, M.X., L. Diwakar, R. Seldon, G. van Cutsem, G.A. Meintjes, C. Morroni, P. Mouton, M.S. Shey, G. Maartens, K.A. Wilkinson, and R.J. Wilkinson. 2007a. Clinical, immunological, and epidemiological importance of antituberculosis T cell responses in HIV-infected Africans. *Clin Infect Dis* 44:1639-1646.

Rangaka, M.X., K.A. Wilkinson, R. Seldon, G. Van Cutsem, G.A. Meintjes, C. Morroni, P. Mouton, L. Diwakar, T.G. Connell, G. Maartens, and R.J.

Wilkinson. 2007b. Effect of HIV-1 infection on T-Cell-based and skin test detection of tuberculosis infection. *Am J Respir Crit Care Med* 175:514-520.

Rao, P.K., G.M. Rodriguez, I. Smith, and Q. Li. 2008. Protein dynamics in iron-starved Mycobacterium tuberculosis revealed by turnover and abundance measurement using hybrid-linear ion trap-Fourier transform mass spectrometry. *Anal Chem* 80:6860-6869.

Ravn, P., A. Demissie, T. Eguale, H. Wondwosson, D. Lein, H.A. Amoudy, A.S. Mustafa, A.K. Jensen, A. Holm, I. Rosenkrands, F. Oftung, J. Olobo, F. von Reyn, and P. Andersen. 1999. Human T cell responses to the ESAT-6 antigen from *Mycobacterium tuberculosis*. *J Infect Dis* 179:637-645.

Richeldi, L., M. Losi, R. D'Amico, M. Luppi, A. Ferrari, C. Mussini, M. Codeluppi, S. Cocchi, F. Prati, V. Paci, M. Meacci, B. Meccugni, F. Rumpianesi, P. Roversi, S. Cerri, F. Luppi, G. Ferrara, I. Latorre, G.E. Gerunda, G. Torelli, R. Esposito, and L.M. Fabbri. 2009. Performance of tests for latent tuberculosis in different groups of immunocompromised patients. *Chest* 136:198-204.

Ritz, N., W.A. Hanekom, R. Robins-Browne, W.J. Britton, and N. Curtis. 2008. Influence of BCG vaccine strain on the immune response and protection against tuberculosis. *FEMS Microbiol Rev* 32:821-841.

Roberts, D.M., R.P. Liao, G. Wisedchaisri, W.G. Hol, and D.R. Sherman. 2004. Two sensor kinases contribute to the hypoxic response of *Mycobacterium tuberculosis*. *J Biol Chem* 279:23082-23087.

Roupie, V., M. Romano, L. Zhang, H. Korf, M.Y. Lin, K.L. Franken, T.H. Ottenhoff, M.R. Klein, and K. Huygen. 2007. Immunogenicity of eight dormancy regulon-encoded proteins of *Mycobacterium tuberculosis* in DNA-vaccinated and tuberculosis-infected mice. *Infect Immun* 75:941-949.

Roxas, B.A., and Q. Li. 2009. Acid stress response of a mycobacterial proteome: insight from a gene ontology analysis. *Int J Clin Exp Med* 2:309-328.

Russell, D.G., C.E. Barry, 3rd, and J.L. Flynn. 2010. Tuberculosis: what we don't know can, and does, hurt us. *Science* 328:852-856.

Rustad, T.R., M.I. Harrell, R. Liao, and D.R. Sherman. 2008. The enduring hypoxic response of *Mycobacterium tuberculosis*. *PLoS One* 3:e1502.

Rustad, T.R., A.M. Sherrid, K.J. Minch, and D.R. Sherman. 2009. Hypoxia: a window into *Mycobacterium tuberculosis* latency. *Cell Microbiol* 11:1151-1159.

Sassetti, C.M., D.H. Boyd, and E.J. Rubin. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 48:77-84.

Schnappinger, D., S. Ehrt, M.I. Voskuil, Y. Liu, J.A. Mangan, I.M. Monahan, G. Dolganov, B. Efron, P.D. Butcher, C. Nathan, and G.K. Schoolnik. 2003. Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. *J Exp Med* 198:693-704.

Scholvinck, E., K.A. Wilkinson, A.O. Whelan, A.R. Martineau, M. Levin, and R.J. Wilkinson. 2004. Gamma interferon-based immunodiagnosis of tuberculosis: comparison between whole-blood and enzyme-linked immunospot methods. *J Clin Microbiol* 42:829-831.

Schuck, S.D., H. Mueller, F. Kunitz, A. Neher, H. Hoffmann, K.L. Franken, D. Repsilber, T.H. Ottenhoff, S.H. Kaufmann, and M. Jacobsen. 2009. Identification of T-cell antigens specific for latent mycobacterium tuberculosis infection. *PLoS One* 4:e5590.

Sharma, D., A. Bose, H. Shakila, T.K. Das, J.S. Tyagi, and V.D. Ramanathan. 2006. Expression of mycobacterial cell division protein, FtsZ, and dormancy proteins, DevR and Acr, within lung granulomas throughout guinea pig infection. *FEMS Immunol Med Microbiol* 48:329-336.

Sherman, D.R., M. Voskuil, D. Schnappinger, R. Liao, M.I. Harrell, and G.K. Schoolnik. 2001. Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha -crystallin. *Proc Natl Acad Sci U S A* 98:7534-7539.

Shi, L.B., Y.J. Jung, S. Tyagi, M.L. Gennaro, and R.J. North. 2003. Expression of Th1-mediated immunity in mouse lungs induces a *Mycobacterium tuberculosis* transcription pattern characteristic of nonreplicating persistence. *Proceedings of the National Academy of Sciences of the United States of America* 100:241-246.

Sidders, B., C. Pirson, P.J. Hogarth, R.G. Hewinson, N.G. Stoker, H.M. Vordermeier, and K. Ewer. 2008. Screening of highly expressed mycobacterial genes identifies Rv3615c as a useful differential diagnostic antigen for the *Mycobacterium tuberculosis* complex. *Infect Immun* 76:3932-3939.

Singh, H., and G.P. Raghava. 2001. ProPred: prediction of HLA-DR binding sites. *Bioinformatics* 17:1236-1237.

Smollett, K.L., A.S. Fivian-Hughes, J.E. Smith, A. Chang, T. Rao, and E.O. Davis. 2009. Experimental determination of translational start sites resolves uncertainties in genomic open reading frame predictions - application to *Mycobacterium tuberculosis*. *Microbiology* 155:186-197.

Soares, A.P., T.J. Scriba, S. Joseph, R. Harbacheuski, R.A. Murray, S.J. Gelderbloem, A. Hawkridge, G.D. Hussey, H. Maecker, G. Kaplan, and W.A. Hanekom. 2008. Bacillus Calmette-Guerin vaccination of human newborns induces T cells with complex cytokine and phenotypic profiles. *J Immunol* 180:3569-3577.

Sorensen, A.L., S. Nagai, G. Houen, P. Andersen, and A.B. Andersen. 1995. Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. *Infect Immun* 63:1710-1717.

Steingart, K.R., N. Dendukuri, M. Henry, I. Schiller, P. Nahid, P.C. Hopewell, A. Ramsay, M. Pai, and S. Laal. 2009. Performance of purified antigens for serodiagnosis of pulmonary tuberculosis: a meta-analysis. *Clin Vaccine Immunol* 16:260-276.

Steingart, K.R., M. Henry, V. Ng, P.C. Hopewell, A. Ramsay, J. Cunningham, R. Urbanczik, M. Perkins, M.A. Aziz, and M. Pai. 2006. Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *Lancet Infect Dis* 6:570-581.

Stern, L.J., and J.M. Calvo-Calle. 2009. HLA-DR: molecular insights and vaccine design. *Curr Pharm Des* 15:3249-3261.

Stewart, G.R., S.M. Newton, K.A. Wilkinson, I.R. Humphreys, H.N. Murphy, B.D. Robertson, R.J. Wilkinson, and D.B. Young. 2005. The stress-responsive chaperone alpha-crystallin 2 is required for pathogenesis of *Mycobacterium tuberculosis*. *Mol Microbiol* 55:1127-1137.

Stewart, G.R., L. Wernisch, R. Stabler, J.A. Mangan, J. Hinds, K.G. Laing, D.B. Young, and P.D. Butcher. 2002. Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. *Microbiology* 148:3129-3138.

Sturniolo, T., E. Bono, J. Ding, L. Raddrizzani, O. Tuereci, U. Sahin, M. Braxenthaler, F. Gallazzi, M.P. Protti, F. Sinigaglia, and J. Hammer. 1999. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* 17:555-561.

Tena-Coki, N.G., T.J. Scriba, N. Peteni, B. Eley, R.J. Wilkinson, P. Andersen, W.A. Hanekom, and B. Kampmann. 2010. CD4 and CD8 T Cell Responses to Mycobacterial Antigens in African Children. *Am J Respir Crit Care Med*

Terasaki, P.I., M. Ozawa, and R. Castro. 2007. Four-year follow-up of a prospective trial of HLA and MICA antibodies on kidney graft survival. *Am J Transplant* 7:408-415.

Timm, J., F.A. Post, L.G. Bekker, G.B. Walther, H.C. Wainwright, R. Manganelli, W.T. Chan, L. Tsenova, B. Gold, I. Smith, G. Kaplan, and J.D. McKinney. 2003. Differential expression of iron-, carbon-, and oxygen-responsive mycobacterial genes in the lungs of chronically infected mice and tuberculosis patients. *Proc Natl Acad Sci U S A* 100:14321-14326.

Todryk, S.M., A.A. Pathan, S. Keating, D.W. Porter, T. Berthoud, F. Thompson, P. Klenerman, and A.V. Hill. 2009. The relationship between human effector and memory T cells measured by ex vivo and cultured ELISPOT following recent and distal priming. *Immunology* 128:83-91.

Trunz, B.B., P. Fine, and C. Dye. 2006. Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. *Lancet* 367:1173-1180.

Tsurui, H., and T. Takahashi. 2007. Prediction of T-cell epitope. *J Pharmacol Sci* 105:299-316.

Ulrichs, T., G.A. Kosmiadi, S. Jorg, L. Pradl, M. Titukhina, V. Mishenko, N. Gushina, and S.H. Kaufmann. 2005. Differential Organization of the Local Immune Response in Patients with Active Cavitary Tuberculosis or with Nonprogressive Tuberculoma. *J Infect Dis* 192:89-97.

van Crevel, R., T.H. Ottenhoff, and J.W. van der Meer. 2002. Innate immunity to *Mycobacterium tuberculosis*. *Clin Microbiol Rev* 15:294-309.

van Pinxteren, L.A., P. Ravn, E.M. Agger, J. Pollock, and P. Andersen. 2000. Diagnosis of tuberculosis based on the two specific antigens ESAT-6 and CFP10. *Clin Diagn Lab Immunol* 7:155-160.

Via, L.E., P.L. Lin, S.M. Ray, J. Carrillo, S.S. Allen, S.Y. Eum, K. Taylor, E. Klein, U. Manjunatha, J. Gonzales, E.G. Lee, S.K. Park, J.A. Raleigh, S.N. Cho, D.N. McMurray, J.L. Flynn, and C.E. Barry, 3rd. 2008. Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. *Infect Immun* 76:2333-2340.

Vijayalakshmi, V., K.J. Murthy, S. Kumar, and A.L. Kiran. 1995. Comparison of the immune responses in children vaccinated with three strains of BCG vaccine. *Indian Pediatr* 32:979-982.

Vincenti, D., S. Carrara, P. De Mori, L.P. Pucillo, N. Petrosillo, F. Palmieri, O. Armignacco, G. Ippolito, E. Girardi, M. Amicosante, and D. Goletti. 2003. Identification of early secretory antigen target-6 epitopes for the immunodiagnosis of active tuberculosis. *Mol Med* 9:105-111.

Vordermeier, H.M., M.A. Chambers, P.J. Cockle, A.O. Whelan, J. Simmons, and R.G. Hewinson. 2002. Correlation of ESAT-6-specific gamma interferon production with pathology in cattle following *Mycobacterium bovis* BCG vaccination against experimental bovine tuberculosis. *Infect Immun* 70:3026-3032.

Vordermeier, M., S.V. Gordon, and A.R. Hewinson. 2009. Antigen mining to define *Mycobacterium bovis* antigens for the differential diagnosis of vaccinated and infected animals: A VLA perspective. *Transbound Emerg Dis* 56:240-247.

Vordermeier, M., A.O. Whelan, and R.G. Hewinson. 2003. Recognition of mycobacterial epitopes by T cells across mammalian species and use of a program that predicts human HLA-DR binding peptides to predict bovine epitopes. *Infect Immun* 71:1980-1987.

Voskuil, M.I., D. Schnappinger, K.C. Visconti, M.I. Harrell, G.M. Dolganov, D.R. Sherman, and G.K. Schoolnik. 2003. Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. *J Exp Med* 198:705-713.

Voskuil, M.I., K.C. Visconti, and G.K. Schoolnik. 2004. *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. *Tuberculosis (Edinb)* 84:218-227.

Walker, V., G. Selby, and I. Wacogne. 2006. Does neonatal BCG vaccination protect against tuberculous meningitis? *Arch Dis Child* 91:789-791.

Walters, S.B., E. Dubnau, I. Kolesnikova, F. Laval, M. Daffe, and I. Smith. 2006. The Mycobacterium tuberculosis PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. *Mol Microbiol* 60:312-330.

Wang, J., Y. Qie, B. Zhu, H. Zhang, Y. Xu, Q. Wang, J. Chen, W. Liu, and H. Wang. 2009a. Evaluation of a recombinant BCG expressing antigen Ag85B and PPE protein Rv3425 from DNA segment RD11 of *Mycobacterium tuberculosis* in C57BL/6 mice. *Med Microbiol Immunol* 198:5-11.

Wang, J.H., R. Meijers, Y. Xiong, J.H. Liu, T. Sakihama, R. Zhang, A. Joachimiak, and E.L. Reinherz. 2001. Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. *Proc Natl Acad Sci U S A* 98:10799-10804.

Wang, M., L. Harhaji, K. Lamberth, M. Harndahl, S. Buus, N.H. Heegaard, M.H. Claesson, and M.H. Nissen. 2009b. Modified human beta 2-microglobulin (desLys(58)) displays decreased affinity for the heavy chain of MHC class I and induces nitric oxide production and apoptosis. *Scand J Immunol* 69:203-212.

Wayne, L.G., and C.D. Sohaskey. 2001. Nonreplicating persistence of *mycobacterium tuberculosis. Annu Rev Microbiol* 55:139-163.

Weber, C.A., P.J. Mehta, M. Ardito, L. Moise, B. Martin, and A.S. De Groot. 2009. T cell epitope: friend or foe? Immunogenicity of biologics in context. *Adv Drug Deliv Rev* 61:965-976.

Weir, R.E., A.R. Morgan, W.J. Britton, C.R. Butlin, and H.M. Dockrell. 1994. Development of a whole blood assay to measure T cell responses to leprosy: a new tool for immuno-epidemiological field studies of leprosy immunity. *J Immunol Methods* 176:93-101.

Weis, S.E., P.C. Slocum, F.X. Blais, B. King, M. Nunn, G.B. Matney, E. Gomez, and B.H. Foresman. 1994. The effect of directly observed therapy on the rates of drug resistance and relapse in tuberculosis. *N Engl J Med* 330:1179-1184.

WHO. 2010. Global tuberculosis control: key findings from the December 2009 WHO report. *Wkly Epidemiol Rec* 85:69-80.

Wilkinson, K.A., O.M. Kon, S.M. Newton, G. Meintjes, R.N. Davidson, G. Pasvol, and R.J. Wilkinson. 2006. Effect of treatment of latent tuberculosis infection on the T cell response to *Mycobacterium tuberculosis* antigens. *J Infect Dis* 193:354-359.

Wilkinson, K.A., R. Seldon, G. Meintjes, M.X. Rangaka, W.A. Hanekom, G. Maartens, and R.J. Wilkinson. 2009. Dissection of regenerating T-Cell responses against tuberculosis in HIV-infected adults sensitized by *Mycobacterium tuberculosis*. *Am J Respir Crit Care Med* 180:674-683.

Wilkinson, K.A., G.R. Stewart, S.M. Newton, H.M. Vordermeier, J.R. Wain, H.N. Murphy, K. Horner, D.B. Young, and R.J. Wilkinson. 2005a. Infection biology of a novel alpha-crystallin of *Mycobacterium tuberculosis*: Acr2. *J Immunol* 174:4237-4243.

Wilkinson, K.A., M.H. Vordermeier, J. Kajtar, S. Jurcevic, R. Wilkinson, J. Ivanyi, and F. Hudecz. 1997a. Modulation of peptide specific T cell responses by non-native flanking regions. *Mol Immunol* 34:1237-1246.

Wilkinson, K.A., R.J. Wilkinson, A. Pathan, K. Ewer, M. Prakash, P. Klenerman, N. Maskell, R. Davies, G. Pasvol, and A. Lalvani. 2005b. Ex vivo

characterization of early secretory antigenic target 6-specific T cells at sites of active disease in pleural tuberculosis. *Clin Infect Dis* 40:184-187.

Wilkinson, R.J., K. Haslov, R. Rappuoli, F. Giovannoni, P.R. Narayanan, C.R. Desai, H.M. Vordermeier, J. Paulsen, G. Pasvol, J. Ivanyi, and M. Singh. 1997b. Evaluation of the recombinant 38-kilodalton antigen of *Mycobacterium tuberculosis* as a potential immunodiagnostic reagent. *J Clin Microbiol* 35:553-557.

Wilkinson, R.J., H.M. Vordermeier, K.A. Wilkinson, A. Sjolund, C. Moreno, G. Pasvol, and J. Ivanyi. 1998a. Peptide-specific T cell response to *Mycobacterium tuberculosis*: clinical spectrum, compartmentalization, and effect of chemotherapy. *J Infect Dis* 178:760-768.

Wilkinson, R.J., K.A. Wilkinson, K.A. De Smet, K. Haslov, G. Pasvol, M. Singh, I. Svarcova, and J. Ivanyi. 1998b. Human T- and B-cell reactivity to the 16kDa alpha-crystallin protein of *Mycobacterium tuberculosis*. *Scand J Immunol* 48:403-409.

Wu, B., C. Huang, L. Garcia, A. Ponce de Leon, J.S. Osornio, M. Bobadilla-del-Valle, L. Ferreira, S. Canizales, P. Small, M. Kato-Maeda, A.M. Krensky, and C. Clayberger. 2007. Unique gene expression profiles in infants vaccinated with different strains of *Mycobacterium bovis* bacille Calmette-Guerin. *Infect Immun* 75:3658-3664.

Yeaman, M.R., K.D. Gank, A.S. Bayer, and E.P. Brass. 2002. Synthetic peptides that exert antimicrobial activities in whole blood and blood-derived matrices. *Antimicrob Agents Chemother* 46:3883-3891.

Young, D.B., H.P. Gideon, and R.J. Wilkinson. 2009. Eliminating latent tuberculosis. *Trends Microbiol* 17:183-188.

Zhang, Y. 2004. Persistent and dormant tubercle bacilli and latent tuberculosis. *Front Biosci* 9:1136-1156.

Zucker, S.D., X. Qin, S.D. Rouster, F. Yu, R.M. Green, P. Keshavan, J. Feinberg, and K.E. Sherman. 2001. Mechanism of indinavir-induced hyperbilirubinemia. *Proc Natl Acad Sci U S A* 98:12671-12676.

Zvi, A., N. Ariel, J. Fulkerson, J.C. Sadoff, and A. Shafferman. 2008. Whole genome identification of *Mycobacterium tuberculosis* vaccine candidates by comprehensive data mining and bioinformatic analyses. *BMC Med Genomics* 1:18.

Appendix

Appendix A

A.1 Synthetic overlapping peptides obtained from Veterinary Laboratories Agency

16mer peptides with 8 αα overlap

A.1.1 ESAT-6

ESAT-6	**Sequence**
ESAT-6 (1-16)	MTEQQWNFAGIEAAAS
ESAT-6 (9-24)	AGIEAAASAIQGNVTS
ESAT-6 (17-32)	AIQGNVTSIHSLLDEG
ESAT-6 (25-40)	IHSLLDEGKQSLTKLA
ESAT-6 (33-48)	KQSLTKLAAAWGGSGS
ESAT-6 (41-56)	AAWGGSGSEAYQGVQQ
ESAT-6 (49-64)	EAYQGVQQKWDATATE
ESAT-6 (57-72)	KWDATATELNNALQNL
ESAT-6 (65-80)	LNNALQNLARTISEAG
ESAT-6 (73-88)	ARTISEAGQAMASTEG
ESAT-6 (81-95)	QAMASTEGNVTGMFA

A.1.2 : CFP-10

CFP-10	**Sequence**
CFP-10 (1-16)	MAEMKTDAATLAQEAG
CFP-10 (9-24)	ATLAQEAGNFERISGD
CFP-10 (17-32)	NFERISGDLKTQIDQV
CFP-10 (25-40)	LKTQIDQVESTAGSLQ
CFP-10 (33-48)	ESTAGSLQGQWRGAAG
CFP-10 (41-56)	GQWRGAAGTAAQAAVV
CFP-10 (49-64)	TAAQAAVVRFQEAANK
CFP-10 (57-72)	RFQEAANKQKQELDEI
CFP-10 (65-80)	QKQELDEISTNIRQAG
CFP-10 (73-88)	STNIRQAGVQYSRADE
CFP-10 (81-96)	VQYSRADEEQQQALSS
CFP-10 (89-100)	EQQQALSSQMGF

A.1.3 : Acr-1 (20 mers with 12αα overlap)

Acr-1	Sequence
Acr-1 (1-20)	MATTLPVQRHPRSLFPEFSE
Acr-1 (9-28)	RHPRSLFPEFSELFAAFPSF
Acr-1 (17-36)	EFSELFAAFPSFAGLRPTFD
Acr-1 (25-44)	FPSFAGLRPTFDTRLMRLED
Acr-1 (33-52)	PTFDTRLMRLEDEMKEGRYE
Acr-1 (41-60)	RLEDEMKEGRYEVRAELPGV
Acr-1 (49-68)	GRYEVRAELPGVDPDKDVDI
Acr-1 (57-76)	LPGVDPDKDVDIMVRDGQLT
Acr-1 (65-84)	DVDIMVRDGQLTIKAERTEQ
Acr-1 (73-92)	GQLTIKAERTEQKDFDGRSE
Acr-1 (81-100)	RTEQKDFDGRSEFAYGSFVR
Acr-1 (89-108)	GRSEFAYGSFVRTVSLPVGA
Acr-1 (97-116)	SFVRTVSLPVGADEDDIKAT
Acr-1 (105-124)	PVGADEDDIKATYDKGILTV
Acr-1 (113-132)	IKATYDKGILTVSVAVSEGK
Acr-1 (121-140)	ILTVSVAVSEGKPTEKHIQI
Acr-1 (129-144)	SEGKPTEKHIQIRSTN

A.2 Synthetic overlapping peptides covering 29 EHR antigens purchased from Peptide protein research Ltd, UK

20mer peptides with 10αα overlap

A.2.1 Rv0188

Rv0188	**Sequence**	**Pool**
Rv0188 (1-20)	MSTVHSSIDQHPDLLALRAS	1
Rv0188 (11-30)	HPDLLALRASFDRAAESTIA	1
Rv0188 (21-40)	FDRAAESTIAHFTFGLALLA	1
Rv0188 (31-50)	HFTFGLALLAGLYVAASPWI	1
Rv0188 (41-60)	GLYVAASPWIVGFSATRGLP	1
Rv0188 (51-70)	VGFSATRGLPTCDLIVGIAV	1
Rv0188 (61-80)	TCDLIVGIAVAYLAYGFASA	1
Rv0188 (71-90)	AYLAYGFASALDRTHGMTWT	1
Rv0188 (81-100)	LDRTHGMTWTLPVLGVWVIF	1
Rv0188 (91-110)	LPVLGVWVIFSPWVLPGVAV	1
Rv0188 (100-120)	SPWVLPGVAVTAGMMWSHII	1
Rv0188 (110-130)	TAGMMWSHIIAGAVVAVLGF	1
Rv0188 (120-143)	AGAVVAVLGFYFGMRTRAAANQG	1

A.2.3: Rv0767c

Rv0767c	Sequence	Pool
Rv0767c (1-20)	VSSDVLVTTPAQRQTEPHAE	1
Rv0767c (11-30)	AQRQTEPHAEAVSRNRRQQA	1
Rv0767c (21-40)	AVSRNRRQQATFRKVLAAAM	1
Rv0767c (31-50)	TFRKVLAAAMATLREKSYAD	1
Rv0767c (41-60)	ATLREKSYADLTVRLVAARA	1
Rv0767c (51-70)	LTVRLVAARAKVAPATAYTY	1
Rv0767c (61-80)	KVAPATAYTYFSSKNHLIAE	1
Rv0767c (71-90)	FSSKNHLIAEVYLDLVRQVP	1
Rv0767c (81-100)	VYLDLVRQVPCVTDVNVPMP	1
Rv0767c (91-110)	CVTDVNVPMPIRVTSSLRHL	1
Rv0767c (101-120)	IRVTSSLRHLALVVADEPEI	2
Rv0767c (111-130)	ALVVADEPEIGAACTAALLD	2
Rv0767c (121-140)	GAACTAALLDGGADPAVRAV	2
Rv0767c (131-150)	GGADPAVRAVRDRIGAEIHR	2
Rv0767c (141-160)	RDRIGAEIHRRITSAIGPGA	2
Rv0767c (151-170)	RITSAIGPGADPGTVFALEM	2
Rv0767c (161-180)	DPGTVFALEMAFFGALVQAG	2
Rv0767c (171-190)	AFFGALVQAGSGTFTYHEIA	2
Rv0767c (181-200)	SGTFTYHEIADRLGYVVGLI	2
Rv0767c (191-213)	DRLGYVVGLILAGANEPSTGGSE	2

A.2. 3: Rv0826

Rv0826	Sequence	Pool
Rv0826 (1-20)	VTQDTSATCPLTSTVQDSSP	1
Rv0826 (11-30)	LTSTVQDSSPVAGQLGRPIG	1
Rv0826 (21-40)	VAGQLGRPIGFRGLAGGCPV	1
Rv0826 (31-50)	FRGLAGGCPVSPLGYESPPL	1
Rv0826 (41-60)	SPLGYESPPLPLGPDSLTWR	1
Rv0826 (51-70)	PLGPDSLTWRYFGDWRGMLQ	1
Rv0826 (61-80)	YFGDWRGMLQGPWAGSMQNM	1
Rv0826 (71-90)	GPWAGSMQNMHPQLGAAVED	1
Rv0826 (81-100)	HPQLGAAVEDHSTFFRERWP	1
Rv0826 (91-110)	HSTFFRERWPRLLRSLYPIG	1
Rv0826 (101-120)	RLLRSLYPIGGVVFDGDRAP	1
Rv0826 (111-130)	GVVFDGDRAPVTGVQVRDYH	2
Rv0826 (121-140)	VTGVQVRDYHITIKGVDGAG	2
Rv0826 (131-150)	ITIKGVDGAGRRYHALNPDV	2
Rv0826 (141-160)	RRYHALNPDVFYWAHATFFV	2
Rv0826 (151-170)	FYWAHATFFVGTLHVAERFC	2
Rv0826 (161-180)	GTLHVAERFCGGLTEAQRRQ	2
Rv0826 (171-190)	GGLTEAQRRQLFDEHVQWYR	2
Rv0826 (181-200)	LFDEHVQWYRMYGMSMRPVP	2
Rv0826 (191-210)	MYGMSMRPVPATWEEFQDYW	2
Rv0826 (201-220)	ATWEEFQDYWDHMCRNVLEN	2
Rv0826 (211-230)	DHMCRNVLENNFAARAVLDL	2
Rv0826 (221-240)	NFAARAVLDLTELPKPPFAQ	2
Rv0826 (231-250)	TELPKPPFAQRVPDWLWAAP	3
Rv0826 (241-260)	RVPDWLWAAPRKLLARFFVW	3
Rv0826 (251-270)	RKLLARFFVWLTVGLYDPPV	3
Rv0826 (261-280)	LTVGLYDPPVRELMGYRWLR	3
Rv0826 (271-290)	RELMGYRWLRRDEWLHRRFG	3
Rv0826 (281-300)	RDEWLHRRFGDIVRLVFALV	3
Rv0826 (291-310)	DIVRLVFALVPFRFRKHPRA	3
Rv0826 (301-320)	PFRFRKHPRARAGWDRATGR	3
Rv0826 (311-330)	RAGWDRATGRIPADAPLVQT	3
Rv0826 (321-340)	IPADAPLVQTPARNLPPPDE	3
Rv0826 (331-350)	PARNLPPPDERDNPTHYCPK	3
Rv0826 (341-351)	RDNPTHYCPKV	3

A.2.4. Rv0847

Rv0847	Sequence	Pool
Rv0847 (1-20)	VVWMRSAIVAVALGVTVAAV	1
Rv0847 (11-30)	VALGVTVAAVAAACWLPQLH	1
Rv0847 (21-40)	AAACWLPQLHRHVAHPNHPL	1
Rv0847 (31-50)	RHVAHPNHPLTTSVGSEFVI	1
Rv0847 (41-60)	TTSVGSEFVINTDHGHLVDN	1
Rv0847 (51-70)	NTDHGHLVDNSMPPCPERLA	1
Rv0847 (61-80)	SMPPCPERLATAVLPRSATP	1
Rv0847 (71-90)	TAVLPRSATPVLLPDVVAAA	1
Rv0847 (81-100)	VLLPDVVAAAPGMTAALTDP	1
Rv0847 (91-110)	PGMTAALTDPVAPAARGPPA	1
Rv0847 (101-120)	VAPAARGPPAAQGSVRTGQD	1
Rv0847 (111-130)	AQGSVRTGQDLLTRFCLARR	1

A.2.5: Rv0849

Rv0849	Sequence	Pool
Rv0849 (1-20)	MGARAIFRGFNRPSRVLMIN	1
Rv0849 (11-30)	NRPSRVLMINQFGINIGFYM	1
Rv0849 (21-40)	QFGINIGFYMLMPYLADYLA	1
Rv0849 (31-50)	LMPYLADYLAGPLGLAAWAV	1
Rv0849 (41-60)	GPLGLAAWAVGLVMGVRNFS	1
Rv0849 (51-70)	GLVMGVRNFSQQGMFFVGGT	1
Rv0849 (61-80)	QQGMFFVGGTLADRFGYKPL	1
Rv0849 (71-90)	LADRFGYKPLIIAGCLIRTG	1
Rv0849 (81-100)	IIAGCLIRTGGFALLVVAQS	1
Rv0849 (91-110)	GFALLVVAQSLPSVLIAAAA	1
Rv0849 (101-120)	LPSVLIAAAATGFAGALFNP	2
Rv0849 (111-130)	TGFAGALFNPAVRGYLAAEA	2
Rv0849 (121-140)	AVRGYLAAEAGERKIEAFAM	2
Rv0849 (131-150)	GERKIEAFAMFNVFYQSGIL	2
Rv0849 (141-160)	FNVFYQSGILLGPLVGLVLL	2
Rv0849 (151-170)	LGPLVGLVLLALDFRITVLA	2
Rv0849 (161-180)	ALDFRITVLAAAGVFGLLTV	2
Rv0849 (171-190)	AAGVFGLLTVAQLVALPQHR	2
Rv0849 (181-200)	AQLVALPQHRADSEREKTSI	2
Rv0849 (191-210)	ADSEREKTSILQDWRVVVRN	2
Rv0849 (201-220)	LQDWRVVVRNRPFLTLAAAM	3
Rv0849 (211-230)	RPFLTLAAAMTGCYALSFQI	3
Rv0849 (221-240)	TGCYALSFQIYLALPMQASI	3
Rv0849 (231-250)	YLALPMQASILMPRNQYLLI	3
Rv0849 (241-260)	LMPRNQYLLIAAMFAVSGLV	3
Rv0849 (251-270)	AAMFAVSGLVAVGGQLRITR	3
Rv0849 (261-280)	AVGGQLRITRWFAVRWGAER	3
Rv0849 (271-290)	WFAVRWGAERSLVVGATILA	3
Rv0849 (281-300)	SLVVGATILAASFIPVAVIP	3
Rv0849 (291-310)	ASFIPVAVIPNGQRFGVAVA	3
Rv0849 (301-320)	NGQRFGVAVAVMALVLSASL	4
Rv0849 (311-330)	VMALVLSASLLAVASAALFP	4
Rv0849 (321-340)	LAVASAALFPFEMRAVVALS	4
Rv0849 (331-350)	FEMRAVVALSGDRLVATHYG	4
Rv0849 (341-350)	GDRLVATHYGFYSTIVGVGV	4
Rv0849 (351-360)	FYSTIVGVGVLVGNLAIGSL	4
Rv0849 (361-370)	LVGNLAIGSLMSAARRLNTD	4
Rv0849 (371-380)	MSAARRLNTDEIVWGGLILV	4
Rv0849 (381-390)	EIVWGGLILVGIVAVAGLRR	4
Rv0849 (391-400)	GIVAVAGLRRLDTFTSGSQN	4
Rv0849 (401-409)	LDTFTSGSQNMTGRWAAPR	4

A.2.6 : Rv0967

Rv0967	**Sequence**	**Pool**
Rv0967 (1-20)	MSKELTAKKRAALNRLKTVR	1
Rv0967 (11-30)	AALNRLKTVRGHLDGIVRML	1
Rv0967 (21-40)	GHLDGIVRMLESDAYCVDVM	1
Rv0967 (31-50)	ESDAYCVDVMKQISAVQSSL	1
Rv0967 (41-60)	KQISAVQSSLERANRVMLHN	1
Rv0967 (51-70)	ERANRVMLHNHLETCFSTAV	1
Rv0967 (61-80)	HLETCFSTAVLDGHGQAAIE	1
Rv0967 (71-90)	LDGHGQAAIEELIDAVKFTP	1
Rv0967 (81-100)	ELIDAVKFTPALTGPHARLG	1
Rv0967 (91-110)	ALTGPHARLGGAAVGESATE	1
Rv0967 (101-119)	GAAVGESATEEPMPDASNM	1

A.2.7 Rv0990c

Rv0990c	Sequence	Pool
Rv0990c (1-20)	VAESSLNPSLVSRISAFLRP	1
Rv0990c (11-30)	VSRISAFLRPDWTRTVRARR	1
Rv0990c (21-40)	DWTRTVRARRFAAAGLVMLA	1
Rv0990c (31-50)	FAAAGLVMLAGVAALRSNPE	1
Rv0990c (41-60)	GVAALRSNPEDDRAEVVVAA	1
Rv0990c (51-70)	DDRAEVVVAAHDLRPGTALT	1
Rv0990c (61-80)	HDLRPGTALTPGDVRLEKRS	1
Rv0990c (71-90)	PGDVRLEKRSATTLPDGSQA	1
Rv0990c (81-100)	ATTLPDGSQADLDAVVGSTL	1
Rv0990c (91-110)	DLDAVVGSTLASPTRRGEVL	1
Rv0990c (101-120)	ASPTRRGEVLTDVRLLGSRL	2
Rv0990c (111-130)	TDVRLLGSRLAESTAGPDAR	2
Rv0990c (121-140)	AESTAGPDARIVPLHLADSA	2
Rv0990c (131-150)	IVPLHLADSALVDLVRVGDV	2
Rv0990c (141-160)	LVDLVRVGDVVDVLAAPVTD	2
Rv0990c (151-170)	VDVLAAPVTDSPAALRLLAT	2
Rv0990c (161-180)	SPAALRLLATDAIVVLVSAQ	2
Rv0990c (171-190)	DAIVVLVSAQQKAQAADSDR	2
Rv0990c (181-200)	QKAQAADSDRVVLVALPARL	2
Rv0990c (191-210)	VVLVALPARLANTVAGAALG	2
Rv0990c (201-218)	ANTVAGAALGQTVTLTLH	2

A.2.8. Rv0991c

Rv0991c	Sequence	Pool
Rv0991c (1-20)	VPTYSYECTQCANRFDVVQA	1
Rv0991c (11-30)	CANRFDVVQAFTDDALTTCE	1
Rv0991c (21-40)	FTDDALTTCERCSGRLRKLF	1
Rv0991c (31-50)	RCSGRLRKLFNAVGVVFKGT	1
Rv0991c (41-60)	NAVGVVFKGTGFYRTDSRES	1
Rv0991c (51-70)	GFYRTDSRESGKKSKSQTNG	1
Rv0991c (61-80)	GKKSKSQTNGSSTSESTKSS	1
Rv0991c (71-90)	SSTSESTKSSGSSGSSGSSE	1
Rv0991c (81-100)	GSSGSSGSSESKASGSTEKS	1
Rv0991c (91-110)	SKASGSTEKSTSSTTAAAAV	1

A.2.9. Rv1284

Rv1284	Sequence	Pool
Rv1284 (1-20)	VTVTDDYLANNVDYASGFKG	1
Rv1284 (11-30)	NVDYASGFKGPLPMPPSKHI	1
Rv1284 (21-40)	PLPMPPSKHIAIVACMDARL	1
Rv1284 (31-50)	AIVACMDARLDVYRMLGIKE	1
Rv1284 (41-60)	DVYRMLGIKEGEAHVIRNAG	1
Rv1284 (51-70)	GEAHVIRNAGCVVTDDVIRS	1
Rv1284 (61-80)	CVVTDDVIRSLAISQRLLGT	1
Rv1284 (71-90)	LAISQRLLGTREIILLHHTD	2
Rv1284 (81-100)	REIILLHHTDCGMLTFTDDD	2
Rv1284 (91-110)	CGMLTFTDDDFKRAIQDETG	2
Rv1284 (101-120)	FKRAIQDETGIRPTWSPESY	2
Rv1284 (111-130)	IRPTWSPESYPDAVEDVRQS	2
Rv1284 (121-140)	PDAVEDVRQSLRRIEVNPFV	2
Rv1284 (131-150)	LRRIEVNPFVTKHTSLRGFV	2
Rv1284 (141-163)	TKHTSLRGFVFDVATGKLNEVTP	2

A.2.10 Rv1471

Rv1471	**Sequence**	**Pool**
Rv1471 (1-20)	VTTRDLTAAQFNETIQSSDM	1
Rv1471 (11-30)	FNETIQSSDMVLVDYWASWC	1
Rv1471 (21-40)	VLVDYWASWCGPCRAFAPTF	1
Rv1471 (31-50)	GPCRAFAPTFAESSEKHPDV	1
Rv1471 (41-60)	AESSEKHPDVVHAKVDTEAE	1
Rv1471 (51-70)	VHAKVDTEAERELAAAAQIR	1
Rv1471 (61-80)	RELAAAAQIRSIPTIMAFKN	1
Rv1471 (71-90)	SIPTIMAFKNGKLLFNQAGA	1
Rv1471 (81-100)	GKLLFNQAGALPPAALESLV	1
Rv1471 (91-110)	LPPAALESLVQQLKAYEVEA	1
Rv1471 (101-123)	QQLKAYEVEAGEATTQNGRAQQA	1

A.2.11 Rv1954c

Rv1954c	Sequence	Pool
Rv1954c (1-20)	MAAGSGGGTVGLVLPRVASL	1
Rv1954c (11-30)	GLVLPRVASLSGLDGAPTVP	1
Rv1954c (21-40)	SGLDGAPTVPEGSDKALMHL	1
Rv1954c (31-50)	EGSDKALMHLGDPPRRCDTH	1
Rv1954c (41-60)	GDPPRRCDTHPDGTSSAAAA	1
Rv1954c (51-70)	PDGTSSAAAALVLRRIDVHP	1
Rv1954c (61-80)	LVLRRIDVHPLLTGLGRGRQ	1
Rv1954c (71-90)	LLTGLGRGRQTVSLRNGHLV	1
Rv1954c (81-100)	TVSLRNGHLVATANRAILSR	2
Rv1954c (91-110)	ATANRAILSRRRSRLTRGRS	2
Rv1954c (101-120)	RRSRLTRGRSFTSHLITSCP	2
Rv1954c (111-130)	FTSHLITSCPRLDDHQHRHP	2
Rv1954c (121-140)	RLDDHQHRHPTRCRAEHAGC	2
Rv1954c (131-150)	TRCRAEHAGCTVATCIPNAR	2
Rv1954c (141-160)	TVATCIPNARDPAPGHQTPR	2
Rv1954c (151-173)	DPAPGHQTPRWGPFRLKPAYTRI	2

A.2.12 Rv1955

Rv1955	Sequence	Pool
Rv1955 (1-20)	VPSGWVSHRLGGSPKCISAL	1
Rv1955 (11-30)	GGSPKCISALSLPSGTVGAP	1
Rv1955 (21-40)	SLPSGTVGAPSKPDNDATRG	1
Rv1955 (31-50)	SKPDNDATRGRTRPTVPPPD	1
Rv1955 (41-60)	RTRPTVPPPDPAAMGTWKFF	1
Rv1955 (51-70)	PAAMGTWKFFRASVDGRPVF	1
Rv1955 (61-80)	RASVDGRPVFKKEFDKLPDQ	1
Rv1955 (71-90)	KKEFDKLPDQARAALIVLMQ	1
Rv1955 (81-100)	ARAALIVLMQRYLVGDLAAG	2
Rv1955 (91-110)	RYLVGDLAAGSIKPIRGDIL	2
Rv1955 (101-120)	SIKPIRGDILELRWHEANNH	2
Rv1955 (111-130)	ELRWHEANNHFRVLFFRWGQ	2
Rv1955 (121-140)	FRVLFFRWGQHPVALTAFYK	2
Rv1955 (131-150)	HPVALTAFYKNQQKTPKTKI	2
Rv1955 (141-160)	NQQKTPKTKIETALDRQKIW	2
Rv1955 (151-170)	ETALDRQKIWKRAFGDTPPI	2

A.2..13 Rv1956

Rv1956	**Sequence**	**Pool**
Rv1956 (1-20)	MSIDFPLGDDLAGYIAEAIA	1
Rv1956 (11-30)	LAGYIAEAIAADPSFKGTLE	1
Rv1956 (21-40)	ADPSFKGTLEDAEEARRLVD	1
Rv1956 (31-50)	DAEEARRLVDALIALRKHCQ	1
Rv1956 (41-60)	ALIALRKHCQLSQVEVAKRM	1
Rv1956 (51-70)	LSQVEVAKRMGVRQPTVSGF	1
Rv1956 (61-80)	GVRQPTVSGFEKEPSDPKLS	1
Rv1956 (71-90)	EKEPSDPKLSTLQRYARALD	1
Rv1956 (81-100)	TLQRYARALDARLRLVLEVP	1
Rv1956 (91-110)	ARLRLVLEVPTLREVPTWHR	1
Rv1956 (101-120)	TLREVPTWHRLSSYRGSARD	1
Rv1956 (111-130)	LSSYRGSARDHQVRVGADKE	1
Rv1956 (121-140)	HQVRVGADKEILMQTNWARH	1
Rv1956 (131-149)	ILMQTNWARHISVRQVEVA	1

A.2.14 Rv1957

Rv1957	Sequence	Pool
Rv1957 (1-20)	MTDRTDADDLDLQRVGARLA	1
Rv1957 (11-30)	DLQRVGARLAARAQIRDIRL	1
Rv1957 (21-40)	ARAQIRDIRLLRTQAAVHRA	1
Rv1957 (31-50)	LRTQAAVHRAPKPAQGLTYD	1
Rv1957 (41-60)	PKPAQGLTYDLEFEPAVDAD	1
Rv1957 (51-70)	LEFEPAVDADPATISAFVVR	1
Rv1957 (61-80)	PATISAFVVRISCHLRIQNQ	1
Rv1957 (71-90)	ISCHLRIQNQAADDDVKEGD	1
Rv1957 (81-100)	AADDDVKEGDTKDETQDVAT	2
Rv1957 (91-110)	TKDETQDVATADFEFAALFD	2
Rv1957 (101-120)	ADFEFAALFDYHLQEGEDDP	2
Rv1957 (111-130)	YHLQEGEDDPTEEELTAYAA	2
Rv1957 (121-140)	TEEELTAYAATTGRFALYPY	2
Rv1957 (131-150)	TTGRFALYPYIREYVYDLTG	2
Rv1957 (141-160)	IREYVYDLTGRLALPPLTLE	2
Rv1957 (151-170)	RLALPPLTLEILSRPMPVSP	2
Rv1957 (161-180)	ILSRPMPVSPGAQWPATRGT	2

A.2.15 Rv2021c

Rv2021c	Sequence	Pool
Rv2021c (1-20)	MAMTLRDMDAVRPVNREAVD	1
Rv2021c (11-30)	VRPVNREAVDRHKARMRDEV	1
Rv2021c (21-40)	RHKARMRDEVRAFRLRELRA	1
Rv2021c (31-50)	RAFRLRELRAAQSLTQVQVA	1
Rv2021c (41-60)	AQSLTQVQVAALAHIRQSRV	1
Rv2021c (51-70)	ALAHIRQSRVSSIENGDIGS	1
Rv2021c (61-80)	SSIENGDIGSAQVNTLRKYV	1
Rv2021c (71-90)	AQVNTLRKYVSALGGELDIT	1
Rv2021c (81-101)	SALGGELDITVRLGDETFTLA	1

A.2.16 Rv2022c

Rv2022c	Sequence	Pool
Rv2022c (1-20)	VNVPWENAHGGALYCLIRGD	1
Rv2022c (11-30)	GALYCLIRGDEFSAWHRLLF	1
Rv2022c (21-40)	EFSAWHRLLFQRPGCAESVL	1
Rv2022c (31-50)	QRPGCAESVLACRHFLDGSP	1
Rv2022c (41-60)	ACRHFLDGSPVARCSYPEEY	1
Rv2022c (51-70)	VARCSYPEEYHPCVISRIAL	1
Rv2022c (61-80)	HPCVISRIALLCDSVGWTAD	1
Rv2022c (71-90)	LCDSVGWTADVERISAWLNG	1
Rv2022c (81-100)	VERISAWLNGLDRETYELVF	1
Rv2022c (91-110)	LDRETYELVFAAIEVLEEEG	1
Rv2022c (101-120)	AAIEVLEEEGPALGCPLVDT	2
Rv2022c (111-130)	PALGCPLVDTVRGSRHKNMK	2
Rv2022c (121-140)	VRGSRHKNMKELRPGSQGRS	2
Rv2022c (131-150)	ELRPGSQGRSEVRILFAFDP	2
Rv2022c (141-160)	EVRILFAFDPARQAIMLAAG	2
Rv2022c (151-170)	ARQAIMLAAGNKAGRWTQWY	2
Rv2022c (161-180)	NKAGRWTQWYDEKIKAADEM	2
Rv2022c (171-190)	DEKIKAADEMFAEHLAQFED	2
Rv2022c (181-201)	FAEHLAQFEDTKPKRRKRKKG	2

A.2.17. Rv2466c

Rv2466c	Sequence	Pool
Rv2466c (1-20)	MLEKAPQKSVADFWFDPLCP	1
Rv2466c (11-30)	ADFWFDPLCPWCWITSRWIL	1
Rv2466c (21-40)	WCWITSRWILEVAKVRDIEV	1
Rv2466c (31-50)	EVAKVRDIEVNFHVMSLAIL	1
Rv2466c (41-60)	NFHVMSLAILNENRDDLPEQ	1
Rv2466c (51-70)	NENRDDLPEQYREGMARAWG	1
Rv2466c (61-80)	YREGMARAWGPVRVAIAAEQ	1
Rv2466c (71-90)	PVRVAIAAEQAHGAKVLDPL	1
Rv2466c (81-100)	AHGAKVLDPLYTAMGNRIHN	1
Rv2466c (91-110)	YTAMGNRIHNQGNHELDEVI	1
Rv2466c (101-120)	QGNHELDEVITQSLADAGLP	2
Rv2466c (111-130)	TQSLADAGLPAELAKAATSD	2
Rv2466c (121-140)	AELAKAATSDAYDNALRKSH	2
Rv2466c (131-150)	AYDNALRKSHHAGMDAVGED	2
Rv2466c (141-160)	HAGMDAVGEDVGTPTIHVNG	2
Rv2466c (151-170)	VGTPTIHVNGVAFFGPVLSK	2
Rv2466c (161-180)	VAFFGPVLSKIPRGEEAGKL	2
Rv2466c (171-190)	IPRGEEAGKLWDASVTFASY	2
Rv2466c (181-210)	WDASVTFASYPHFFELKRTR	2
Rv2466c (191-217)	PHFFELKRTRTEPPQFD	2

A.2.18 Rv2517c

Rv2517c	Sequence	Pool
Rv2517c (1-20)	VNSAIIKIAKWAQSQQWTVE	1
Rv2517c (11-30)	WAQSQQWTVEDDASGYTRFY	1
Rv2517c (21-40)	DDASGYTRFYNPQGVYIARF	1
Rv2517c (31-50)	NPQGVYIARFPATPSNEYRR	1
Rv2517c (41-60)	PATPSNEYRRMRDLLGALKK	1
Rv2517c (51-70)	MRDLLGALKKAGLTWPPPSK	1
Rv2517c (61-80)	AGLTWPPPSKKERRAQHRKE	1
Rv2517c (71-93)	KERRAQHRKEGAQ	1

A.2.19 Rv2660c

Rv2660c	Sequence	Pool
Rv2660c (1-20)	VIAGVDQALAATGQASQRAA	1
Rv2660c (11-30)	ATGQASQRAAGASGGVTVGV	1
Rv2660c (21-40)	GASGGVTVGVGVGTEQRNLS	1
Rv2660c (31-50)	GVGTEQRNLSVVAPSQFTFS	1
Rv2660c (41-60)	VVAPSQFTFSSRSPDFVDET	1
Rv2660c (51-70)	SRSPDFVDETAGQSWCAILG	1
Rv2660c (61-75)	AGQSWCAILGLNQFH	1

A.2.20 Rv2662

Rv2662	Sequence	Pool
Rv2662 (1-20)	MDDLTRLRRELLDRFDVRDF	1
Rv2662 (11-30)	LLDRFDVRDFTDWPPASLRA	1
Rv2662 (21-40)	TDWPPASLRALIATYDPWID	1
Rv2662 (31-50)	LIATYDPWIDMTASPPQPVS	1
Rv2662 (41-60)	MTASPPQPVSPGGPRLRLVR	1
Rv2662 (51-70)	PGGPRLRLVRLTTNPSARAA	1
Rv2662 (61-80)	LTTNPSARAAPIGNGGDSSV	1
Rv2662 (71-90)	PIGNGGDSSVCAGEKQCRPP	1

A.2.21 Rv2663

Rv2663	Sequence	Pool
Rv2663 (1-20)	VEVRASARKHGINDDAMLHA	1
Rv2663 (11-30)	GINDDAMLHAYRNALRYVEL	1
Rv2663 (21-40)	YRNALRYVELEYHGEVQLLV	1
Rv2663 (31-50)	EYHGEVQLLVIGPDQTGRLL	1
Rv2663 (41-60)	IGPDQTGRLLELVIPADEPP	1
Rv2663 (51-70)	ELVIPADEPPRIIHANVLRP	1
Rv2663 (61-80)	EPPRIIHANVLRPKFYDYLR	1

A.2.22 Rv2664

Rv2664	**Sequence**	**Pool**
Rv2664 (1-20)	VKHKTDIDEWLDTIEPNPAD	1
Rv2664 (11-30)	LDTIEPNPADAHDASHLRRI	1
Rv2664 (21-40)	AHDASHLRRIIAAKEAVQTA	1
Rv2664 (31-50)	IAAKEAVQTAESELRAAVNA	1
Rv2664 (41-60)	ESELRAAVNAARAAGDTWAA	1
Rv2664 (51-70)	ARAAGDTWAAIGVALGITRQ	1
Rv2664 (61-84)	IGVALGITRQAAFQRFGPHSTASP	1

A.2.23 Rv2780

Rv2780	Sequence	Pool
Rv2780 (1-20)	MRVGIPTETKNNEFRVAITP	1
Rv2780 (11-30)	NNEFRVAITPAGVAELTRRG	1
Rv2780 (21-40)	AGVAELTRRGHEVLIQAGAG	1
Rv2780 (31-50)	HEVLIQAGAGEGSAITDADF	1
Rv2780 (41-60)	EGSAITDADFKAAGAQLVGT	1
Rv2780 (51-70)	KAAGAQLVGTADQVWADADL	1
Rv2780 (61-80)	ADQVWADADLLLKVKEPIAA	1
Rv2780 (71-90)	LLKVKEPIAAEYGRLRHGQI	1
Rv2780 (81-100)	EYGRLRHGQILFTFLHLAAS	1
Rv2780 (91-110)	LFTFLHLAASRACTDALLDS	1
Rv2780 (101-120)	RACTDALLDSGTTSIAYETV	1
Rv2780 (111-130)	GTTSIAYETVQTADGALPLL	1
Rv2780 (121-140)	QTADGALPLLAPMSEVAGRL	2
Rv2780 (131-150)	APMSEVAGRLAAQVGAYHLM	2
Rv2780 (141-160)	AAQVGAYHLMRTQGGRGVLM	2
Rv2780 (151-170)	RTQGGRGVLMGGVPGVEPAD	2
Rv2780 (161-180)	GGVPGVEPADVVVIGAGTAG	2
Rv2780 (171-190)	VVVIGAGTAGYNAARIANGM	2
Rv2780 (181-200)	YNAARIANGMGATVTVLDIN	2
Rv2780 (191-210)	GATVTVLDINIDKLRQLDAE	2
Rv2780 (201-220)	IDKLRQLDAEFCGRIHTRYS	2
Rv2780 (211-230)	FCGRIHTRYSSAYELEGAVK	2
Rv2780 (221-240)	SAYELEGAVKRADLVIGAVL	2
Rv2780 (231-250)	RADLVIGAVLVPGAKAPKLV	2
Rv2780 (241-260)	VPGAKAPKLVSNSLVAHMKP	3
Rv2780 (251-270)	SNSLVAHMKPGAVLVDIAID	3
Rv2780 (261-280)	GAVLVDIAIDQGGCFEGSRP	3
Rv2780 (271-290)	QGGCFEGSRPTTYDHPTFAV	3
Rv2780 (281-300)	TTYDHPTFAVHDTLFYCVAN	3
Rv2780 (291-310)	HDTLFYCVANMPASVPKTST	3
Rv2780 (301-320)	MPASVPKTSTYALTNATMPY	3
Rv2780 (311-330)	YALTNATMPYVLELADHGWR	3
Rv2780 (321-340)	VLELADHGWRAACRSNPALA	3
Rv2780 (331-350)	AACRSNPALAKGLSTHEGAL	3
Rv2780 (341-350)	KGLSTHEGALLSERVATDLG	3
Rv2780 (351-361)	LSERVATDLGVPFTEPASVLA	3

A.2.24. Rv2693c

Rv2693c	Sequence	Pool
Rv2693c (1-20)	VNANRTSAQRLLAQAGGVSG	1
Rv2693c (11-30)	LLAQAGGVSGLVYSSLPVVT	1
Rv2693c (21-40)	LVYSSLPVVTFVVASSAAGL	1
Rv2693c (31-50)	FVVASSAAGLLPAIGFALSM	1
Rv2693c (41-60)	LPAIGFALSMAGLILLWRLL	1
Rv2693c (51-70)	AGLILLWRLLRRESARPVVA	1
Rv2693c (61-80)	RRESARPVVAGFCGVAVCAL	1
Rv2693c (71-90)	GFCGVAVCALIAYLVGQSKG	1
Rv2693c (81-100)	IAYLVGQSKGYFLLGIWMSL	1
Rv2693c (91-110)	YFLLGIWMSLLWAVVFTLSI	1
Rv2693c (101-120)	LWAVVFTLSILIRRPIVGYL	1
Rv2693c (111-130)	LIRRPIVGYLWSWLSGRDRA	2
Rv2693c (121-140)	WSWLSGRDRAWRDVSRAVFA	2
Rv2693c (131-150)	WRDVSRAVFAFDVATLGWTL	2
Rv2693c (141-160)	FDVATLGWTLVFAARFIVQR	2
Rv2693c (151-170)	VFAARFIVQRHLYDADKTGW	2
Rv2693c (161-180)	HLYDADKTGWLGVARIGMGW	2
Rv2693c (171-190)	LGVARIGMGWPLTALAALAT	2
Rv2693c (181-200)	PLTALAALATYAAIKAAQRA	2
Rv2693c (191-210)	YAAIKAAQRAILASHDAAAV	2
Rv2693c (201-220)	ILASHDAAAVGGAAEFDADA	2
Rv2693c (211-223)	GGAAEFDADAGRE	2

A.2.25 Rv3334

Rv3334	Sequence	Pool
Rv3334 (1-20)	MKISEVAALTNTSTKTLRFY	1
Rv3334 (11-30)	NTSTKTLRFYENSGLLPPPA	1
Rv3334 (21-40)	ENSGLLPPPARTASGYRNYG	1
Rv3334 (31-50)	RTASGYRNYGPEIVDRLRFI	1
Rv3334 (41-60)	PEIVDRLRFIHRGQAAGLAL	1
Rv3334 (51-70)	HRGQAAGLALQEVRQILAIH	1
Rv3334 (61-80)	QEVRQILAIHDRGEAPCAHV	1
Rv3334 (71-90)	DRGEAPCAHVRQLLSTRIDE	2
Rv3334 (81-100)	RQLLSTRIDEVRAQIAELIA	2
Rv3334 (91-110)	VRAQIAELIALEGHLQTLLD	2
Rv3334 (101-120)	LEGHLQTLLDHASYGPPTEH	2
Rv3334 (111-130)	HASYGPPTEHDHSTVCWILE	2
Rv3334 (121-140)	DHSTVCWILESDLDEPTAIE	2
Rv3334 (131-150)	WILESDLDEPTAIEVSDIHA	2

A.2.26: Rv3406

Rv3406	Sequence	Pool
Rv3406 (1-20)	MTDLITVKKLGSRIGAQIDG	1
Rv3406 (11-30)	GSRIGAQIDGVRLGGDLDPA	1
Rv3406 (21-40)	VRLGGDLDPAAVNEIRAALL	1
Rv3406 (31-50)	AVNEIRAALLAHKVVFFRGQ	1
Rv3406 (41-60)	AHKVVFFRGQHQLDDAEQLA	1
Rv3406 (51-70)	HQLDDAEQLAFAGLLGTPIG	1
Rv3406 (61-80)	FAGLLGTPIGHPAAIALADD	1
Rv3406 (71-90)	HPAAIALADDAPIITPINSE	1
Rv3406 (81-100)	APIITPINSEFGKANRWHTD	1
Rv3406 (91-110)	FGKANRWHTDVTFAANYPAA	1
Rv3406 (101-120)	VTFAANYPAASVLRAVSLPS	2
Rv3406 (111-130)	SVLRAVSLPSYGGSTLWANT	2
Rv3406 (121-140)	YGGSTLWANTAAAYAELPEP	2
Rv3406 (131-150)	AAAYAELPEPLKCLTENLWA	2
Rv3406 (141-160)	LKCLTENLWALHTNRYDYVT	2
Rv3406 (151-170)	LHTNRYDYVTTKPLTAAQRA	2
Rv3406 (161-180)	TKPLTAAQRAFRQVFEKPDF	2
Rv3406 (171-190)	FRQVFEKPDFRTEHPVVRVH	2
Rv3406 (181-200)	RTEHPVVRVHPETGERTLLA	2
Rv3406 (191-210)	PETGERTLLAGDFVRSFVGL	2
Rv3406 (201-220)	GDFVRSFVGLDSHESRVLFE	3
Rv3406 (211-230)	DSHESRVLFEVLQRRITMPE	3
Rv3406 (221-240)	VLQRRITMPENTIRWNWAPG	3
Rv3406 (231-250)	NTIRWNWAPGDVAIWDNRAT	3
Rv3406 (241-260)	DVAIWDNRATQHRAIDDYDD	3
Rv3406 (251-270)	QHRAIDDYDDQHRLMHRVTL	3
Rv3406 (261-280)	QHRLMHRVTLMGDVPVDVYG	3
Rv3406 (271-290)	MGDVPVDVYGQASRVISGAP	3
Rv3406 (276-295)	VDVYGQASRVISGAPMEIAG	3

A.2.27 Rv1986

Rv1986	Sequence	Pool
Rv1986 (1-20)	VNSPLVVGFLACFTLIAAIG	1
Rv1986 (11-30)	ACFTLIAAIGAQNAFVLRQG	1
Rv1986 (21-40)	AQNAFVLRQGIQREHVLPVV	1
Rv1986 (31-50)	IQREHVLPVVALCTVSDIVL	1
Rv1986 (41-60)	ALCTVSDIVLIAAGIAGFGA	1
Rv1986 (51-70)	IAAGIAGFGALIGAHPRALN	1
Rv1986 (61-80)	LIGAHPRALNVVKFGGAAFL	1
Rv1986 (71-90)	VVKFGGAAFLIGYGLLAARR	1
Rv1986 (81-100)	IGYGLLAARRAWRPVALIPS	1
Rv1986 (91-110)	AWRPVALIPSGATPVRLAEV	1
Rv1986 (101-120)	GATPVRLAEVLVTCAAFTFL	2
Rv1986 (111-130)	LVTCAAFTFLNPHVYLDTVV	2
Rv1986 (121-140)	NPHVYLDTVVLLGALANEHS	2
Rv1986 (131-150)	LLGALANEHSDQRWLFGLGA	2
Rv1986 (141-160)	DQRWLFGLGAVTASAVWFAT	2
Rv1986 (151-170)	VTASAVWFATLGFGAGRLRG	2
Rv1986 (161-180)	LGFGAGRLRGLFTNPGSWRI	2
Rv1986 (171-190)	LFTNPGSWRILDGLIAVMMV	2
Rv1986 (181-199)	LDGLIAVMMVALGISLTVT	2

A.2.28 Rv2658c

Rv2658c	**Sequence**	**Pool**
Rv2658c (1-20)	MADAVKYVVMCNCDDEPGAL	1
Rv2658c (11-30)	CNCDDEPGALIIAWIDDERP	1
Rv2658c (21-40)	IIAWIDDERPAGGHIQMRSN	1
Rv2658c (31-50)	AGGHIQMRSNTRFTETQWGR	1
Rv2658c (41-60)	TRFTETQWGRHIEWKLECRA	1
Rv2658c (51-70)	HIEWKLECRACRKYAPISEM	1
Rv2658c (61-80)	CRKYAPISEMTAAAILDGFG	1
Rv2658c (71-90)	TAAAILDGFGAKLHELRTST	1
Rv2658c (81-100)	AKLHELRTSTIPDADDPSIA	1
Rv2658c (91-110)	IPDADDPSIAEARHVIPFSA	1
Rv2658c (101-120)	EARHVIPFSALCLRLSQLGG	1

A.2.29 Rv2659c

Rv2659c	Sequence	Pool
Rv2659c (1-20)	VTQTGKRQRRKFGRIRQFNS	1
Rv2659c (11-30)	KFGRIRQFNSGRWQASYTGP	1
Rv2659c (21-40)	GRWQASYTGPDGRVYIAPKT	1
Rv2659c (31-50)	DGRVYIAPKTFNAKIDAEAW	1
Rv2659c (41-60)	FNAKIDAEAWLTDRRREIDR	1
Rv2659c (51-70)	LTDRRREIDRQLWSPASGQE	1
Rv2659c (61-80)	QLWSPASGQEDRPGAPFGEY	1
Rv2659c (71-90)	DRPGAPFGEYAEGWLKQRGI	1
Rv2659c (81-100)	AEGWLKQRGIKDRTRAHYRK	1
Rv2659c (91-110)	KDRTRAHYRKLLDNHILATF	1
Rv2659c (101-120)	LLDNHILATFADTDLRDITP	1
Rv2659c (111-130)	ADTDLRDITPAAVRRWYATT	1
Rv2659c (121-140)	AAVRRWYATTAVGTPTMRAH	2
Rv2659c (131-150)	AVGTPTMRAHSYSLLRAIMQ	2
Rv2659c (141-160)	SYSLLRAIMQTALADDLIDS	2
Rv2659c (151-170)	TALADDLIDSNPCRISGAST	2
Rv2659c (161-180)	NPCRISGASTARRVHKIRPA	2
Rv2659c (171-190)	ARRVHKIRPATLDELETITK	2
Rv2659c (181-200)	TLDELETITKAMPDPYQAFV	2
Rv2659c (191-210)	AMPDPYQAFVLMAAWLAMRY	2
Rv2659c (201-220)	LMAAWLAMRYGELTELRRKD	2
Rv2659c (211-230)	GELTELRRKDIDLHGEVARV	2
Rv2659c (221-240)	IDLHGEVARVRRAVVRVGEG	2
Rv2659c (231-250)	RRAVVRVGEGFKVTTPKSDA	2
Rv2659c (241-260)	FKVTTPKSDAGVRDISIPPH	3
Rv2659c (251-270)	GVRDISIPPHLIPAIEDHLH	3
Rv2659c (261-280)	LIPAIEDHLHKHVNPGRESL	3
Rv2659c (271-290)	KHVNPGRESLLFPSVNDPNR	3
Rv2659c (281-300)	LFPSVNDPNRHLAPSALYRM	3
Rv2659c (291-310)	HLAPSALYRMFYKARKAAGR	3
Rv2659c (301-320)	FYKARKAAGRPDLRVHDLRH	3
Rv2659c (311-330)	PDLRVHDLRHSGAVLAASTG	3
Rv2659c (321-340)	SGAVLAASTGATLAELMQRL	3
Rv2659c (331-350)	ATLAELMQRLGHSTAGAALR	3
Rv2659c (341-360)	GHSTAGAALRYQHAAKGRDR	3
Rv2659c (351-370)	YQHAAKGRDREIAALLSKLA	3
Rv2659c (361-375)	EIAALLSKLAENQEM	3

Appendix B

*B.1 Hypoxic time course data (Fold induction) including rank at 7 day**

S.No	Name	Gene name	4 hours	1 day	4 day	7 day	Rank At 7 day
1	Rv0140	Rv0140	0.116	0.962	1.870	3.180	**47**
2	**Rv0188**	**Rv0188**	**0.871**	**0.883**	**1.866**	**2.638**	**76**
3	Rv0233	nrdB	1.562	2.360	2.988	2.492	**86**
4	Rv0251c	hsp	0.247	3.304	3.722	4.117	**16**
5	Rv0268c	Rv0268c	0.613	0.795	1.535	2.800	**69**
6	Rv0327c	cyp135A1	-0.230	2.048	1.905	2.924	**65**
7	Rv0350	dnaK	0.421	1.875	1.693	2.335	**96**
8	Rv0384c	clpB	0.268	2.527	2.626	2.941	**63**
9	Rv0474	Rv0474	0.712	2.254	3.966	3.390	**33**
10	Rv0520	Rv0520	0.133	1.125	1.837	3.296	**38**
11	Rv0521	Rv0521	0.232	1.292	1.521	2.317	**99**
12	Rv0754	PE_PGRS11	0.755	2.730	3.046	3.056	**54**
13	Rv0766c	cyp123	0.090	2.339	3.645	3.040	**55**
14	**Rv0767c**	**Rv0767c**	**0.172**	**2.666**	**4.720**	**4.019**	**19**
15	Rv0791c	Rv0791c	0.018	2.647	3.409	3.040	**56**
16	Rv0793	Rv0793	0.140	2.130	2.706	2.730	**72**
17	**Rv0826**	**Rv0826**	**0.345**	**1.889**	**3.697**	**3.477**	**31**
18	Rv0846c	Rv0846c	0.645	3.214	2.777	3.538	**30**
19	**Rv0847**	**lpqS**	**2.175**	**5.352**	**5.035**	**5.564**	**1**
20	Rv0848	cysK2	1.771	4.725	3.596	4.040	**17**
21	**Rv0849**	**Rv0849**	**1.280**	**3.809**	**3.439**	**2.592**	**83**
22	Rv0940c	Rv0940c	0.699	1.901	3.028	2.684	**73**
23	Rv0967	Rv0967	0.324	3.843	3.992	4.806	**5**
24	Rv0976c	Rv0976c	1.526	2.052	3.660	3.105	**51**
25	Rv0986	Rv0986	-0.715	1.305	1.045	2.526	**85**
26	**Rv0990c**	**Rv0990c**	**1.707**	**2.604**	**3.178**	**3.902**	**21**
27	**Rv0991c**	**Rv0991c**	**0.233**	**3.418**	**3.406**	**5.230**	**2**
28	Rv1169c	PE11	-0.113	2.702	1.799	3.224	**44**
29	Rv1221	sigE	-0.657	1.283	2.085	2.771	**70**
30	**Rv1284**	**Rv1284**	**0.488**	**1.854**	**3.020**	**3.107**	**50**
31	Rv1285	cysD	0.947	3.186	3.082	2.403	**92**
32	Rv1403c	Rv1403c	0.397	2.559	3.426	3.984	**20**
33	Rv1405c	Rv1405c	1.030	2.537	3.029	3.370	**34**
34	**Rv1471**	**trxB1**	**-0.172**	**2.847**	**3.101**	**4.556**	**8**
35	Rv1587c	Rv1587c	0.384	3.030	3.012	2.313	**100**
36	Rv1738	Rv1738	5.225	5.137	3.775	4.987	**4**

37	Rv1806	PE20	0.350	0.849	1.397	4.393	**9**
38	Rv1813c	Rv1813c	3.608	4.479	4.113	4.029	**18**
39	Rv1875	Rv1875	0.156	0.878	1.950	3.745	**24**
40	Rv1909c	furA	0.092	1.210	1.588	2.989	**59**
41	**Rv1954c**	**Rv1954c**	**0.514**	**1.389**	**2.868**	**2.678**	**74**
42	**Rv1955**	**Rv1955**	**0.192**	**2.506**	**4.304**	**4.383**	**10**
43	**Rv1956**	**Rv1956**	**0.355**	**2.228**	**2.726**	**3.115**	**49**
44	**Rv1957**	**Rv1957**	**-0.152**	**1.872**	**3.831**	**3.542**	**28**
45	**Rv1986**	**Rv1986**	**1.099**	**3.765**	**3.603**	**3.058**	**53**
46	Rv1990c	Rv1990c	0.664	1.686	2.434	3.151	**48**
47	Rv1994c	Rv1994c	1.210	2.839	2.219	3.220	**45**
48	Rv1995	Rv1995	2.410	4.472	2.265	2.317	**98**
49	Rv1996	Rv1996	4.857	3.237	3.956	2.945	**62**
50	Rv2007c	fdxA	3.909	3.618	2.718	2.623	**81**
51	Rv2011c	Rv2011c	0.125	1.739	1.626	2.906	**66**
52	Rv2012	Rv2012	0.491	1.887	3.218	3.317	**36**
53	**Rv2021c**	**Rv2021c**	**-0.310**	**1.152**	**2.032**	**2.483**	**87**
54	**Rv2022c**	**Rv2022c**	**-0.055**	**0.970**	**2.102**	**2.948**	**61**
55	Rv2025c	Rv2025c	0.836	2.099	3.042	2.878	**68**
56	Rv2030c	Rv2030c	5.855	3.327	2.593	2.984	**60**
57	**Rv2031c**	**hspX**	**5.335**	**4.015**	**4.948**	**4.217**	**14**
58	Rv2032	acg	4.032	3.461	1.974	3.210	**46**
59	Rv2034	Rv2034	0.681	3.744	5.049	4.997	**3**
60	Rv2035	Rv2035	0.267	3.089	3.793	3.828	**23**
61	Rv2036	Rv2036	0.730	2.593	2.043	3.310	**37**
62	Rv2050	Rv2050	0.786	2.333	2.791	3.238	**41**
63	Rv2465c	rpiB	-0.347	0.895	1.524	2.994	**57**
64	**Rv2466c**	**Rv2466c**	**0.054**	**2.933**	**3.045**	**4.153**	**15**
65	Rv2504c	scoA	0.813	1.125	1.427	2.669	**75**
66	**Rv2517c**	**Rv2517c**	**0.387**	**1.655**	**2.591**	**2.633**	**78**
67	Rv2558	Rv2558	1.504	1.167	3.097	3.636	**27**
68	Rv2617c	Rv2617c	0.065	0.398	1.358	2.482	**88**
69	Rv2623	TB31.7	3.999	3.296	3.348	3.234	**42**
70	Rv2626c	Rv2626c	3.895	4.346	4.127	4.755	**6**
71	Rv2627c	Rv2627c	2.318	2.634	2.522	3.280	**40**
72	Rv2628	Rv2628	2.312	1.853	1.693	2.427	**91**
73	Rv2642	Rv2642	0.034	2.632	4.110	4.741	**7**
74	Rv2643	arsC	0.371	2.258	2.828	3.729	**25**
75	**Rv2658c**	**Rv2658c**	**2.145**	**4.680**	**4.933**	**4.286**	**13**
76	**Rv2659c**	**Rv2659c**	**1.489**	**3.565**	**3.435**	**3.286**	**39**
77	**Rv2660c**	**Rv2660c**	**1.099**	**2.501**	**2.958**	**2.471**	**89**
78	**Rv2662**	**Rv2662**	**0.658**	**1.684**	**1.367**	**3.348**	**35**
79	**Rv2663**	**Rv2663**	**0.439**	**1.297**	**1.241**	**2.901**	**67**
80	**Rv2664**	**Rv2664**	**0.361**	**1.284**	**2.116**	**2.762**	**71**

81	**Rv2693c**	**Rv2693c**	**0.180**	**0.773**	**2.242**	**2.348**	**94**
82	Rv2694c	Rv2694c	0.149	1.591	2.496	2.624	**79**
83	Rv2699c	Rv2699c	-0.485	0.574	1.209	2.633	**77**
84	**Rv2780**	**ald**	**2.438**	**3.826**	**3.199**	**2.448**	**90**
85	Rv2913c	Rv2913c	0.117	0.965	3.290	3.540	**29**
86	Rv2962c	Rv2962c	0.742	2.577	2.407	2.334	**97**
87	Rv2963	Rv2963	1.404	4.271	3.982	3.069	**52**
88	Rv3130c	Rv3130c	4.534	3.615	3.266	3.712	**26**
89	Rv3131	Rv3131	2.908	2.750	1.953	2.368	**93**
90	Rv3269	Rv3269	2.794	2.144	3.167	2.930	**64**
91	Rv3288c	usfY	0.812	2.897	2.242	2.609	**82**
92	Rv3290c	lat	0.791	2.642	2.099	3.224	**43**
93	**Rv3334**	**Rv3334**	**0.189**	**2.628**	**2.447**	**3.399**	**32**
94	**Rv3406**	**Rv3406**	**0.248**	**0.715**	**2.846**	**4.310**	**11**
95	Rv3503c	fdxD	1.221	1.719	2.962	2.552	**84**
96	Rv3515c	fadD19	0.305	2.867	3.017	2.990	**58**
97	Rv3536c	Rv3536c	0.456	1.550	3.337	3.887	**22**
98	Rv3597c	lsr2	-0.025	0.347	2.143	2.345	**95**
99	Rv3681c	whiB4	0.994	1.392	1.807	2.623	**80**
100	Rv3862c	whiB6	0.615	2.317	3.983	4.301	**12**

**M. tuberculosis* transcript levels from bacilli exposed to 7 days of hypoxia *in vitro*. Fold Induction (FI) at 4 hours, 24 hours, 4 days and 7 days, with respective ranks: (R_{FI})

** Data obtained from (Rustad et al., 2008)

B.2 Raw intensity data normalized to SigA (abundance) *

S.No	Time(h)	Gene name	0 hour	4hour	96hs (4 days)	168 hours (7 days)	Rank at 96hr	Rank at 168hrs	Average of ranks	**Rank[$R_{(N)}$]**
1	Rv0140	Rv0140	0.61	1.21	1.26	5.10	46	13	29.5	**26**
2	**Rv0188**	**Rv0188**	**0.54**	**1.15**	**2.10**	**2.78**	**26**	**26**	**26**	**22**
3	Rv0233	nrdB	0.32	0.89	1.21	1.76	51	51	51	**50**
4	Rv0251c	hsp	0.84	1.34	11.87	13.22	3	3	3	**2**
5	Rv0268c	Rv0268c	0.51	0.34	0.83	1.39	66	59	62.5	**67**
6	Rv0327c	cyp135A1	0.01	0.00	0.14	0.18	98	100	99	**100**
7	Rv0350	dnaK	1.19	1.17	5.58	4.17	4	17	10.5	**7**
8	Rv0384c	clpB	0.32	0.58	1.25	1.58	47	54	50.5	**49**
9	Rv0474	Rv0474	0.37	0.50	4.37	4.84	7	15	11	**9**
10	Rv0520	Rv0520	0.11	0.10	0.48	0.63	83	84	83.5	**85**
11	Rv0521	Rv0521	0.18	0.21	0.46	0.53	86	89	87.5	**88**
12	Rv0754	PE_PGRS11	0.03	0.02	0.54	0.65	80	82	81	**80**
13	Rv0766c	cyp123	0.08	0.16	1.32	1.07	41	66	53.5	**57**
14	**Rv0767c**	**Rv0767c**	**0.07**	**0.03**	**3.49**	**2.63**	**10**	**28**	**19**	**18**
15	Rv0791c	Rv0791c	0.10	0.04	0.89	0.96	61	70	65.5	**71**
16	Rv0793	Rv0793	0.06	0.09	0.47	0.44	84	92	88	**89**
17	**Rv0826**	**Rv0826**	**0.34**	**0.91**	**3.43**	**3.07**	**11**	**25**	**18**	**16**
18	Rv0846c	Rv0846c	0.05	0.06	0.51	0.65	82	83	82.5	**83**
19	**Rv0847**	**lpqS**	**0.21**	**0.74**	**21.75**	**9.97**	**1**	**5**	**3**	**3**
20	Rv0848	cysK2	0.13	0.18	2.68	1.82	18	48	33	**29**
21	**Rv0849**	**Rv0849**	**0.07**	**0.11**	**0.46**	**0.46**	**87**	**91**	**89**	**90**
22	Rv0940c	Rv0940c	0.12	0.13	0.85	2.01	65	41	53	**55**
23	Rv0967	Rv0967	0.37	0.35	14.37	22.70	2	2	2	**1**
24	Rv0976c	Rv0976c	0.08	0.20	0.65	0.52	73	90	81.5	**81**
25	Rv0986	Rv0986	0.08	0.03	0.26	0.63	96	85	90.5	**92**
26	**Rv0990c**	**Rv0990c**	**0.20**	**0.60**	**1.16**	**2.16**	**52**	**39**	**45.5**	**41**
27	**Rv0991c**	**Rv0991c**	**0.29**	**0.39**	**5.16**	**6.96**	**6**	**11**	**8.5**	**5**
28	Rv1169c	PE11	0.17	0.19	0.70	2.52	70	33	51.5	**51**
29	Rv1221	sigE	1.26	0.52	3.98	26.16	8	1	4.5	**4**
30	**Rv1284**	**Rv1284**	**0.13**	**0.13**	**0.58**	**0.62**	**77**	**86**	**81.5**	**82**
31	Rv1285	cysD	0.25	0.42	1.27	1.03	44	68	56	**61**
32	Rv1403c	Rv1403c	0.05	0.09	0.92	1.34	60	62	61	**64**
33	Rv1405c	Rv1405c	0.18	0.41	1.55	1.90	34	44	39	**36**
34	**Rv1471**	**trxB1**	**0.15**	**0.08**	**1.85**	**2.18**	**31**	**38**	**34.5**	**31**
35	Rv1587c	Rv1587c	0.20	0.17	2.49	2.05	20	40	30	**27**
36	Rv1738	Rv1738	0.11	4.74	0.54	1.40	79	58	68.5	**75**
37	Rv1806	PE20	0.06	0.08	0.39	2.22	89	37	63	**68**
38	Rv1813c	Rv1813c	0.20	2.03	2.02	1.38	30	60	45	**40**
39	Rv1875	Rv1875	0.31	0.33	0.86	9.15	64	6	35	**32**
40	Rv1909c	furA	0.19	0.18	0.64	0.79	74	75	74.5	**78**

41	**Rv1954c**	**Rv1954c**	**0.02**	**0.01**	**0.42**	**0.22**	**88**	**99**	**93.5**	**94**
42	**Rv1955**	**Rv1955**	**0.14**	**0.13**	**5.36**	**3.84**	**5**	**19**	**12**	**12**
43	**Rv1956**	**Rv1956**	**0.16**	**0.12**	**2.34**	**2.53**	**22**	**32**	**27**	**23**
44	**Rv1957**	**Rv1957**	**0.26**	**0.16**	**1.04**	**1.80**	**56**	**49**	**52.5**	**53**
45	**Rv1986**	**Rv1986**	**0.16**	**0.21**	**2.10**	**1.35**	**25**	**61**	**43**	**39**
46	Rv1990c	Rv1990c	0.19	0.48	1.25	1.75	48	52	50	**48**
47	Rv1994c	Rv1994c	0.41	0.96	1.46	8.08	38	8	23	**20**
48	Rv1995	Rv1995	0.01	0.11	0.27	0.29	95	97	96	**96**
49	Rv1996	Rv1996	0.09	1.11	0.38	0.60	92	88	90	**91**
50	Rv2007c	fdxA	0.25	5.17	0.65	1.15	72	64	68	**74**
51	Rv2011c	Rv2011c	0.19	0.14	1.11	1.67	53	53	53	**56**
52	Rv2012	Rv2012	0.10	0.13	1.10	1.77	54	50	52	**52**
53	**Rv2021c**	**Rv2021c**	**0.32**	**0.16**	**1.29**	**1.17**	**42**	**63**	**52.5**	**54**
54	**Rv2022c**	**Rv2022c**	**0.22**	**0.16**	**0.74**	**0.85**	**69**	**73**	**71**	**76**
55	Rv2025c	Rv2025c	0.23	0.39	1.54	1.41	35	56	45.5	**42**
56	Rv2030c	Rv2030c	0.03	1.28	0.13	0.36	99	96	97.5	**99**
57	**Rv2031c**	**hspX**	**0.14**	**4.09**	**0.63**	**1.40**	**75**	**57**	**66**	**72**
58	Rv2032	acg	0.03	0.55	0.17	0.38	97	95	96	**97**
59	Rv2034	Rv2034	0.16	0.24	3.37	7.18	14	9	11.5	**11**
60	Rv2035	Rv2035	0.21	0.27	2.24	3.68	23	20	21.5	**19**
61	Rv2036	Rv2036	0.17	0.33	0.87	1.89	63	45	54	**58**
62	Rv2050	Rv2050	0.94	1.30	3.17	4.57	17	16	16.5	**15**
63	Rv2465c	rpiB	0.58	0.47	1.00	2.24	58	36	47	**45**
64	**Rv2466c**	**Rv2466c**	**0.35**	**0.54**	**3.29**	**10.81**	**16**	**4**	**10**	**6**
65	Rv2504c	scoA	0.22	0.28	0.75	3.29	68	24	46	**44**
66	**Rv2517c**	**Rv2517c**	**0.40**	**0.56**	**1.50**	**4.94**	**36**	**14**	**25**	**21**
67	Rv2558	Rv2558	0.29	0.72	2.03	1.82	29	47	38	**35**
68	Rv2617c	Rv2617c	0.38	0.58	1.23	3.63	50	21	35.5	**33**
69	Rv2623	TB31.7	0.27	1.42	0.58	0.99	78	69	73.5	**77**
70	Rv2626c	Rv2626c	0.16	1.55	0.88	2.30	62	35	48.5	**46**
71	Rv2627c	Rv2627c	0.06	0.22	0.12	0.40	100	94	97	**98**
72	Rv2628	Rv2628	0.24	0.92	0.52	0.81	81	74	77.5	**79**
73	Rv2642	Rv2642	0.24	0.24	3.42	7.14	12	10	11	**10**
74	Rv2643	arsC	0.08	0.09	0.60	2.38	76	34	55	**59**
75	**Rv2658c**	**Rv2658c**	**0.04**	**0.18**	**1.25**	**0.66**	**49**	**80**	**64.5**	**70**
76	**Rv2659c**	**Rv2659c**	**0.10**	**0.14**	**1.47**	**0.72**	**37**	**79**	**58**	**62**
77	**Rv2660c**	**Rv2660c**	**0.01**	**0.01**	**0.33**	**0.28**	**93**	**98**	**95.5**	**95**
78	**Rv2662**	**Rv2662**	**0.10**	**0.06**	**0.30**	**0.95**	**94**	**71**	**82.5**	**84**
79	**Rv2663**	**nrdB**	**0.21**	**0.20**	**0.76**	**2.57**	**67**	**31**	**49**	**47**
80	**Rv2664**	**Rv2664**	**0.11**	**0.11**	**0.69**	**1.56**	**71**	**55**	**63**	**69**
81	**Rv2693c**	**Rv2693c**	**0.31**	**0.24**	**1.70**	**1.96**	**32**	**43**	**37.5**	**34**
82	Rv2694c	Rv2694c	0.88	1.04	3.60	5.60	9	12	10.5	**8**
83	Rv2699c	Rv2699c	0.24	0.16	0.39	0.66	90	81	85.5	**86**
84	**Rv2780**	**ald**	**0.86**	**3.41**	**3.35**	**3.58**	**15**	**22**	**18.5**	**17**
85	Rv2913c	Rv2913c	0.14	0.10	1.03	0.78	57	76	66.5	**73**

86	Rv2962c	Rv2962c	0.09	0.13	0.38	0.42	91	93	92	**93**
87	Rv2963	Rv2963	0.11	0.23	1.40	0.74	40	78	59	**63**
88	Rv3031	Rv3130c	1.27	2.06	2.07	2.65	27	27	27	**24**
89	Rv3130c	Rv3131	0.07	3.66	0.47	0.62	85	87	86	**87**
90	Rv3269	Rv3269	0.64	3.96	1.44	0.90	39	72	55.5	**60**
91	Rv3288c	usfY	0.19	0.36	2.12	1.06	24	67	45.5	**43**
92	Rv3290c	lat	0.60	1.18	2.37	8.68	21	7	14	**13**
93	**Rv3334**	**Rv3334**	**0.21**	**0.12**	**2.51**	**2.01**	**19**	**42**	**30.5**	**28**
94	**Rv3406**	**Rv3406**	**0.11**	**0.10**	**1.56**	**1.85**	**33**	**46**	**39.5**	**37**
95	Rv3503c	fdxD	0.18	0.45	1.26	0.75	45	77	61	**65**
96	Rv3515c	fadD19	0.14	0.17	0.97	1.09	59	65	62	**66**
97	Rv3536c	Rv3536c	0.15	0.09	1.28	3.40	43	23	33	**30**
98	Rv3597c	lsr2	1.21	0.81	2.04	2.58	28	30	29	**25**
99	Rv3681c	whiB4	0.32	0.61	1.06	2.58	55	29	42	**38**
100	Rv3862c	whiB6	0.09	0.06	3.38	4.14	13	18	15.5	**14**

**M. tuberculosis* transcript levels from bacilli exposed to 7 days of hypoxia *in vitro*. Median intensity normalized to *SigA* levels (NI) at 4 hours, 24hours, 4 days and 7 days. Also included rank based on the transcript levels at day 4 and day 7; average score of ranks at day 4 and day 7; with respective ranks: (R_{NI}).

** Data obtained from (Rustad et al., 2008)

B.3 Allelic frequency in South African Xhosa Population **(du Toit et al., 1988)**

HLA Antigens	**DR1**	**DR3**	**DR4**	**DR7**	**DR8**	**DR9**	**DR11 (DR5)**	**DR13 (DR6)**	**DR15 (DR2)**
Allelic frequency	0.0839	0.4193	0.1335	0.1429	0.0124	0.0155	0.3137	0.3292	0.2174

*B.4 Properties of EHR proteins (n=26)***

	Rv0188	**Rv0767**	**Rv0826c**	**Rv0847**
Function	Conserved transmembrane proteins	Hypothetical protein	Hypothetical protein	Lipoprotein lpqS
Essentiality	Non-essential	Non-essential	Non-essential	Non-essential
Up-regulation in				
Starvation	+			
High temperature	+	+	+	
Domain		Bacterial regulatory protein, tetR family		
Gene ontology				
Gene Family				
Expression	Altered enhanced expression in Rv3676 deleted mutants			Lower level of expression in phoP (Rv0757) mutant (Walter et al., 2007)
Co-expression		Rv2034, Rv2663, Rv0841, Rv2669, Rv2662, Rv2650, Rv0840, Rv1577, Rv2660, Rv3180	Rv2011c, Rv2012, Rv2016, Rv1405c, Rv2651c, Rv0195, Rv2497c, Rv0841, Rv0998, Rv0595	Rv0848(cysK2), Rv0846c, Rv0849, Rv2963, Rv0190, Rv3406, Rv0850, Rv1533, Rv2122(hisE), Rv3405
Protein Structure		HTH (Helix-turn-helix)-type transcriptional regulator		Contains possible signal sequence and PS00013 Prokaryotic membrane lipoprotein lipid attachment site

	Rv0849	Rv0967	Rv0990c	Rv0991c
Function	Integral membrane protein, involved in the transport (drugs) across membrane	Hypothetical protein	Hypothetical protein	Serine rich protein
Essentiality	Non-essential	Non-essential	Non-essential	Non-essential
Up-regulation in				
Starvation				
High temperature				
Domain	MSF	Uncharacterised BCT, COG1937 (start 29 end 70)	SAF domain – protein domain ; 54-116 αα	
Gene ontology		1) Biological process: regulation of transcription, DND-dependent. 2) Molecular function: DNA-binding; copper ion binding Cellular component: cytoplasm	1)COG:N cell motility 2)COG: O Posttranslational modification, protein turnover , chaperones 3)Cellular process : adaptation to atypical conditions	
Gene Family				
Expression				Expression is under s^H control (Walters et al., 2007)
Co-expression	Rv0850, Rv0848, Rv0847, Rv0846c, Rv2963, Rv0190, Rv0451c, Rv0677c, Rv0676c, Rv0678	Rv0982(mprB), Rv0991c, Rv2078, Rv0987, Rv0969, Rv0356c, Rv0978c, Rv0983, Rv0979c, Rv0961	Rv1813c, Rv2008c, Rv1996, Rv2007c(fdxA), Rv0825c, Rv0252(nirB), Rv1997(ctpF), Rv2031c(hspX), Rv3366 (spoU), Rv960	Rv0987, Rv0967, Rv0982(mprB), Rv0988, Rv0983(pepD), Rv0961, Rv2078, Rv0975(fadE13), Rv0318c, Rv0972c(fadE12)
Protein Structure				
Operon	Rv0848 (cysK2) /Rv0849/Rv0850			
Additional Note	Similar to the multidrug resistance protein mdtH	Also similar to MTB proteins Rv0190 and Rv1766		

MSF Major facilitator superfamily

BCR Breakpoint cluster region

COG Clusters of Orthologous Groups of proteins

	Rv1284	**Rv1471**	**Rv1954c**	**Rv1955**
Function	Beta-carbonic anhydrase (canA)	Thioredoxin trxB1	Hypothetical protein	Hypothetical protein. Part of the TA module, Possible Toxin
Essentiality	Essential	Non-essential	Non-essential	Non-essential
Up-regulation in				
Starvation	+		+	+
High temperature		+	+	+
Domain	Pro_CA (Carbonic anhydrase) 23-162(pfam) 26-64 (bio health)	Thioredoxin – protein domain 2-103		Phage derived protein Gp49-like (DUF891) (c terminal) Pfam05973, Gp49-phage derived protein GP49-like (DUF891)
Gene ontology	1) Biological process : Carbon utilization 2) Molecular function: Carbohydrate dehydratase activity and Zinc ion binding	1) Biological process : cell redox homostasis; glycerol ether metabolism; oxidation reduction. 2) Molecular function : electron carrier acitivity; protein disulfide oxidoreductase activity. 3) Energy metablism: electron transport		
Co-expression	Rv3287c (rsbW), Rv2638, Rv1990c, Rv3746c(PE34), Rv3288c (usfY), Rv0264c, Rv1646 (PE17), Rv2034, Rv1404, Rv1910c	Rv1875, Rv2466c, Rv1874, Rv2707, Rv2699c, Rv2465c, Rv3224c(sigh), Rv3463, Rv1298(rpmE), Rv2706c	Rv1956, Rv1972, Rv3082c(virS), Rv2742, Rv1955, Rv0837c, Rv1777, Rv0779c, Rv1031(kdpC), Rv0304c (PPE5)	Rv1956, Rv1990c, Rv1989c, Rv3287c(rsbW), Rv2516c/Rv1991c/Rv2663/Rv3288c/Rv2517c/Rv2034
Operon		Rv1470trxA/1471trxB1/1472echA12/1473	Rv1954c and Rv1955	Rv1954c and Rv1955
Additional Note		Contains PS00194 Thiroredoxin family active site. Belongs to the thiredoxin family	End Overlaps next ORF upstream, Rv1955	N terminal overlaps with another ORF , Rv1954c

DUF Domain of Unknown function

	Rv1956	Rv1957	Rv2021c	Rv2022c
Function	Transcriptional regulatory protein (part of TA module)	Hypothetical protein	Transcriptional regulatory protein	Hypothetical protein
Essential	Non-Essential	Non-essential	Non-essential	Non-essential
Up-regulation in				
Starvation	+			
High temperature				
Domain	Protein domain- Helix turn helix.			
Gene ontology				
Gene Family				
Expression				
Co-expression	Rv1955, Rv1957, Rv2642, Rv1954c, Rv3123, Rv0840c(pip), Rv1990c, Rv2016, Rv2035, Rv0841	Rv1956, Rv1955, Rv0055(rpsR), Rv1954c, RvRv2959, Rv2018, Rv1981c, Rv2563, Rv3521, Rv2717c	Rv2022c, Rv1951c, Rv1960c, Rv3603c, Rv1365c(rsdA), Rv2075, Rv1015c, Rv2132, Rv3391(acrA1), Rv1535	Rv2034, Rv3180c, Rv3832c, Rv3182, Rv0750, Rv2657c, Rv2662, Rv2517c, Rv2827c, Rv1660 (pks10)
Protein Structure	Helix-turn-helix motif at αα 53-73 (+4.78 SD)		helix-turn-helix at 45-66 αα (score 1472,+4.20 SD)	
Operon	Rv1956 and Rv1957	Rv1956 and Rv1957	Rv2021c/2022c/2023c	Rv2021c/2022c/2023c
Additional Note	Molecular function: sequence specific DNA binding		Molecular function: sequence-specific DNA binding	

SD Standard deviation

	Rv2466	**Rv2517**	**Rv2660c**	**Rv2662**
Function	Hypothetical protein	Hypothetical protein	Hypothetical protein	Hypothetical protein
Essentiality	Non-essential	Non-essential	Non-essential	Non-essential
Up-regulation in				
Starvation		+	+	
High temperature	+	+	+	+
Domain	dsbA-like thioredoxin domain (pfam01323)			
Gene ontology	1) Molecular function-protein disulfide oxidoreductase activity			
Gene Family	DSBA oxidoreductase			
Expression				
Co-expression	Rv0140, Rv3054c, Rv2465c, Rv3206c(moeB1), Rv3223c(sigH), Rv1874, Rv2374, Rv3334, Rv3463, Rv2707	Rv2516c, Rv2034, Rv3182, Rv3334, Rv3188, Rv2663, Rv2827c, Rv2035, Rv1990c, Rv3917c (parB)	Rv0841, Rv2034, Rv0767c, Rv2662, Rv0122, Rv0840c (pip), Rv2827c, Rv2651c, Rv2650c, Rv2663	Rv0841,Rv2650c, Rv0840c, Rv2663, Rv2656c, Rv0767c, Rv2034, Rv2651, Rv3183, Rv2669
Protein Structure				
Operon		Rv2516c-2517c	Rv2660c-2661c	
Additional Note	Contains PS00195 glutaredoxin active site.			

	Rv2663	Rv2664	Rv2693	Rv2780c
Function	Hypothetical protein	Hypothetical protein	Conserved integral membrane alanine and leucine rich protein	Secreted l-alanine dehydrogenase ald (40kDa antigen) TB43
Essentiality	Non-essential	Non-essential	Non-essential	Non-essential
Up-regulation in				
Starvation	+			+
High temperature				
Domain				1) Alanine dehydrogenease/PNT, c terminal domain. 2) Alanine dehydrogenase/PNT, N-terminal domain
Gene ontology				1) Biological process-oxidation reduction 2) Molecular function- alanine dehydrogenase activity; binding; electron carrier activity 3) Cellular component – extracellular region 4) Functional annotation - Amino acid transport and metabolism
Co-expression	Rv2664, Rv2034, Rv0767c, Rv2650c, Rv2662, Rv0841, Rv1990c, Rv2035	Rv2663, Rv2516c, Rv2034, Rv2517c, Rv2827c, Rv3182, Rv2650c, Rv0767c, Rv2688c, Rv2662	Rv3965 (mmr), Rv1853 (ureD), Rv0401, Rv2471 (aglA), Rv1635c, Rv2209, Rv0615, Rv2291 (sseB), Rv3903c, Rv2117	Rv3287c (rsbW), Rv3290c (lat), Rv3288c (usfY), Rv0188, Rv2557, Rv0810c, Rv2659c, Rv0274, Rv0088, Rv0064
Protein Structure				
Operon	Rv2663-Rv2664	Rv2663-Rv2664	Rv2693c-2694c	
Additional Note			1) Predicted transmembrane protein- identified in culture filtrates of MTB H37Rv (Malen et al., 2008) 2) Signal peptide prediction and cleavable signal sequence confirmed experimentally. (Malen et al., 2008)	1) Secreted l-alanine dehydrogenase ald (40HDA antigen)/ TB43. 2) Contains PS00836 and PS00837 alanine dehydrogenase and pyridine nucleotide transhydrogenase signature 1 and 2 3) Epitope prediction :- FCGRIHTRYSSAYEL – B cell epitopes are predicted (Gaseitsiwe et al., 2008)

	Rv3334	Rv3406
Function	transcriptional regulatory protein	probable dioxygeneae
Essentiality	Non-essential	Non-essential
Up-regulation in		
Starvation	+	
High temperature	+	
Domain	1) MerR family regulatory protein. 2) MerR, DNA binding	
Gene ontology	1) Biological process – regulation of transcription, DNA dependent. 2) Molecular function – nucleotide binding; transcription factor activity	1) Biological process: oxidation reduction. 2) Molecular function: electron carrier activity; oxidoreductase activity, acting on a single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen
Gene Family		
Expression		
Co-expression	Rv2034, Rv2517c, Rv3188, Rv3182, Rv1990c, Rv2035, Rv2827c, Rv2516c, Rv2663, Rv1991c	Rv0474, Rv0475 (hbhA), Rv3061c (fadE22), Rv1533, Rv0068, Rv3405c, Rv0244c (fadE5), Rv0476, Rv1168c(PPE17), Rv0477
Protein Structure	Contains probable helix-turn	
Operon		Rv3406-3407-3408
Additional Note	1) DNA micro array shows lower level of expression in H37Rv during Mg2+ starvation 2) Similar to many regulatory proteins (notably mercury resistance operon regulators)	

\+ Upregulation

** Table compiled with data obtained from : TubercuList (http://genolist.pasteur.fr/TubercuList/ and http://tuberculist.epfl.ch/), TBDB (http://www.tbdb.org/), TDR Targets database (http://tdrtargets.org/), MTBreg (http://www.doe-mbi.ucla.edu/Services/MTBreg/), BiohealthBase (www.biohealthbase.org), webTB (http://www.webtb.org/), BioCye database collection (http://biocyc.org/MTBRV) and literature (http://www.ncbi.nlm.nih.gov/pubmed/)

B.5 M. tuberculosis transcript levels in bacilli exposed to 7 days hypoxia in vitro

(Rustad et.al, 2008)

		Log base2 Hypoxia/Log [(a)]				Median transcript intensity normalised to *SigA* [(b)]			
Rv#	Gene name	4hrs	1day	4days	7days	4hrs	1day	4days	7days
Rv3875	ESAT-6	-0.34	-2.17	-1.88	-1.40	5.97	4.22	3.49	4.33
Rv3874	CFP-10	0.14	-1.20	-0.56	-1.03	7.42	2.3	1.97	4.04
Rv2659c	Rv2659c	1.49	3.56	3.43	3.29	0.14	1.66	1.47	0.72
Rv2658c	Rv2658c	2.15	4.68	4.93	4.29	0.18	1.69	1.25	0.66
Rv2031c	Acr1	5.34	4.02	4.95	4.22	4.09	1.84	0.63	1.40
Rv1986	Rv1986	1.10	3.77	3.60	3.06	0.21	2.38	2.10	1.35

(a) Shows the fold induction over hypoxic time course with reference to phase of aerobic cultures.

(b) Shows the median transcript intensity normalized to *Sig*A over hypoxic time course

Appendix C

Appendix Figure C.1 Comparison of absolute IFN-g response in 7day whole blood incubation assay and QFT-GIT assay

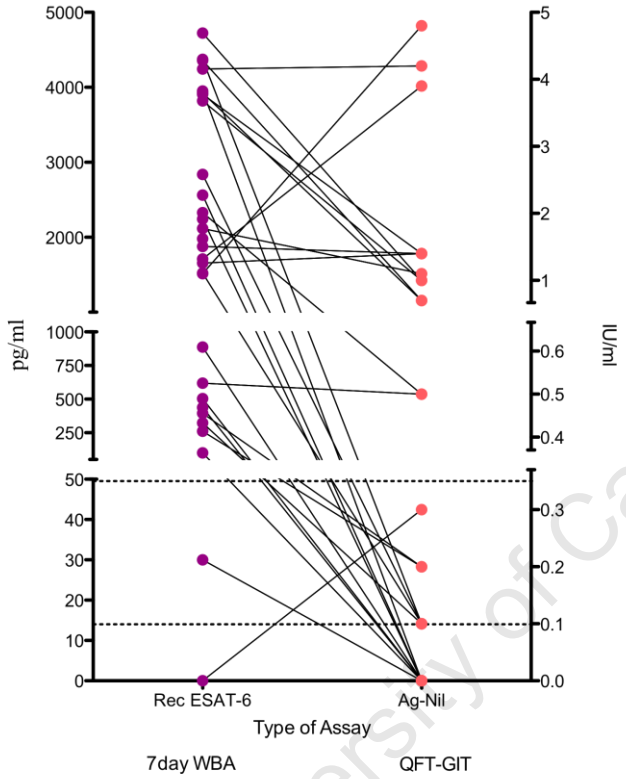


Figure shows direct comparison of the absolute values of IFN-γ in 7 day WBA using recombinant protein (Rec ESAT-6) represented in pg/ml and QFT-GIT (AG-Nil) in IU/ml. Dotted line indicate positive cut off for QFT as per manufacture's instruction.