Identification of candidate genes and testing for association with tuberculosis in humans.

Chantal Louiza Babb

Dissertation presented for the degree of Doctor of Philosophy in Medical Biochemistry at the University of Stellenbosch

Promoter: Prof. Eileen Hoal

Co-promoter: Prof. Paul van Helden

December 2007

Declaration

I, the undersigned, hereby declare that the work contained in this dissertation is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature: …….. Date: ……..

Chantal Babb
US number: 14431165

Summary

This research investigated human candidate genes for susceptibility to tuberculosis and the effect of various factors on time to sputum conversion in the admixed South African Coloured (SAC) population. Population stratification was formally tested and excluded. Population based case-control studies were the primary analysis method with a variety of genotyping methods.

Candidate polymorphisms in RANTES, CCR5, CCR2 and SDF1, were not associated with tuberculosis susceptibility. Initially the RANTES polymorphism -403 was found to be associated with tuberculosis susceptibility but after the testing of additional samples the association was lost, illustrating the challenges with association studies.

The C-type lectins DC-SIGN, encoded by the gene *CD209*, and L-SIGN are important pathogen-recognition receptors of the human innate immune response. Both lectins have been shown to interact with *Mycobacterium tuberculosis*. *CD209* promoter polymorphisms, -336 and -871, were both found to be associated with tuberculosis susceptibility. The haplotype containing *CD209* -871G and -336A was strongly associated with the control group. The *CD209* -336A allele has been found to be associated with increased DC-SIGN expression, which may be the underlying reason for an increased efficiency of host phagocytes.

Susceptibility to tuberculosis in mice has recently been attributed to the *Ipr1* gene. Eight polymorphisms in the human homologue, SP110, were investigated, including two novel polymorphisms. No significant associations were found with any of the polymorphisms investigated, including two polymorphisms that were previously found to be associated with tuberculosis susceptibility in West African populations.

A cohort of 249 cases from a longitudinal study of first time pulmonary tuberculosis patients was available. The cohort was used to investigate if the vitamin D receptor gene (VDR) polymorphisms *FokI, ApaI* and *TaqI* were associated with tuberculosis susceptibility or time to sputum conversion, and to investigate other clinical and demographic factors affecting the rate of response to treatment. No association between the VDR genotype and tuberculosis was found in the case-control study. The cohort allowed for a reliable conversion time to be determined for smear (n=220) and culture (n=222). Analysis was carried out to determine which factors, including VDR *FokI*, *ApaI*, and *TaqI* genotypes, contribute to faster mycobacterial resolution in sputum. This was done by survival curves and Cox regression models. The results indicate that the extent of disease at diagnosis was predictive of both smear and culture conversion times in the final models. Smoking status and VDR genotype contributed independently to smear conversion time, with *ApaI* 'AA' and *TaqI* 'T' containing genotypes being predictive of a faster response to tuberculosis therapy. We can conclude that the time taken for an individual to convert to sputum negativity while on DOTS therapy, can be independently predicted by the VDR genotype. This may have implications for future immunomodulatory therapies.

Identifying what contributes to susceptibility to tuberculosis will provide us with a better understanding of the human immune response to tuberculosis which may lead to the development of accurately targeted therapeutics and vaccines.

Opsomming

Kandidaatgene vir die vatbaarheid vir tuberkulose en die effek van verskeie faktore op sputum oorgangstyd was in hierdie navorsingsstudie ondersoek in die Suid-Afrikaanse Kleurlingbevolking (SAC). Dié bevolking was ook getoets vir populasie-stratifikasie, waarvan daar geen bewyse gevind is nie. Populasiegebaseerde pasiënt-kontrole studies was die primêre metode van analise en verskeie genotipering metodes was gebruik.

Polimorfismes in kandidaatgene soos RANTES, CCR5, CCR2 en SDF1 was nie met die vatbaarheid van tuberkulose geassosieer nie. Oorspronklik was daar 'n assosiasie met die RANTES -403 polimorfisme, maar met die genotipering van addisionele individue het die assosiasie verdwyn. Resultate verkry vir die polimorfisme illustreer die uitdagings waaraan assosiasie studies onderworpe is.

Die C-tipe lektiene DC-SIGN, wat gekodeer word deur *CD209*, en L-SIGN is belangrike patogeen herkenningsreseptore in die aangebore immuunreaksie. Interaksies tussen beide lektiene en *Mycobacterium tuberculosis* is voorheen gerapporteer. Die *CD209* promoter polimorfismes, -336 en -871, was met die vatbaarheid van tuberkulose geassosieer. 'n Haplotipe bestaande uit die *CD209* -871G en -336A allele was sterk met die kontrole groep geassosieer. Die *CD209* -336A alleel was geassosieer met 'n toename in die DC-SIGN proteïen vlakke, wat moontlik 'n onderliggende rede is vir die toename in die effektiwiteit van die gasheer se fagosiete.

Vatbaarheid vir tuberkulose is onlangs toegeskryf aan die *Ipr1* geen in muise. Agt polimorfismes, insluitend 2 voorheen onbekendes, was in die mens homoloog *SP110* bestudeer. Geen positiewe beduidende assosiasie was met enige van die polimorfismes gevind nie ten spyte van die feit dat twee van hierdie polimorfismes voorheen met tuberkulose vatbaarheid geassosieer was in bevolkings van Wes-Afrika.

'n Versameling van 249 TB pasiënte van 'n longitudinale studie was beskikbaar. Dié groep was gebruik om polimorfismes *Fok*I, *Apa*I and *TaqI* in die vitamien D reseptor geen (VDR) te bestudeer ten opsigte van vatbaarheid vir tuberkulose of sputum oorgangstyd sowel as ander kliniese en demografiese faktore wat die tempo van respons op behandeling kan affekteer. In hierdie studie was daar geen assosiasie gevind tussen die ontwikkeling van tuberkulose en die VDR genotipes nie. Die bepaling van 'n betroubare oorgangstyd vir beide smeer en kultuur van die groep was moontlik. Analises was uitgevoer om te bepaal watter faktore bydrae tot vinniger resolusie van *Mycobacteria* in sputum. Resultate verkry het aangedui dat die aard van die siekte tydens diagnose voorspelbaar was van die oorgangstye van beide smeer en kultuur in die finale modelle. Die rookstatus van individue sowel as die VDR genotipes het onafhanklik bygedrae tot die oorgangstyd van die smeer, met *ApaI* 'AA' en *TaqI* 'T' bevattende genotipes wat 'n vinniger reaksie op tuberkulose behandeling voorspel het. Ter opsomming, die tyd wat dit 'n individu op DOTS terapie neem om na sputum negatief oor te gaan kan onafhanklik deur die VDR genotipe voorspel word. Dit kan moontlik implikasies hê vir ander immunomodulerende terapië in die toekoms.

Die identifisering van faktore wat bydra tot die vatbaarheid van turberkulose sal ons in staat stel om 'n beter begrip te hê van die immuunrespons teen tuberkulose wat moontlik kan lei tot die ontwikkeling van akkurate behandelings en inentings.

For my parents, Wayne and Dimph

Acknowledgments

There are numerous people that I am grateful to for assisting me in finishing this, my thesis. Firstly and most importantly Prof. Eileen Hoal van Helden and Prof. Paul van Helden, for hiring someone from half way round the world, thank you for giving me the opportunity. I cannot express how I have appreciated it. Eileen for all your guidance and support over the last four years and to Paul for all your encouragement. I am extremely grateful, you have both gone beyond what was required of you as supervisors.

To everyone who helped with the Surrogate Marker database; Prof. Nulda Beyers for all her assistance, Carmen Pheiffer and her help with the database, writing queries and bouncing ideas off each other, Lize van der Merwe for explaining statistics to me, thank you for your patience, Kathy Lawrence for the Surrogate Marker workshops and her assistance with all my database questions, Gill Black and Prof. Gerhard Walzl for all their helpful input.

To everyone from, or previously at, the Department of Science and Technology/ National Research Foundation (DST/NRF) Centre of Excellence for Biomedical Tuberculosis Research including, Theresa Roos, Lizma Streicher, Rakeshnie Ramoutar, Teri Roberts, Kim Stanley, Rabia Johnson, André Loxton, Hanno Nel, Erika Truter and especially Marlo Möller, my partner in crime. To everyone I have not mentioned by name, thank you for all your support. To Elizabeth Hanekom-Keet for training me on the sequencer. To René Veikondis for running the sequencer lab, you could not have arrived at a better time. To the staff of Desmond Tutu Tuberculosis centre, the nurses and doctors involved with the care, assessment and recruitment of patients, especially Magdalene Kennedy. Last but definitely not least to the patients that volunteered to take part in these research projects.

I am grateful to the Pasteur Institute, Paris, France for giving me the fellowship within the laboratory of Lluis Quintana-Murci and Luis Barreiro with support from Prof. Brigitte Gicquel.

I acknowledge all the funders that made this research possible, including the South African Medical Research Council and DST/NRF Centre of Excellence. The Surrogate Marker cohort received funding from GlaxoSmithKline Action TB and a European and Developing Countries Clinical Trials Partnership grant to Paul van Helden.

To my housemate, for putting up with me for the last six months. To my family and friends, those near and far, thank you for all the late night calls, long letters, the welcomed visits and holidays spent together.

List of Publications

Babb C, Keet EH, van Helden PD, Hoal EG. SP110 polymorphisms are not associated with pulmonary tuberculosis in a South African population. Human Genetics 2007; 121: 521-22.

Babb C, van der Merwe L., Beyers N, Pheiffer C, Walzl G, Duncan K, van Helden P, Hoal EG. Vitamin D receptor gene polymorphisms and sputum conversion time in pulmonary tuberculosis patients. Tuberculosis 2007; 87: 295-302.

Barreiro LB, Neyrolles O, **Babb CL**, Tailleux L, Quach H, McElreavey K, Helden PD, Hoal EG, Gicquel B, Quintana-Murci L. Promoter Variation in the DC-SIGN Encoding Gene CD209 Is Associated with Tuberculosis. PLoS Medicine 2006; 3: e20.

Barreiro LB, Neyrolles O, **Babb CL**, van Helden PD, Gicquel B, Hoal EG, Quintana-Murci L. Length Variation of DC-SIGN and L-SIGN Neck-Region has no Impact on Tuberculosis Susceptibility. Human Immunology 2007; 68: 106-12.

Lombard Z, Brune AE, Hoal EG, **Babb C**, van Helden PD, Epplen JT, Bornman L. HLA class II disease associations in southern Africa. Tissue Antigens 2006; 67: 97-110.

van Helden PD, Moller M, **Babb C**, Warren R, Walzl G, Uys P, Hoal E. TB epidemiology and human genetics in Novartis Foundation Symposium No 279: Innate Immunity to Pulmonary Infection 17-41, John Wiley and Sons Ltd, Chichester 2006.

List of Abbreviations

A	adenine base
AFB	acid fast bacilli
AIC	Akaike information criterion
AIDS	Acquired Immune Deficiency Syndrome
AraLAM	Lipoarabinomannan lacking the mannose cap
ARTI	annual risk of tuberculosis infection
BCG	Bacillus Calmette-Guérin
bp	base pair
C	cytosine base
CCR	chemokine receptor
CI	confidence interval
CRD	carbohydrate recognition domain
DC	dendritic cell
DC-SIGN	dendritic cell-specific intercellular adhesion molecule (ICAM)-3 grabbing non-intergrin
DNA	Deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphates
DoH	South African Department of Health
DOTS	direct observed treatment short-course
DR	drug resistance
Freq	Frequency
G	guanine base
HCV	hepatitis C virus
HIV	human immuno-deficiency virus
HLA	human leukocyte antigens
HW	Hardy Weinberg
HWP	Hardy Weinberg equilibrium probability
IFN	interferon
IL	interleukin
Ipr1	"intracellular pathogen receptor 1" mouse locus
kb	kilo base
LAM	lipoarabinomannan
LD	linkage disequilibrium
LOD	log of the odds
L-SIGN	dendritic cell-specific intercellular adhesion molecule (ICAM)-3 grabbing non-intergrin related
LTBI	latent tuberculosis infection
MAF	minor allele frequency
ManLAM	mannosylated lipoarabinomannan
MCP-1	monocyte chemo-attractant protein-1
MDR-TB	multi-drug resistant tuberculosis

MIP-1	macrophage inflammatory protein 1
ml	millilitres
MODS	Micro Observation Direct Susceptibility
M. tuberculosis	*Mycobacterium tuberculosis*
myHsp-70	mycobacterial heat shock protein 70
n	number of samples
nc	not computable
NCBI	National Centre for Biotechnology Information
NTM	non-tuberculous mycobacteria
OR	odds ratio
PCR	polymerase chain reaction
PRR	protein recognition receptor
RANTES	regulated upon activation, normal T cell expressed and secreted
RFLP	restriction enzyme fragment length polymorphism
SA	South Africa
SAC	South African Coloured
SAND domain	Sp100, AIRE-1, NucP41/75 and DEAF-1 domain
SARS	severe acute respiratory syndrome
SDF1	Stromal Derived Factor
SM	surrogate marker
SNP	single nucleotide polymorphisms
sst1	"super susceptibility to TB1" mouse locus
T	thymine base
TB	tuberculosis
TBM	tuberculosis meningitis
TDT	transmission disequilibrium test
TLR	toll like receptors
Tm	melting temperature
TST	tuberculin skin test
UTR	untranslated region
WBC	white blood cells
WHO	World Health Organisation
wt	wild type
XDR-TB	Extensive drug resistance tuberculosis
ZN	Ziehl Neelsen
μ	level of stratification
μl	micro litre
$\chi 2$	chi square

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

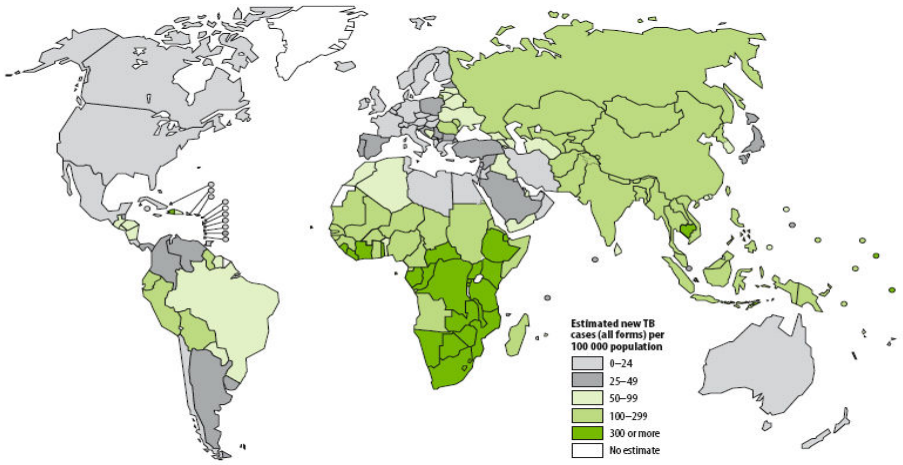
1.1 Overview of tuberculosis

1.1.1 Scale of the problem

One-third of the world's population is estimated to be infected with *Mycobacterium tuberculosis*, the etiological agent of tuberculosis [153]. However in the absence of immunosuppression, only 10% will develop disease [228], resulting in 8.8 million new cases each year and a death toll of between 2 and 3 million people per year [99, 268]. This disease tops the World Health Organisation (WHO) list of deaths caused by a single infectious agent.

With tuberculosis being such a major health problem, the WHO declared tuberculosis as a global emergency in 1993. One might imagine that more progress would have been made in eradicating this disease, since it has been known for centuries, accurately described since 1852, and a vaccine and antibiotics are available. Major progress was made with the development of chemotherapies from the mid 1900's and it was believed that it would be eradicated by the end of the last century. But this has not happened (Figure 1.1). It is also surprising that the last major development of tuberculosis treatment was over 40 years ago with the licensing of rifampicin for the treatment of tuberculosis in 1966 [274].

Figure 1.1: World tuberculosis incidence rates for 2005 from WHO Report 2007 Global Tuberculosis Control [398].

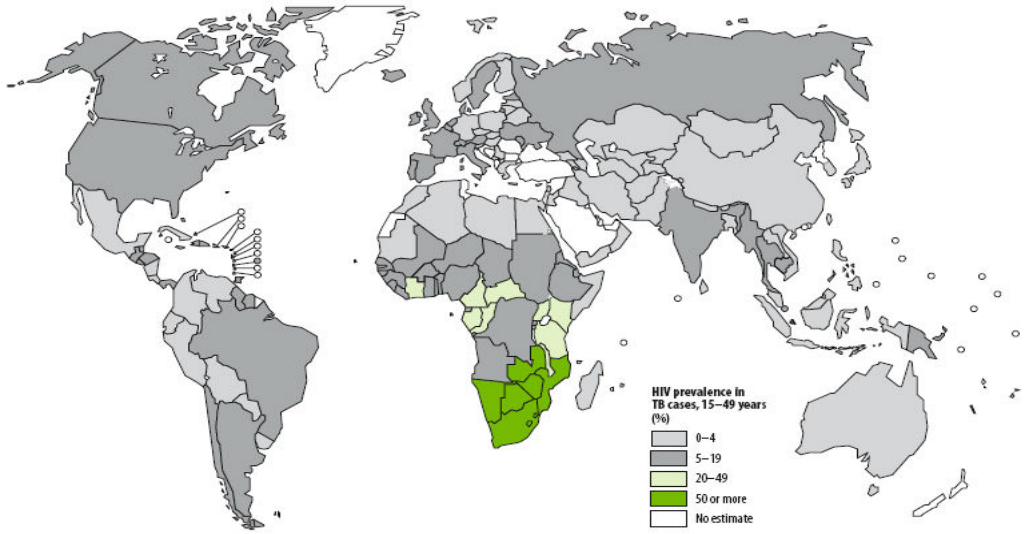


With the emergence of extensive drug resistant tuberculosis (XDR-TB) [103] there is increased urgency to develop new therapies and a need for a better understanding of this global burden. Renewed drives to increase awareness of the 're-emergence' of this burden include the Global Plan to Stop TB 2006-2015 programme [269, 323]. In addition, the World Health Assembly passed a resolution aimed at accelerating progress towards halving tuberculosis deaths and prevalence by 2015 by urging WHO Member States to develop and implement long-term plans for tuberculosis prevention and control through the full implementation of the Global Plan to Stop TB program.

The current vaccine, bacillus Calmette-Guérin (BCG), is prepared from an attenuated live bovine tuberculosis bacillus, *Mycobacterium bovis*, which was developed in the 1920's. The vaccine can be administered at birth and a single inoculation can produce long-lasting protection [36] and works well in protecting children against tuberculosis. Caution must be taken when administering the vaccination in high incidence human immuno-deficiency virus (HIV) populations as it can cause disease in HIV infected individuals when administered. Disseminated BCG can occasionally occur in non-HIV infected individuals [339]. The vaccine does not prevent infection and is of questionable benefit in many parts of the globe [22].

HIV is one of the main factors behind the increase in tuberculosis worldwide, particularly in sub-Saharan Africa. Latent tuberculosis infection (LTBI) individuals have a lifetime risk of from 5 to 8% of developing tuberculosis, whereas individuals infected with HIV have an annual risk of 8 to 10% for developing tuberculosis [302]. Over 14 million people are thought to be co-infected worldwide (Figure 1.2). In South Africa, 58% of new adult tuberculosis cases were also HIV positive according to the 2007 WHO report [398]. Despite being treatable, the combined infection has a poor outcome and is hampered by poor diagnosis and delayed reporting to clinics.

Figure 1.2: Estimated HIV prevalence in new adult TB cases 2005 as reported by the WHO [398].



Apart from HIV infection, there are a number of other risk factors for developing tuberculosis, such as sex, age, stress, immuno-compromising infections, malnutrition, previous exposure to mycobacteria, host immune defences [172], substance abuse (for example alcohol or smoking) and socio-economic status. A better understanding of the environmental risk factors has identified key factors that can be controlled in combating the spread of tuberculosis. Some lifestyle changes could reduce risk and transmission of *M. tuberculosis*, which is important in communities where tuberculosis is abundant.

A major stumbling block in the treatment of tuberculosis is that it is largely a disease of the poor. The highest burden countries are often the countries least capable of coping with the burden. It

should therefore be no surprise that of the 22 high burden countries, 9 of them are on the United Nations least developed nations list. These countries are not able to provide a steady supply of medications; the health services are often unable to cope with the case load; and regular food shortages, skills shortages and civil unrest add to the problems these countries face. Poor infrastructure usually means that new cases are either not detected or detected late. Without adequate rapid detection of new cases there is increased transmission within the community which exacerbates the problem. However this does not mean that tuberculosis will occur only in developing nations; it can be found all over the world and all populations are at risk.

1.1.2 Pathology of *M. tuberculosis* infection in humans

After contact with the pathogen, there is a complex and multistage process of interaction between host and pathogen and the true impact and importance of each step is still unclear. For example it is unclear why BCG vaccination efficacy decreases with time [91, 330] such that it works well for children but not adults.

The mycobacteria generally enter the alveolar space and thereafter the parenchyma, where the mycobacteria replicate in the alveolar macrophages or in resident lung macrophages. Macrophages are generally able to phagocytose and kill micro-organisms, but some pathogens, such as *M. tuberculosis*, have developed mechanisms to resist intracellular destruction. Dendritic cells (DC) and macrophages are the major phagocyte populations that recognise invading bacteria through innate pattern recognition receptors (PRR) [360]. It is hypothesised that the choice of receptor used to enter the macrophage influences the cellular response of the host [59]. Macrophages are the preferred habitat for *M. tuberculosis* where either a host-defence mechanism is stimulated or mycobacterial-induced inhibition of the phagosome-lysosome occurs.

The host-defence mechanism is partly stimulated through toll-like receptors (TLR) on macrophages, which then induce the production of inflammatory chemokines and cytokines, signalling infection [59, 94]. The early phase immune response of interleukin (IL)-12, from natural killer cells drives production of interferon gamma (IFNγ). IL-12 is an important component of driving the activation, differentiation and expansion of T- helper 1 (Th1) cells, which are a major source of IFNγ during the innate immune response. IFNγ prompts maturation of the phagosomes to microbicidal phagolysomes but *M. tuberculosis* can arrest the phagosomes at an early stage of maturation. This is done by preventing the fusion of the phagosomes with lysosomes [149]. This mycobacteria-induced inhibition of phagosome-lysosome fusion allows the pathogen to avoid both effective antigen presentation and elicitation of the adaptive immune response and direct anti-microbial activity of the innate immune response [163].

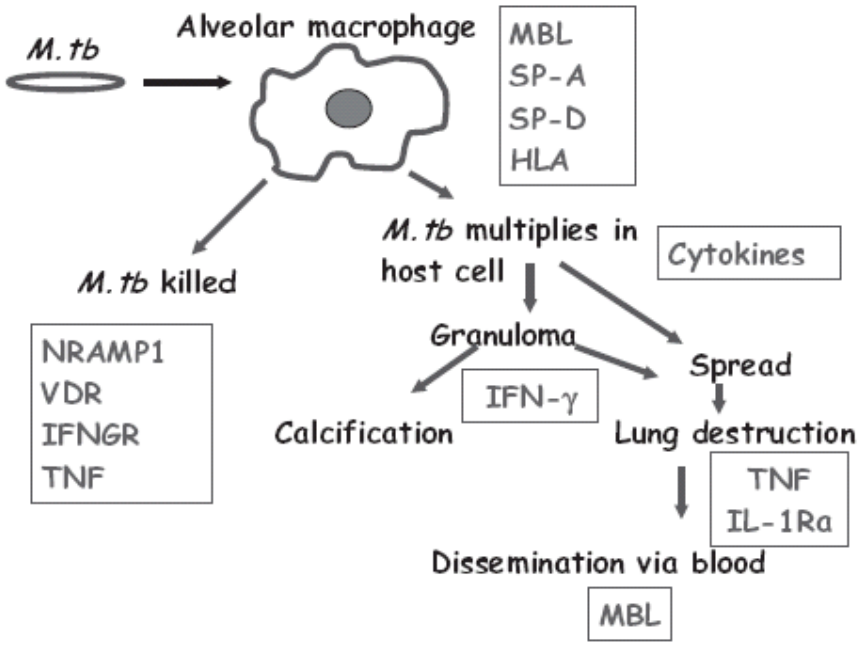
IL-12 forms a major link between the innate and adaptive immune response, since it is produced by subsets of macrophages and DCs and acts on T cells [360, 363]. IFNγ production results in the migration of monocyte-derived macrophages and resident DCs to the focal site of infection in the lungs and results in granuloma formation. Granulomas form a wall encompassing the bacilli which reside within macrophages. The granulomas function as an immune microenvironment to facilitate interactions between T cells, macrophages and cytokines along

with other host cells such as B cells, endothelial cells, fibroblasts and probably stromal cells [59, 354].

The granuloma becomes a home for *M. tuberculosis* where some are thought to avoid elimination, resulting in latent tuberculosis infection (LTBI). Extensive replication of bacteria is believed to result in loss of granuloma structure with extensive necrosis, cavity formation and finally disease. The immune response is intricate and varied. Pathways can be disrupted and interrupted at a number of steps, resulting in disease (Figure 1.1). The bacterium is able to manipulate many of these pathways and the full extent of the host-pathogen interaction is still not clear.

Within the lymph nodes, CD4 and CD8 T cells are primed against mycobacterial antigens. The priming of the CD4 T cells is an important step in controlling *M. tuberculosis* infection, as it is a pathway for processing major histocompatibility complex (MHC) II antigens, which *M. tuberculosis* can disrupt [354]. CD4 T cells are important in tuberculosis infection, which is why HIV has such a detrimental effect, as it causes the loss of CD4 cells and subsequent immunodeficiency. CD4 cells produce IFNγ, which activates macrophages to kill the intracellular bacilli [149]. In addition IL-12, IL-10 and IL-15 may be produced [354]. IL-12 is important in nitric oxide synthase 2 (NOS2) production.

Figure 1.3: A simplified representation of the tuberculosis disease process and some of the genes that may be involved at different stages [363].



1.1.3 Clinical overview of tuberculosis

The clinical features of tuberculosis will be discussed only briefly as books have been written on the subject and there are good reviews available [80, 263, 280].

The main forms of tuberculosis are pulmonary and extra pulmonary [249]. The most frequent in South Africa is pulmonary. Pulmonary tuberculosis is generally the most common form of tuberculosis with patients presenting with some or various combinations of symptoms, such as a persistent cough (>3 weeks), night sweats, loss of appetite, weight loss, shortness of breath, fever, chest pains, tiredness and / or sputum containing blood. The use of chest radiography is common in clinical practice to assist with diagnosis. Extra-pulmonary tuberculosis occurs outside of the lungs and is more difficult to diagnose [111].

Tuberculous meningitis (TBM) [75, 347] is a rare form of extrapulmonary tuberculosis, but is the most frequent form of central nervous system tuberculosis. Occurring mainly in children, it has a poor prognosis, particularly if coma occurs. TBM is difficult to diagnose as the clinical features are often non-specific. Diagnosis can be aided by testing central spinal fluid for high protein, low glucose and a raised number of lymphocytes. Frequently patients will be treated on the basis of clinical suspicion alone, as many of the cases cannot be microbiologically confirmed. When TBM is suspected, treatment should be initiated even without microbiological confirmation as there is a time lapse to receive results and any delay in treatment can result in permanent brain damage.

Disseminated BCG or BCG-osis disease occurs in children shortly after vaccination. It is rare and mainly occurs in immuno-compromised individuals, but it can occur in apparently healthy individuals. This is an extreme phenotype of mycobacterial disease and has been associated with mutations in the IFNγ / IL-12 /IL-23 pathway [193, 254], where patients are unable to produce normal levels of IFNγ. A retrospective study of 25 children from our study area in the Western Cape has recently been reviewed by Hesseling *et al* 2006 [128].

Diagnosis of LTBI can be done by tuberculin skin test (TST), which was developed in 1910, but this test may be hampered in BCG vaccinated individuals and those exposed to atypical mycobacteria [400]. The development of immune-based rapid blood tests such as the T-SPOT.TB and QuantiFERON-TB Gold will hopefully improve the diagnosis of LTBI [276]. However in developing countries there are not enough resources to provide treatment for LTBI cases and the focus is necessarily on individuals with active tuberculosis.

Generally, diagnosis of active tuberculosis will be done with a combination of a clinical examination, chest radiograph and microbiological results, as only about 60% of cases can be microbiologically confirmed. Gram positive staining of the acid fast bacilli (AFB) is done by Ziehl Neelsen (ZN) staining. Owing to its simplicity and speed it is the most widely used technique for diagnosis. Besides a microscope, no major equipment is required and work can be done on a bench not requiring any specific containment facilities. Little technical skill is required and the low running costs make it ideal for developing countries. However, sensitivity and

specificity is not particularly good, and culture based diagnosis (using Lowenstein-Jensen medium (LJ), BACTEC or Mycobacteria Growth Indicator Tube (MGIT)) is preferable.

Culture is also preferable since the isolation of live *M. tuberculosis* is usually required for drug susceptibility testing. The importance of sensitivity testing has gained a new perspective with the increasing prevalence of drug resistant (DR-), multi-drug resistant (MDR-) and extensively drug resistant tuberculosis (XDR-TB). Unless the patient is suspected of having been in contact with a DR-TB case, the only indication that a patient may have DR-TB is that they do not respond to standard treatment. Any delay in diagnosis of drug resistance will allow for transmission to other individuals, therefore perpetuating the problem. This highlights the importance of initial clinical isolates of cases being tested for drug sensitivity.

The current, optimal treatment is considered to be direct observed treatment short-course (DOTS), promoted by the WHO since 1990 [84, 98]. When correctly implemented it is considered to be the best method of treating tuberculosis and preventing resistance. Where it has been implemented correctly, with commitment from the government and health services, the treatment success rate has been 78% compared to 45% in areas which did not use the DOTS strategy [395]. For new patients, treatment consists of a combination of four medications, namely the first line drugs isoniazid, rifampicin, pyrazinamide and ethambutol, taken for at least 6 months. The initial two months of treatment is known as the intensive phase where all four antibiotics are given and this is followed by a four month continuation phase with isoniazid and rifampicin only. The intensive phase of treatment is extended if the patient has not responded within the first two months or if DR-TB is suspected.

The six months required for treatment often results in problems with compliance, which is vital for successful cure and to halt acquired DR-TB. DR-TB can be acquired through poor adherence to the DOTS regime by a patient or the TB programme. Disrupted drug supply (due to logistic problems or political unrest), poor transport networks or job requirements for example can cause interrupted treatment. The side-effects of the treatment may discourage patients from completing treatment, or they may stop taking their medication when they feel better and think they are cured. However, acquired DR might not be the major problem, as more evidence is appearing that drug resistant strains are more capable of being transmitted than previously thought [197, 373, 374]. If this is true, there is more need for rapid diagnosis of DR-TB to prevent further transmission [364], particularly when dealing with vulnerable communities. The spread of XDR-TB to a community that has high HIV rates would have, and has had, devastating effects [103].

1.1.4 *Mycobacterium tuberculosis*

M. tuberculosis is the infectious agent that usually causes tuberculosis. It is generally spread in particles that are either coughed or sneezed out by an active case. Half of those exposed to the organism become infected [286] and recent studies show that DR-TB strains are as transmissible as sensitive strains [10, 33, 374].

Strain identification may be important, as there is also evidence that various strains elicit different responses from the host [129, 133, 184, 192, 270]. Worldwide, strain type frequencies differ

between countries. It is hypothesised that this could explain why frequencies of the tuberculosis forms vary between countries and populations [192] and that the pathogen has adapted to specific populations [100, 125]. This may also be reflected in the geographical variation of the protective efficacy of the BCG vaccine [9].

1.1.5 Tuberculosis in South Africa

The revised South African national tuberculosis control programme incorporating the DOTS strategy was first established in 1996. DOTS implementation in South Africa has improved over the last decade, increasing the detection rate, but progress has been slow over the last few years. In 1998 South Africa was listed as one of 16 countries that was struggling to control tuberculosis [395] and was ranked ninth globally [397] but in 2007 was ranked seventh globally [398] as a high burden country. There are a number of factors that contribute to this, mainly the increasing rates of HIV, DR-TB, insufficient investment in personnel and infrastructure and poor quality record keeping. With the implementation of DOTS, better statistics are available, as there is improved disease surveillance. However, it is clear that the situation is not improving as desired.

The South African Department of Health (DoH) reported a tuberculosis incidence of 1037/100 000 for the Western Cape in 2005, which is the highest incidence of all South African provinces (Table 1.1). However, the HIV rate of the Western Cape was the lowest of all provinces, with a reported prevalence of 1.9% compared to a national rate of 10.8% for 2005 (which are believed to be minimum estimates). With the absence of HIV until recently [32], the epidemic in this population is historically not driven by HIV and it is possible that the population is more genetically susceptible than most to tuberculosis (Chapter 2).

Table 1.1: Tuberculosis incidence (all forms) and HIV prevalence per province as reported by the South African Department of Health for 2005 (from www.hst.org.za Accessed March 2007).

Province	**TB incidence per 100 000**	**HIV prevalence % (total population)**
Western Cape	1037.2	1.9
KwaZulu Natal	885.9	16.5
Northern Cape	846	5.4
Northwest	711.5	10.9
Free State	708.3	12.6
South Africa	645	10.8
Eastern Cape	637.9	8.9
Gauteng	487.8	10.8
Mpumalanga	333.6	15.2
Limpopo	237.2	8

The annual rate of tuberculosis infection (ARTI) is considered to be one of the most important indicators of the trend of a tuberculosis problem. Tuberculin surveys provide information on populations ARTI [13]. The estimated ARTI in the Western Cape Ravensmead and Uitsig community is 3.9% [161]. Not surprisingly, a randomized survey found that of 2401 adults analysed, 76% had a positive TST [71] and a survey of children (<15 years old) found 32% were TST positive [72].

Study site: Ravensmead and Uitsig, Western Cape

The Ravensmead, Uitsig and neighbouring communities are in close proximity to Tygerberg Hospital and have been undergoing observational tuberculosis studies since 1992, four years prior to the implementation of DOTS in South Africa. These suburbs are predominantly South African Coloured (SAC) which represents a present day homogenous population that previously received contributions from Khoisan, Malaysian, Bantu and European populations [242, 361]. The previous studies were on population health, juvenile tuberculosis, epidemic spread, DR-TB monitoring and host susceptibility. Clinical studies have shown that there is a high prevalence of childhood tuberculosis and that they frequently present late to the health services in the area [31, 195, 367]. This research has resulted in contributions to improve the diagnosis of tuberculosis and disseminated tuberculosis in children [194, 196].

The strain typing of sputum samples that have been collected from the community since 1993 provides insight into the tuberculosis epidemic in the area. There is high strain diversity in this community [382, 383], and it has been shown that transmission frequently takes place outside of the household [56, 299, 371, 372] and that patients can be infected with multiple strains of *M. tuberculosis* at once [384]. Another important finding is that a large proportion of patients that require retreatments after cure, are reinfected and are not reactivation cases [368, 370]. Individuals who have had a previous episode of disease have a four-fold increased risk of developing another tuberculosis episode compared to those who have never had active tuberculosis disease [370]. The investigations into strain diversity and variability have lead to greater understanding of the dynamics of tuberculosis within this community and can be translated to other high tuberculosis prevalence settings.

Some research into genetic susceptibility to tuberculosis using population case-control and TDT studies has also been done in this community and revealed that polymorphisms in the genes IFNγ [284] and solute carrier family 11A member 1 (*SLC11A1*, formerly known as *NRAMP1*, natural resistance-associated macrophage protein 1) [132] are associated with susceptibility to tuberculosis. The *IFNγ* +874T allele, which was overrepresented in controls, is preferentially bound by transcription factor NFKβ, and is associated with raised IFNγ expression *in vitro* [284]. A case-control using 313 cases and 235 controls found a significant association (p=0.017) between the *IFNγ* +874T allele and protection against tuberculosis, which was then confirmed in a TDT using 131 families from the same population (p=0.005) [284].

Polymorphisms in the gene *SLC11A1,* were also found to be associated with tuberculosis susceptibility in a SAC case-control analysis. A 5'$(GT)_9$ allele in the promoter had a significantly higher frequency in controls versus cases (p=0.002). Homozygotes for a 3'UTR TGTG deletion

(1729+55del4), were over-represented in the cases compared to the controls (p=0.013). After stepwise logistic regression, the 5' and 3' polymorphisms were found to contribute separate main effects. *SLC11A1* is the human homologue of the mouse *Nramp1* gene which will be discussed in Section 1.3.

A number of polymorphisms that were associated with other diseases were hypothesised to be candidates in tuberculosis susceptibility, namely butyrophilin-like 2 gene (*BTNL2*) [218] and caspase recruitment domain-containing protein 15 gene (*CARD15*) [219]. *BTNL2* was hypothesised to have a possible function as a T cell co-stimulatory molecule [358]. *BTNL2* has previously been associated with sarcoidosis, a disease of unknown etiology but presents with coughing and granulomas, rather similar to tuberculosis [291]. *CARD15* is the gene encoding nucleotide-binding oligomerisation domain 2 protein (NOD2) which has been recognised as a non-redundant recognition mechanism of *M. tuberculosis* and polymorphisms in *CARD15* were associated with Crohns disease, also a granulomatous disorder. After large case-control studies, it was concluded that the polymorphisms in both *BTNL2* and *CARD15* were not associated with tuberculosis susceptibility in our population.

The frequency of polymorphisms in the genes glutathione S-transferase (*GST*) and N-acetyltransferase 2 (*NAT2*) were determined in various population groups from SA, including the SAC. These enzymes metabolise many environmental and chemotherapeutic agents, and polymorphisms in these enzymes may contribute to different disease profiles or responses to toxic or chemotherapeutic agents. The *NAT2* 'fast' allele, which has been found to have a significant effect on acetylation, was present at a high frequency in this population [6]. It is hypothesised that the efficacy of antituberculosis chemotherapy may vary between various ethnic groups depending on their *NAT2* status. In other words, the acetylation of the isoniazid depends on *NAT2* status, therefore because of the *NAT2* allele frequency, isoniazid efficacy will vary between populations. This could have implications for drug development and trials. The *NAT2* status in our population and the effect on tuberculosis therapy is being investigated further and is not part of this thesis.

1.2 Host Genetics and Tuberculosis

1.2.1 Genetic susceptibility to tuberculosis in human populations

Prior to the discovery of the pathological agent *M. tuberculosis* by Koch in 1882, many believed that tuberculosis was actually an inherited disorder. As a result of it being thereafter attributed to an infectious agent, there was little focus on the host. The known environmental risk factors such as age, immuno-compromising diseases (diabetes or HIV), or substance abuse, do not explain why only half of the individuals that are exposed to the organism will become infected, and why only 10% will ever develop disease. A complex interplay between environment, pathogen and host can account for the development of active disease. The importance of the genetic components and heritability of tuberculosis susceptibility is provided by twin [61, 143, 147, 308] and adoptee [322] studies. There is a large body of evidence pointing to the major role of genetic factors in the human response to a number of infectious pathogens [5, 26, 34, 62, 131, 233], and

these genes could impact on treatment and vaccine efficacy [234]. There is also increasing evidence that host genetic factors determine differences in host susceptibility to mycobacterial infection and might contribute therefore to the pattern of clinical disease [26, 34, 49, 62]. The common disease-common variant (CDCV) hypothesis is believed to apply to tuberculosis. The CDCV hypothesis suggests that genetic susceptibility to a common disorder will be influenced by relatively common polymorphisms [46, 273].

1.2.2 Evolution and infection

It has been claimed that during the industrial revolution in Western Europe one-quarter to one-fifth of all deaths occurred as a result of *M. tuberculosis* infection. This makes it very likely that *M. tuberculosis* has had an effect on the evolution of the human genome. Not only *M. tuberculosis* infection would have played a role, but also other diseases such as *Plasmodium falciparum* (malaria), the bubonic plague and smallpox. Haldane, one of the pioneers in recognising the role pathogens could have on the evolution of the human genome, suggested that micro-organisms have been the main agents of natural selection for the past 5000 years [119]. The spread of advantageous alleles is fundamental to evolutionary processes and includes the spatial dynamics of co-evolution between pathogens and their hosts [241]. Diseases that exert the largest death toll will produce the most pronounced evolutionary effect [222]. Time of exposure to the disease will then play an important role in the effect and an example of this occurring would be in American Indian reservations [353]. There were a large number of deaths attributed to tuberculosis after introduction of the disease to the population, but after a generation or two the death rates due to tuberculosis decreased drastically. This is presumably due to strong selective pressure against tuberculosis-susceptible genotypes. For example, in the Qu'Appelle Indian reservation there was a reported tuberculosis incidence rate of 9000 per 100 000 in 1890, which decreased to 800 per 100 000 by 1926, and further decreased to 273 per 100 000 in 1931 [88, 186]. Within the reservation, the death rate due to tuberculosis was double that of surrounding white communities in 1931 [88].

1.3 Strategies to identify candidate-genes for susceptibility to tuberculosis

The human genome project has revealed extensive variation between populations and individuals. Identifying which variations are associated with susceptibility to a complex disorder requires selection of appropriate candidate genes to investigate [256]. The selection can be done through a number of different approaches and these have included genome wide studies, mouse models and human susceptibility models.

Genome-wide studies are systemic and comprehensive and are capable of identifying genes that exert a major effect on population-wide disease susceptibility. Data from genome-wide scans can be used to narrow down the number of candidate genes. With genome-wide scans there is usually a large number of genes within the identified regions, and determining which could be important in tuberculosis infection can be done by hypothesis. This method has been fairly successful in determining disease-susceptibility genes for leprosy susceptibility [213, 307, 352]. Bellamy *et al* conducted a genome-wide scan for tuberculosis susceptibility and found two regions, 15q and

Xq, that showed suggestive evidence of linkage to tuberculosis [27]. They found that no single marker was strongly linked as a susceptibility gene among the African populations [27]. A second genome wide scan was conducted in a Brazilian population and found suggestive evidence for regions 6p21.32, 17q22 and 20p13 [211]. A more recent genome-wide scan in 96 Moroccan families found the 8q12-13q region to be significantly linked to tuberculosis [17]. Currently, linkage disequilibrium (LD) mapping of the 8q region is being conducted to identify tuberculosis-susceptibility genes. In a genome-wide scan for leprosy susceptibility, chromosome region 6q25 was identified [213], and linkage disequilibrium mapping found that polymorphisms in the *PARK2/PACRG* gene conferred susceptibility to leprosy [212].

For both of the earlier genome-wide scans done in humans, further fine-mapping has taken place. In the 15q11-13 region, fine mapping results suggested that *UBE3A* or a closely flanking gene may be a tuberculosis-susceptibility locus in African families [50]. After a segregation analysis was done in a Brazilian cohort, it was determined that only one or two loci may determine individual tuberculosis susceptibility [304].

Genome-wide scans of tuberculosis susceptibility have also been conducted in mouse models and resulted in the identification of regions on mouse chromosomes 1 (*Nramp1*, *sst1* (super susceptibility to TB 1) and *Trl-1*), 3 (*tbs1* and *Trl-2*), 7 (*Trl3*), 19 (*Trl-4*) and 9 (*tbs2*) [159, 166, 214, 215, 375]. Fine mapping of these regions could lead to the identification of particular genes, whose human homologues can then be identified for further study.

The mouse *Nramp1* region has been successfully transferred to human association studies where the human *NRAMP1*, renamed as *SLC11A1*, has been found to be associated with susceptibility to tuberculosis in a variety of populations [2, 14, 29, 79, 104, 132, 181, 187, 262, 292, 317, 342, 401]. A recent meta-analysis showed statistically significant associations between the *NRAMP1* variants and tuberculosis susceptibility in Asian and African subjects but not in Europeans [170].

The mouse chromosome 11 has synteny with human chromosome 17q11.1-q12, which contains a number of potential candidate genes involved in the immune response. These include regulated upon activation, normal T cell expressed and secreted (*RANTES*), *NOS2A* which encodes inducible nitric oxide synthase (iNOS), genes encoding members of the family of small inducible chemokines and genes for signal transducers and activators of transcription [141]. Fine-mapping has identified region 17q11.2 as containing potential candidate genes that could account for the variable disease and pathology associated with tuberculosis and leprosy.

Fine mapping of the *sst1* locus on the mouse chromosome has recently identified *Ipr1* in the locus as the mouse susceptibility gene to *M. tuberculosis* [159, 250]. The human homologue is *SP110*, whose function is not clear. It is believed to be involved in regulation of transcription by facilitating DNA binding, using the conserved SAND (Sp100, AIRE-1, NucP41/75 and DEAF-1) domain. The *sst1* phenotype differs from *Nramp1* and reflects a true murine susceptibility gene as it causes lung-specific effects in tuberculosis susceptibility.

A new approach that is coming into its own is genome–wide association (GWA) studies. GWA offers the potential of identifying causal genes with modest effects, which would probably be the

case in tuberculosis susceptibility. GWA has already been successfully used with Crohn's disease [124] and Inflammatory Bowel disease [82]. However, there are still a number of factors that need to be considered when contemplating a GWA study, such as marker set, LD and statistical analysis methods that consider multiple testing [130, 380]. In the mean time rigorously designed studies are required to determine if this type of approach is feasible.

Hypothesis driven candidate gene selection can be effective in identifying susceptibility to tuberculosis genes. It requires an understanding of the pathology of the disease to identify which genes could be important in the control of *M. tuberculosis* infection. As discussed earlier (section 1.1.2), the response to *M. tuberculosis* infection is complex. There are a number of pathways involved and why there is progression to disease in only 10% of the infected individuals is still not clear. Identification of proteins and molecules that are important in the various responses to tuberculosis has already been successful done with, for example *SLC11A1*.

Candidate genes chosen by hypothesis

The gene encoding RANTES is located on chromosome 17 in the chemokine cluster, a region which has been shown to contribute to susceptibility to tuberculosis and leprosy in Brazilians [141] and has been associated with tuberculosis in a number of instances in the last few years [294, 313, 388]. This chemokine plays a significant role in inducing activation and proliferation of macrophages and T cells as well as lymphocytes and monocytes, which are essential for granuloma formation [294].

Chemokine receptor 5 (*CCR5*), Chemokine receptor 2 (*CCR2*) and Stromal derived factor (*SDF1*) polymorphisms have been found to have an association with HIV infection and disease progression, although the effect of the polymorphisms on HIV disease varies between populations [114]. As chemokines play a critical role in the immune and inflammatory response to infectious diseases, these polymorphisms might be related to susceptibility to tuberculosis [63]. The effects and frequencies of these polymorphisms vary in different populations, particularly the frequencies of *CCR5* polymorphisms [179, 311].

Dendritic Cell Specific Intercellular adhesion molecule (ICAM)-Grabbing Nonintegrin (DC-SIGN) is known to be a major receptor for *M. tuberculosis* and other pathogens on human DCs [365]. The recognition of pathogens by DCs is crucial in mounting the immune response. DCs bind strongly to mycobacteria such as *M. tuberculosis* and *M. bovis* BCG through the mannose capped cell wall component (ManLAM) of the pathogen, but does not bind to LAM, which lacks the mannose cap (AraLAM). ManLAM binding to DC-SIGN induces production of the anti-inflammatory cytokine IL-10 [157, 338, 366]. Immature DCs bind to *M. tuberculosis* and *M. tuberculosis* BCG and this inhibits maturation of the DC. These factors both contribute to *M. tuberculosis* virulence. ManLAM is abundant in slow growing virulent mycobacteria, such as *M. tuberculosis* and *M. leprae*, whereas AraLAM is abundant in fast growing atypical, avirulent mycobacteria, such as *M. smegmatis*. It is suggested that *M. tuberculosis* targets DC-SIGN both to infect DCs and to down-regulate DC mediated immune responses. Investigating polymorphisms and haplotypes in our population could show the importance of this gene in tuberculosis susceptibility.

Vitamin D receptor (VDR) is synthesised in monocytes and activated T and B lymphocytes. Its ligand, the active metabolite of Vitamin D, calcitriol, is produced in the kidney and by activated monocytes and macrophages, particularly in granulomas. Through its interaction with Vitamin D, the retinoid X receptor (RXR) and the Vitamin D response element (VDRE), VDR exerts several immunomodulatory effects. These include the activation of monocytes and cell mediated immunity, modulation of the Th1-Th2 host immune response, suppression of lymphocyte proliferation and restriction of *M. tuberculosis* survival in macrophages [37, 70]. Various *VDR* polymorphisms and their involvement in tuberculosis susceptibility have been investigated with varying results [28, 37, 180, 285, 391]. A study in West Africa on four *VDR* polymorphisms found no association in a case control analysis but an association was found with a particular haplotype in the transmission disequilibrium test (TDT) [37]. Evidence of the subtlety of the effect of *VDR* polymorphisms in the immune response was found by Roth *et al* [285], who detected an association with time to sputum conversion in tuberculosis patients after diagnosis.

1.4 Genetic association studies

There are two study designs that are used in genetic association studies [168] namely, family-based [165] or population-based studies using unrelated individuals [18]. Currently the most widely used family-based method is TDT studies and for population studies the case-control methodology.

Family-based TDT studies are robust against population substructure and significant findings always imply both association and linkage [165]. The TDT study design was originally developed to test for linkage in the presence of association but now, as it is robust to population stratification, the most common usage is testing for association in the presence of linkage [165, 300]. However population-based studies are regaining popularity as there are several methods that can compensate for population stratification [74, 261]. In addition case-control studies require a smaller sample size compared to family-based studies which requires parents and/or siblings also to be sampled.

Case-control analyses were the primary testing methodology applied in this thesis. There are a number of important factors that need to be considered for a successful case-control study [18, 58, 168]. These will mainly impact the power of a study and therefore its validity. Case-control studies compare the allele frequency of a polymorphism between two well defined groups; cases with the disease and unrelated healthy controls. An increased frequency of the polymorphism in cases indicates that the presence of the polymorphism may increase risk of disease. Understandably, cases and controls should be from the same ethnic group to control for the genetic background of the groups. When the population under investigation is large and there is random mating, control genotypes should be in Hardy Weinberg equilibrium (HWE) (Appendix A1.6.1). Reasons for a genotype not being in HWE include: random chance, miscalled genotyping, and a heterogeneous population [168, 396].

Classification of the phenotype must be done correctly, but clinical heterogeneity is important only if a genetic variant influences a disease subset [256]. For example prior to collection, it is

vital to define which form of tuberculosis and which controls will be assessed. In tuberculosis, controls could be latently infected, i.e. TST positive, or negative and presumably not infected. As discussed in section 1.1.6 our population has a high exposure rate, therefore our controls are likely to almost all be LTBI [71, 72]. This would be sufficient for our analysis as we are investigating progression to tuberculosis disease and not *M. tuberculosis* infection status. However, we are aware that a number of our controls may develop tuberculosis after the sample was collected, in particular with younger participants. We recognise this, but it has been shown that with a slight misclassification of controls the overall impact will be small and can be compensated for by increasing the sample size [58].

Analysis in case-control studies is generally done using 3x2 contingency tables, where the genotypes are counted in columns of cases and controls and assessed by a Chi square (χ^2) test. This method has reasonable power to detect underlying risk. This is the general genetic model, but dominant, recessive and multiplicative models can be assessed in a similar manner [168]. Fisher's exact analysis on alleles and dominant/recessive models can be conducted and can estimate the disease risk (odds ratio) conferred by the polymorphism [18, 168]. When there is biological support, more complex analysis, using Cochran-Armitage test, can be used for additive models [18, 168] and does not rely on an assumption of HWE.

Linkage disequilibrium (LD) refers to the non-random association between two loci. Typically a number of polymorphisms spanning a gene will be independently tested, but when there is LD this can be redundant. LD is defined as the non-random pattern of association between alleles at different loci in a population [225, 240]. If LD is high between two alleles, they are considered to be linked and can sometimes be used as a substitute for the state of nearby loci [46]. A polymorphism associated with disease susceptibility might be directly associated, but if in complete LD with the causative polymorphism, the polymorphisms genotyped is then said to be indirectly associated with disease susceptibility [256].

There are two important measures of LD; D' and r^2 [18, 386]. In the context of association studies, r^2 (the correlation coefficient), is more useful as it reflects the statistical power to detect LD [18]. r^2 is inversely proportional to the sample size required to detect statistically significant LD and also relies on allele frequency [386]. D' is typically higher than r^2, where a D' of one indicates complete LD between the polymorphisms. One of the reasons it is not useful in association studies, is that even when an allele is very rare, D' can be large, which is usually of little practical interest. D' is also sensitive to a few recombination's between the loci [18].

When there is complete LD between single nucleotide polymorphisms (SNPs) and one SNP is capable of providing information on the other polymorphism, then these SNPs are considered to be tagging or tagSNPs [122]. The polymorphism, for which the tagSNP provides complete or additional information, is called the tagged SNP. Identifying the tagSNPs spanning a gene can increase genotyping efficacy, reducing costs.

Determination of LD over the genome has recognised regions that are in high or low LD. Regions of high LD and low haplotype diversity are defined as Haplotype blocks [46]. Haplotyping tagging, which involves selection of the minimum subsets of SNPs which represent

the haplotypes inferred from the original set [122] within a block, can also increase genotyping efficacy.

LD and haplotype diversity has been found to vary between populations. In particular African populations have smaller blocks of LD [271] and there is high haplotype diversity within the African compared to European populations. Admixed populations have unique LD patterns, such as long range LD which can be an advantage in association studies [4, 393]. Using the long range haplotypes in admixed populations will provide more power to detect genes that may be associated with disease susceptibility [315].

There are a number of online databases such as HapMap, that provide information on LD and tagSNPs. The information from the reference populations has been found to be transferable to different populations [116]. However, this should be used cautiously in admixed populations, such as the SAC, which have numerous ancestral populations [393]. It is therefore the best strategy to use information from all available reference populations is used.

Stratification

Although our cohort is considered a present-day homogeneous community that has received genetic contribution from different populations multiple generations in the past [242, 361], population stratification between cases and controls can be a confounding factor leading to a spurious positive association. However, the use of admixed populations in association mapping studies can be very useful in identifying disease-causing genetic variants that differ in frequency across parental populations. Nevertheless, when the admixture event is too recent, allelic frequencies can differ coincidentally among cases and controls, reflecting a non-uniform genetic contribution from the parental populations to each subpopulation (i.e., cases and controls), rather than a genuine association between a given genetic variant and the phenotype under study. In this case, the study-cohort is said to present population stratification. To formally test and quantify the levels of background genetic differences [260], if any, between cases and controls, genomic control correction can be done by genotyping a panel of independent SNP markers. Currently there is no agreement as to the number of random SNPs that should be genotyped to determine the level of stratification within a given population. Genomic control can be used to assess the genetic background, and can correct cryptic relatedness between cases and controls [74, 376]. The mean χ^2 statistic among the random SNPs is compared for allele frequencies between cases and controls, which represents the levels of stratification (μ) between the two groups [272].

1.5 Aims of the study

Broadly speaking, this study investigated candidate genes for susceptibility to tuberculosis infection and progression to disease in the SAC population of the Western Cape, South Africa. Population based case-control studies were the primary analysis method, and this is expected to detect genes that exert a low to moderate effect on susceptibility to tuberculosis disease. Tuberculosis cases were defined as individuals with past or current active tuberculosis disease and

controls were individuals with no previous or present clinical presentation of disease. Controls would more than likely be infected with *M. tuberculosis* because of the high ARTI in this community [71, 72, 161].

Preliminary population screens determined the genotype frequency and the minimum sample size required for sufficient statistical power to detect an association. For genotypes that were at too low a frequency, the OR detectable was greatly increased, requiring more samples to be typed to detect a smaller OR. Provided the frequency of the genotype was common in the general population, the study had enough power to detect a genotype that exerts a smaller effect. Inferred haplotype based association studies were done when a number of polymorphisms over a gene were genotyped.

A specific aim was to study the frequencies and potential association with tuberculosis of candidate polymorphisms in genes *RANTES*, *CCR5*, *CCR2* and *SDF1* using a case-control methodology.

A further aim was to investigate the Dendritic cell-specific intercellular adhesion molecule-3 grabbing nonintegrin (*DC-SIGN*/ encoded by gene *CD209*), in collaboration with a group at the Pasteur Institute Paris, France. Resequencing of *CD209* was done to identify polymorphisms in the SAC population to determine whether there were any unique SNPs or haplotypes, and determine their frequencies. Subsequent genotyping analysis was done on the Victor (Perkin Elmer) using a case-control cohort to determine if there was an association with tuberculosis. This cohort was used to assess possible population stratification in our admixed population, using genomic control. A length variation in *L-SIGN*, a close homologue to DC-SIGN, was also investigated.

In addition to testing candidate genes that were chosen by hypothesis, a candidate gene selected from mouse studies was chosen, namely, the human *SP110* gene, which is the homologue of the mouse intracellular pathogen resistance 1 (*Ipr1*) gene which is within the *sst1* locus.

Finally we aimed to investigate the *VDR* gene, which may be linked to susceptibility or progression to disease. This analysis was done with a specific cohort of patients, namely the Surrogate Marker (SM) cohort. The SM study was a longitudinal study of first time pulmonary tuberculosis patients who were evaluated during and after treatment. They were used in a case-control analysis to identify the effect of the *VDR* polymorphisms on phenotype, in particular, time to sputum conversion after diagnosis of tuberculosis. This required more intricate statistical analysis compared to the case-control study and a specialised statistician was required.

Identifying genes that exert some effect on host susceptibility to tuberculosis will give a greater understanding of disease pathogenesis. Clearly the immune response is important in the host reaction to *M. tuberculosis* infection and the genes expressed in the host lungs may influence the host's ability to contain infection and resist disease. The environmental factors play a significant role but are not the only determinants in developing disease, where the interplay of these factors with the host is vital. The different allele frequencies between cases and controls may provide

important information about critical responses of the host to *M. tuberculosis*. It is hoped that our findings will eventually contribute to the development of novel therapeutics and vaccines.

1.6 Summary

The study population, SAC, is an admixed population that has contributions from Africa, Asia and Europe, discussed in Chapter 2. Gene polymorphism frequencies will differ in our unique admixed population compared to previously investigated populations. The use of an admixed population in genetic association studies is believed to be advantageous, but does have a number of factors that need be taken into consideration, such as possible population stratification. It is important to have some knowledge of the background of the population to determine its heterogeneity. By means of extensive literature searches, a table of the HLA class II allele frequencies in various South African populations was constructed (Table 2.1). I also contributed to the writing of the articles from this chapter. The case-control samples all came from thhe SAC community and only Chapter 6 had a unique set of case samples, from them the same population group and area.

Candidate genes, *CCR5*, *CCR2*, *SDF1* and *RANTES*, which were chosen by hypothesis, were investigated. Known candidate polymorphisms that had been associated with other diseases were chosen and the results from the SAC population are reported in Chapter 3. I did the genotyping for the genes *CCR2*, *CCR5* and *SDF1* and the statistical analyses for all the polymorphisms tested.

Functional studies had identified DC-SIGN as a major *M. tuberculosis* receptor on human dendritic cells [338] and it was therefore considered a good candidate gene. The genomic region of *CD209* was sequenced in SAC samples to identify informative markers and avoid ascertainment bias. This resulted in the identification of eight polymorphisms that were genotyped in a case-control study. The work was extended by looking at a length variation in the genes *CD209* and *L-SIGN*, which is a close homologue of *CD209*. Results are reported and discussed in Chapter 4. I prepared all the samples that were subsequently genotyped, did the majority of the *CD209* polymorphism genotyping and contributed to the writing of the article.

SP110 was identified as a candidate gene for susceptibility to tuberculosis through mouse studies. The mouse homologue, *Ipr1*, was located in the *sst1* locus, and mediates innate immunity to tuberculosis in mice [250]. As with *CD209*, the *SP110* gene was resequenced to identify polymorphisms. This allowed for identification of polymorphisms that were novel and the determination of LD. Results are reported in Chapter 5.

Finally a case-control analysis was done to test for an association with tuberculosis and polymorphisms in the *VDR* gene. There are many conflicting studies on the association of *VDR* polymorphisms and tuberculosis, and its contribution to tuberculosis susceptibility is highly debated. A unique sample set of cases that were all first time pulmonary tuberculosis patients were used in this chapter. Apart from the case-control study, a more complex study was conducted on this well defined cohort of patients. This was done to investigate whether our

population would confirm a Peruvian study which found that *VDR* polymorphisms were associated with time to sputum conversion in tuberculosis patients [285]. A cohort of first time pulmonary tuberculosis patients from our SAC population, which had sputum tested at various time points, allowed for a similar analysis and results are described in Chapter 6. A large number of variables, including a reliably estimated time to sputum conversion for both smear and culture, were used to assess the contribution of *VDR* polymorphisms to sputum conversion time. The final analysis required new analytical methods and the construction of Cox models which were conducted by L. van der Merwe with my guidance and co-operation.

Chapter 7 is the concluding section where general discussion of the results is given, and possible future study directions. Problems and difficulties that were encountered are also discussed. As there were many novel methods and types of equipment used in assessing the genotypes, the methods are presented in Appendix 1. All the different technologies are directed at the same goal; the genotyping of polymorphisms. The use of programs available on the internet has greatly advanced analysis methods and a variety of free source software was used. Knowledge of the variety and type of analytical/ statistical algorithms supporting them ensures correct and appropriate use. The information from this unique population was also deposited in publicly accessible databases.

Chapter 2

South African Coloured Population

The major part of this Chapter was published in the following review and article:

Lombard Z, Brune AE, Hoal EG, **Babb C**, van Helden PD, Epplen JT, Bornman L (2006)
HLA class II disease associations in southern Africa.
Tissue Antigens 67:97-110

Barreiro LB, Neyrolles O, **Babb CL**, Tailleux L, Quach H, McElreavey K, Helden PD, Hoal EG, Gicquel B, and Quintana-Murci L (2006)
Promoter Variation in the DC-SIGN Encoding Gene CD209 Is Associated with Tuberculosis.
PLoS Medicine 3(2):e20.

The African continent is home to more than 2000 distinct ethnic groups and languages, responsible for a considerable genetic diversity among native African populations [349]. Southern Africa has seen waves of immigration from the African continent to the north in pre-recorded times, and Caucasoid and Asian during the past three and a half centuries. This population diversity within a uniform environment creates a unique opportunity to study population genetics and complex diseases of genetic etiology.

The southern African population includes native African groups such as the Zulu, Xhosa, Sotho, Tswana, Ndebele, Swazi, Pedi, Lemba, Shona, Khoi-Khoi (Khoi) and San; the white population descended from Dutch, British, German and French immigrants; admixture population groups such as the South African Coloured (SAC), and groups of Asian descent [223, 242].

Archaeological evidence indicates that southern Africa was first peopled by the ancestors of the San people, who exhibit the most ancient of African mitochondrial DNA lineages, the L1 lineage [349]. The Khoi migrated into southern Africa about 2000 years ago from the north and co-existed with the San. Bantu-speaking pastoralists moving south entered the eastern part of southern Africa from approximately AD 250 [223]. Pressure from these migration waves and subsequent European colonization drove the remaining San into the semi-desert areas of southern Africa. It is likely that there was substantial admixture between southern African blacks and the Khoi and San, as evidenced by the presence of San-specific Gm markers in the majority of black South African groups, particularly the Xhosa population [123, 242]. South Africa's largest population group, the Zulu, belongs to the Nguni linguistic division of the Niger-Congo people, as do the Xhosa [255]. The Gokomere people, a Bantu-speaking group of migrant farmers, inhabited the Great Zimbabwe site from about 500 A.D., displacing earlier Khoisan people. These were the ancestors of the Shona, who constitute approximately 80% of the current Zimbabwean population [242].

The SAC are an ethnic group of mixed ancestry. This distinct group emerged about 300 years ago and received genetic input from Malaysia and Indonesia, as well as from the European Caucasoid and indigenous Khoi, San and other black Africans [242, 361]. The Indian population in South Africa originated from different regions of the Indian subcontinent, including the Hindu group originating from South India, Muslims from the Gujerat and Maharashtra and the Tamil community, and is considered to be a hybrid population with some genetic contributions from Caucasoid, Black and Chinese [242, 267].

The genetic diversity in southern Africa is illustrated by the varying frequencies of both the human leukocyte antigen (HLA) -class I [45] and -class II [182] alleles in the different groups. An extensive literature search illustrated the diversity of HLA class II alleles within the various southern African populations (Table 2.1). When focusing on the SAC population group it can be seen that their HLA class II allele frequencies are unique in that they do not correspond with any particular population group. For example the DRw5 serological specificity reflects closer to the black population groups whereas DR7 and DQ3 reflect more Indian than black input. The DRw53 frequency appears to be unique to the SAC population, being midway between that of the black and white population group. Comparisons of polymorphism frequencies that are available in the HapMap database with the SAC population has revealed similar results, where

certain single nucleotide polymorphism (SNP) frequencies are similar to the Caucasian population but others are more similar to the Asian or West African populations (unpublished data M. Möller). This illustrates that there have been contributions from a number of different populations to create the genetic makeup of the SAC population. This also has implications for the linkage disequilibrium and haplotypes that will be present in this population, which as discussed in Chapter 1 is useful for genetic association studies.

The 'Coloured' population originated with the settlement of colonialists in the Cape over three centuries ago. Within the Cape there were the local indigenous populations, such as the Khoi and San, to which was added the white settlers and their slaves. After the subsequent biological and cultural assimilation the term Coloured population arose, as defined by the local community. Apart from a small fraction, known as the Cape Malay, which remained largely distinct because of their Islamic religion, the majority were Christians. Initially the distinction between different groups of people was based on religion, with Christianity being the dominant. Only over time was the distinction made through colour and racial background [55]. The name remains a label of the past and is not regarded as offensive, but is commonly used by the population group itself. In the absence of a viable alternative, the term SAC will be used in this thesis to refer to the population studied. It generally refers to people of the Cape who are not Bantu-speaking and do not have distinctive ancestry along both the maternal and paternal lines [223]. The term Coloured is also an official population group that is counted in Statistics SA censuses, which are reported to the Parliament of the Republic of South Africa. In 2006, 8.9% of the South African population was of the Coloured population group [1].

An alternative term that could be used is admixed population. However we feel this is too broad a term and the origins of the population might be misinterpreted. Understanding the ethnic diversity is important in understanding differences in disease incidence and this has been extensively discussed in the literature [19, 112, 277]. However, researchers should be aware of cultural sensitivities and the potential misuse of population specific genetic information in racial and social characterizations [315].

Table 2.1: HLA class II allele frequencies in southern Africa populations.

Serological Specificity	Allele	Black [76, 138]	Blacks in Cape Town [138]	Zulu [190, 217, 245] Zulu + Sotho [290]	SA Xhosa [77, 78, 248, 287]	Zimbabwe Shona [77]	Khoi [138]	San [138]	SAC [90, 198, 199, 248, 287, 296]	Natal Indians [246, 369]	Natal Tamils [216]	Natal Hindu [216]	Natal Muslim [216]	SA White [78, 138, 198]
DR1		0.03	0.047	0.047, 0.027	0.0428 , 0.111	0.075	-	0	0.147, 0.11, 0.057, 0.12	0.055 , 0.026	0.05	0.05	0.18	0.1056
	DRB1*0101	0.084	-	0.07	0.02	-	0.017	-	0.03	-	-	-	-	-
	DRB1*0102	-	-	-	0.01	-	0.052	-	0.02	-	-	-	-	-
	DRB1*0103	-	-	-	-	-	0.058	-	0.01	-	-	-	-	-
DR2		-	-	0.23	0.1153 , 0.21	0.216	-	-	0.326, 0.190 , 0.344	0.223 , 0.394	0.35	0.43	0.49	0.1495
DR2	DRB1*1501	0.165	-	-	0.0769	-	0.300	0.015	0.0833	-	-	-	-	-
	DRB1*1502	-	-	-	-	-	0.012	-	0.0282	-	-	-	-	-
	DRB1*1503	-	-	-	-	-	-	-	0.0244	-	-	-	-	-
	DRB1*1505	-	-	-	0.0576	-	-	-	-	-	-	-	-	-
	DRB1*02	-	-	0.13	-	-	-	-	-	-	-	-	-	-
DR3		0.29	0.104	0.34	0.2379 , 0.413	0.14	-	0.019	0.216 , 0.1099	0.0672 , 0.121	0.12	0.11	0.2	0.0979
	DRB1*03	-	-	0.36	-	-	-	-	-	-	-	-	-	-
DR3	DRB1*0301	0.195	-	-	0.1026	-	-	-	0.083	-	-	-	-	-
	DRB1*0302	-	-	-	0.1538	-	0.012	0.007	0.028	-	-	-	-	-
DR4		0.028, 0.13	0.104	0.123, 0.096	0.069, 0.141	0.039	-	0.435	0.21, 0.11, 0.25	0.0189, 0.205	0.27	0.08	0.06	0.1516, 0.28
	DRB1*0401	-	-	-	0.04	-	0.135	0.409	0.07	-	-	-	-	-
	DRB1*0403	-	-	-		-	-	-	0.01	-	-	-	-	-
	DRB1*0404	-	-	-	0.01	-	0.036	0.044	0.01	-	-	-	-	-
	DRB1*0405	-	-	-	0.01	-	0.006	-	0.02	-	-	-	-	-

Serological Specificity	Allele	Black [76, 138]	Blacks in Cape Town [138]	Zulu [190, 217, 245] Zulu + Sotho [290]	SA Xhosa [77, 78, 248, 287]	Zimbabwe Shona [77]	Khoi [138]	San [138]	SAC [90, 198, 199, 248, 287, 296]	Natal Indians [246, 369]	Natal Tamils [216]	Natal Hindu [216]	Natal Muslim [216]	SA White [78, 138, 198]
	DRB1*0408	-	-	-	-	-	-	-	0.01	-	-	-	-	-
DR4	DRB1*0401/05	0.061	-	-	-	-	-	-	-	-	-	-	-	-
	DRB1*04	-	-	0.06	-		-	0.063	-	-	-	-	-	-
DR5		-	-	0.33	0.172, 0.323	0.134	-	-	0.16, 0.30	0.084 , 0.166	0.18	0.2	0.16	0.132
DR5	DRB1*1101/1201	0.23	-	-	-	-	-	-	-					
DRw6		-	-	0.16	0.181, 0.305	0.186	-	-	0.11, 0.15, 0.21	0.0737				0.082
DR6		-	-	0.16	-	-	-	-	0.21	0.146	0.15	0.14	0.1	
DR6	DRB1*1302	0.2	-	-	-	-	0.115	0.116	-					
DR7		0.07	0.145	0.123, 0.154	0.0742, 0.137	0.085	-	0.073	0.24, 0.14	0.153, 0.297	0.28	0.3	0.24	0.132
	DRB1*07	-	-	0.13	-	-	-	-	-	-	-	-	-	-
DR7	DRB1*0701	0.11	-	-	0.0577	-	0.041	0.063	0.1387	-	-	-	-	-
DRw8		-	-	0.087, 0.039	0.006, 0.021	0.019	-	-	0.027, 0.054	0.012	-	-	-	0.032
DR8		0.01	0.019	0.01	-	-	-	0.056	0.05	0.079	0.1	0.09	0.04	-
	DRB1*08	-	-	0.10	-	-	-	-	-	-	-	-	-	-
DR8	DRB1*0802	0.109	-	-	0.006	-	-	-	-	-	-	-	-	-
	DRB1*0803	-	-	-	0.006	-	-	-	-	-	-	-	-	-
	DRB1*0804	-	-	-	0.006	-	0.029	0.0362	-	-	-	-	-	-
DRw9		-	-	0.02	0.0078, 0.012	-	-	-	0.0091, 0.023	0.003	-	-	-	0.017
DR9		0.01	0.01	0.005, 0.008	0.01	0.015	-	0.019	0.02	0.008	0.01	0	0	-
	DRB1*09	-	-	0.02	-	-	-	-	-	-	-	-	-	-

Serological Specificity	Allele	Black [76, 138]	Blacks in Cape Town [138]	Zulu [190, 217, 245] Zulu + Sotho [290]	SA Xhosa [77, 78, 248, 287]	Zimbabwe Shona [77]	Khoi [138]	San [138]	SAC [90, 198, 199, 248, 287, 296]	Natal Indians [246, 369]	Natal Tamils [216]	Natal Hindu [216]	Natal Muslim [216]	SA White [78, 138, 198]
DR 9	DRB1*0901	0.047	-	-	0.01	-	-	0.0145	0.0167	-	-	-	-	-
DRw10		-	-	0.02	0.028, 0.051	0.029	-	-	0.021, 0.0082, 0.017	0.045	-	-	-	0.014
DR10		0.02	0.026	0.03	-	-	-	0	0.016	0.104	0.11	0.07	0.1	-
	DRB1*10	-	-	0.01	-	-	-	-	-	-	-	-	-	-
	DRB1*1001	-	-	-	0.03	-	0.017	-	0.033	-	-	-	-	-
DR10	NONE	0.014	-	-	-	-	-	-	-	-	-	-	-	-
DR11		0.247	0.182	-	-	-	-	0.063	-	-	-	-	-	-
	DRB1*11	-	-	0.36	-	-	0.006	-	-	-	-	-	-	-
	DRB1*1101	-	-	-	0.10	-	-	0.015	0.130	-	-	-	-	-
	DRB1*1102	-	-	-	0.03	-	0.017	0.022	0.017	-	-	-	-	-
	DRB1*1104	-	-	-	0.04	-	0.006	-	0.028	-	-	-	-	-
	DRB1*12	-	-	0.08	-	-	-	-	-	-	-	-	-	-
	DRB1*1201	-	-	-	0.03	-	-	-	0.022	-	-	-	-	-
	DRB1*1202	-	-	-		-	0.012	-	0.050	-	-	-	-	-
	DRB1*13	-	-	0.33	-	-	-	-	-	-	-	-	-	-
	DRB1*1301	-	-	-	0.06	-	0.070	0.123	0.022	-	-	-	-	-
	DRB1*1302	-	-	-	0.07	-	-	-	0.017	-	-	-	-	-
	DRB1*1303	-	-	-	0.01	-	-	-	0.006	-	-	-	-	-
	DRB1*1305	-	-	-	0.01	-	-	-	-	-	-	-	-	-
	DRB1*1308	-	-	-		-	-	-	0.006	-	-	-	-	-
	DRB1*14	-	-	0.01	-	-	-	-	-	-	-	-	-	-
	DRB1*1401	-	-	-	0.03	-	0.012	0.007	0.022	-	-	-	-	-
	DRB1*1404	-	-	-	0.01	-	-	-	-	-	-	-	-	-
	DRB1*1601	-	-	-	-	-	0.006	-	-	-	-	-	-	-
	DRB1*1602	-	-	-	-	-	-	-	0.006	-	-	-	-	-

Serological Specificity	Allele	Black [76, 138]	Blacks in Cape Town [138]	Zulu [190, 217, 245] Zulu + Sotho [290]	SA Xhosa [77, 78, 248, 287]	Zimbabwe Shona [77]	Khoi [138]	San [138]	SAC [90, 198, 199, 248, 287, 296]	Natal Indians [246, 369]	Natal Tamils [216]	Natal Hindu [216]	Natal Muslim [216]	SA White [78, 138, 198]
DR12		0.05	0.083	-	-	-	-	0.019	-	-	-	-	-	-
DR13		0.14	0.078	-	-	-	-	0.245	-	-	-	-	-	-
DR14		0.03	0.021	-	-	-	-	0	-	-	-	-	-	-
DR15		0.05	0.099	-	-	-	-	0.019	-	-	-	-	-	-
DR16		0.00	0.021	-	-	-	-	0	-	-	-	-	-	-
DR51	DRB5*0101	0.165	-	-	-	-	-	-	-	-	-	-	-	-
DR52	DRB3*0101	0.552	-	-	-	-	-	-	-	-	-	-	-	-
DR53	DRB4*0101	0.211	-	-	-	-	-	-	-	-	-	-	-	-
DQ1		0.43	0.408	-	-	-	-	0.313	0.660	0.72	-	-	-	-
DQw1		-	-	0.66	0.44	-	-	-	0.420	-	0.69	0.77	0.88	0.39
DQ2		0.17	0.193	-	-	-	-	0.102	0.270	0.23	-	-	-	-
DQw2		-	-	0.23	0.11	-	-	-	0.150	-	0.22	0.27	0.12	-
DQ3		0.02	0.069	-	-	-	-	0.283	0.437	0.38	-	-	-	-
DQw3		-	-	0.31	0.25	-	-	-	0.44, 0.25	-	0.42	0.32	0.28	0.27
	DQB1*02	-	-	0.311	0.006	-	-	-	-	-	-	-	-	-
	DQB1*0201	-	-	-	0.167	-	0.111	0.084	0.193	-	-	-	-	-
	DQB1*0301	-	-	-	0.103	-		-	0.217	-	-	-	-	-
	DQB1*0302	-	-	-	0.039	-	0.144	0.497	0.089	-	-	-	-	-
	DQB1*0303	-	-	-	0.013	-		-	0.028	-	-	-	-	-
	DQB1*04	-	-	0.189	-	-		-	-	-	-	-	-	-
	DQB1*0402	-	-	-	0.160	-	0.055	0.058	0.050	-	-	-	-	-
	DQB1*05	-	-	0.23	-	-			-	-	-	-	-	-
	DQB1*0501	-	-	-	0.121	-	0.105		0.136	-	-	-	-	-
	DQB1*0502	-	-	-	-	-	0.017		0.033	-	-	-	-	-
	DQB1*06	-	-	0.554	-	-			0.006	-	-	-	-	-
	DQB1*0601	-	-	-	-	-	0.006		0.033	-	-	-	-	-

Serological Specificity	Allele	Black [76, 138]	Blacks in Cape Town [138]	Zulu [190, 217, 245] Zulu + Sotho [290]	SA Xhosa [77, 78, 248, 287]	Zimbabwe Shona [77]	Khoi [138]	San [138]	SAC [90, 198, 199, 248, 287, 296]	Natal Indians [246, 369]	Natal Tamils [216]	Natal Hindu [216]	Natal Muslim [216]	SA White [78, 138, 198]
	DQB1*0602	-	-	-	0.221	-	0.273	0.130	0.132	-	-	-	-	-
	DQB1*0603	-	-	-	0.103	-	0.041	0.029	0.050	-	-	-	-	-
	DQB1*0604	-	-	-	0.032	-	0.017	0.108		-	-	-	-	-
	DQB1*0605	-	-	-	0.026	-	0.116	0.029	0.006	-	-	-	-	-
	DQB1*0609	-	-	-		-	-	-	0.006	-	-	-	-	-
	DQ*0602	-	-	0.27	-	-	-	-	-	-	-	-	-	-
	DQ*0603-0608	-	-	0.162	-	-	-	-	-	-	-	-	-	-
	DQB1*07	-	-	0.338	-	-	-	-	-	-	-	-	-	-
	DQB1*08	-	-	0.04	-	-	-	-	-	-	-	-	-	-
DQ4		0.23	0.047	-	-	-	-	0.058	-	0.076	-	-	-	-
DQ7		0.16	0.255	-	-	-	-	0.150	-	-	-	-	-	-
DRw52		-	-	-	0.633	-	-	-	0.475	-	-	-	-	0.4462
DRw53		-	-	-	0.150	-	-	-	0.260	-	-	-	-	0.34
	DQA1*01	-	-	-	0.064	-	-	-	0.048	-	-	-	-	-
	DQA1*0101	-	-	-	0.071	-	0.122	0.015	0.083	-	-	-	-	-
	DQA1*0102	-	-	-	0.251	-	0.426	0.166	0.133	-	-	-	-	-
	DQA1*0103	-	-	-	0.100	-	-	-	0.050	-	-	-	-	-
	DQA1*0201	-	-	-	0.032	-	0.081	0.020	0.116	-	-	-	-	-
	DQA1*03	-	-	-	-	-	0.017	-	-	-	-	-	-	-
	DQA1*0301	-	-	-	0.090	-	0.214	0.605	0.121	-	-	-	-	-
	DQA1*0401	-	-	-	0.1538	-	0.023	0.029	0.033	-	-	-	-	-
	DQA1*0501	-	-	-	0.1529	-	0.126	0.036	0.229	-	-	-	-	-
	DQA1*0601	-	-	-	-	-	-	-	0.033	-	-	-	-	-
	DPB1*0101	-	-	-	-	-	0.175	0.235	0.128	-	-	-	-	-
	DPB1*0102	-	-	-	0.1026	-	-	-	-	-	-	-	-	-
	DPB1*0201	-	-	-	0.0897	-	0.164	0.029	0.100	-	-	-	-	-
	DPB1*0202	-	-	-	0.0128	-	-	-	-	-	-	-	-	-

Serological Specificity	Allele	Black [76, 138]	Blacks in Cape Town [138]	Zulu [190, 217, 245] Zulu + Sotho [290]	SA Xhosa [77, 78, 248, 287]	Zimbabwe Shona [77]	Khoi [138]	San [138]	SAC [90, 198, 199, 248, 287, 296]	Natal Indians [246, 369]	Natal Tamils [216]	Natal Hindu [216]	Natal Muslim [216]	SA White [78, 138, 198]
	DPB1*0210	-	-	-	-	-	-	-	0.006	-	-	-	-	-
	DPB1*0301	-	-	-	-	-	0.057	0.04	0.033	-	-	-	-	-
	DPB1*0401	-	-	-	0.06	-	0.114	0.29	0.172	-	-	-	-	-
	DPB1*0402	-	-	-	0.04	-	0.260	0.12	0.115	-	-	-	-	-
	DPB1*0403	-	-	-	-	-	-	-	0.006	-	-	-	-	-
	DPB1*0501	-	-	-	-	-	0.006	-	0.039	-	-	-	-	-
	DPB1*0601	-	-	-	0.01	-	-	-	-	-	-	-	-	-
	DPB1*0801	-	-	-	-	-	0.006	0.022	-	-	-	-	-	-
	DPB1*0901	-	-	-	-	-	-	-	0.022	-	-	-	-	-
	DPB1*1101	-	-	-	0.02	-	0.058	0.007	0.006	-	-	-	-	-
	DPB1*1201	-	-	-	-	-	-	-	0.006	-	-	-	-	-
	DPB1*1301	-	-	-	0.04	-	0.006	-	0.072	-	-	-	-	-
	DPB1*1401	-	-	-	-	-	0.023	0.007	0.028	-	-	-	-	-
	DPB1*1501	-	-	-	0.01	-	0.017	0.015	0.006	-	-	-	-	-
	DPB1*1601	-	-	-	-	-	-	-	0.011	-	-	-	-	-
	DPB1*1701	-	-	-	0.01	-	-	-	0.006	-	-	-	-	-
	DPB1*1801	-	-	-	0.06	-	0.006	-	0.017	-	-	-	-	-
	DPB1*1901	-	-	-	-	-	0.006	-	0.011	-	-	-	-	-
	DPB1*2001	-	-	-	0.02	-	-	-	0.006	-	-	-	-	-
	DPB1*2101	-	-	-	0.01	-	-	-	0.006	-	-	-	-	-
	DPB1*2201	-	-	-	-	-	0.006	-	-	-	-	-	-	-
	DPB1*2301	-	-	-	0.006	-	0.006	0.015	0.006	-	-	-	-	-
	DPB1*2501	-	-	-	0.006	-	-	-	-	-	-	-	-	-
	DPB1*2601	-	-	-	0.074	-	-	-	0.006	-	-	-	-	-
	DPB1*2701	-	-	-	0.019	-	-	0.007	0.017	-	-	-	-	-
	DPB1*2801	-	-	-	0.013	-	-	-	0.006	-	-	-	-	-
	DPB1*3001	-	-	-	-	-	0.012	0.080	-	-	-	-	-	-

Serological Specificity	Allele	Black [76, 138]	Blacks in Cape Town [138]	Zulu [190, 217, 245] Zulu + Sotho [290]	SA Xhosa [77, 78, 248, 287]	Zimbabwe Shona [77]	Khoi [138]	San [138]	SAC [90, 198, 199, 248, 287, 296]	Natal Indians [246, 369]	Natal Tamils [216]	Natal Hindu [216]	Natal Muslim [216]	SA White [78, 138, 198]
	DPB1*3101	-	-	-	0.013	-	-	-	0.006	-	-	-	-	-
	DPB1*3201	-	-	-	0.039	-	-	-	-	-	-	-	-	-
	DPB1*3301	-	-	-	0.032	-	-	-	-	-	-	-	-	-
	DPB1*3401	-	-	-	0.013	-	0.017	0.022	-	-	-	-	-	-
	DPB1*3901	-	-	-	0.006	-	-	-	0.006	-	-	-	-	-
	DPB1*4001	-	-	-	0.006	-	-	-	-	-	-	-	-	-
	DPB1*4101	-	-	-	0.026	-	-	-	-	-	-	-	-	-
	DPB1*4401	-	-	-	0.006	-	-	-	-	-	-	-	-	-
	DPB1*4601	-	-	-	0.006	-	-	-	-	-	-	-	-	-
	DPB1*4701	-	-	-	0.019	-	-	-	-	-	-	-	-	-
	DPB1*4901	-	-	-	0.019	-	-	-	-	-	-	-	-	-
	DPB1*5001	-	-	-	0.045	-	-	-	0.006	-	-	-	-	-
	DPB1*5101	-	-	-	0.013	-	-	-	0.006	-	-	-	-	-
	DPB1*5501	-	-	-	-	-	0.035	0.108	0.006	-	-	-	-	-

Population stratification

Although this population is considered a relatively homogeneous community that has received genetic contribution from different populations multiple generations ago [242, 361], population stratification between cases and controls is possible. This can be a confounding factor leading to a spurious positive association. Indeed, the use of admixed populations in association mapping studies can be very useful to identify disease-causing genetic variants that differ in frequency across parental populations. However, when the admixture event is too recent, allelic frequencies can differ coincidentally between cases and controls, reflecting a non-uniform genetic contribution from the parental populations to each subpopulation (i.e., cases and controls), rather than a genuine association between a given genetic variant and the phenotype under study. When this occurs, the study-cohort is said to present population stratification. To formally test and quantify the levels of background genetic differences [260], if any, between cases and controls, the entire cohort was genotyped for a panel of independent SNP markers which are (1) not in linkage disequilibrium with the candidate gene locus (*CD209* in Chapter 4) and with any other known gene, (2) randomly distributed along the genome, and (3) polymorphic among the major ethnic groups.

In collaboration with the Pasteur Institute in Paris, France we assessed the level of population stratification within the SAC population. Twenty five independent SNPs were genotyped (Appendix 1) in the cohort of 711 individuals used in Chapter 4. The mean χ^2 statistic among the 25 SNPs, for the comparison of allele frequencies between cases and controls, which represents the levels of stratification (μ) between the two groups [272]. This was 1.25 ($p=0.26$), implying that the two groups were not significantly stratified (Table 2.2).

The SAC population in Ravensmead and Uitsig (Chapter 1) is eminently suitable for genetic association studies assessing susceptibility to tuberculosis because of their uniform ethnicity and socio-economic status, high incidence of tuberculosis and low prevalence of HIV [32]. Being admixed, it represents a community originating from populations with different susceptibilities to tuberculosis and offers a unique opportunity to dissect the contributing genetic variants and their probable geographic/ethnic origins.

Table 2.2: Frequency distribution in the study cohort of 25 polymorphisms used to test for population stratification in the SAC population cohort.

rs number	Location	ObsHET°	PredHETΨ	HWP	MAF Cases (%)	MAF Controls (%)	ϰ	p
rs2048022	Chr4	0.509	0.497	0.588	43.4	48.3	3.469	0.0625
rs1380229	Chr8	0.429	0.468	0.034	36.1	38.4	0.815	0.3667
rs650389	Chr10	0.276	0.287	0.369	15.2	19.5	4.518	0.0335
rs870384	Chr12	0.511	0.499	0.591	48.3	46.8	0.311	0.5771
rs695982	Chr12	0.416	0.419	0.888	32.1	27.7	3.376	0.0662
rs708682	Chr15	0.216	0.212	0.826	11.8	12.3	0.086	0.7693
rs715774	Chr15	0.244	0.257	0.209	14.3	16.1	0.883	0.3473
rs1433456	Chr15	0.313	0.317	0.784	19.9	19.7	0.011	0.9169
rs807131	Chr17	0.478	0.453	0.183	35.5	33.9	0.398	0.5279
rs11672183	Chr19	0.214	0.209	0.649	12	11.7	0.018	0.8947
rs2024628	Chr20	0.513	0.494	0.333	42.2	46.5	2.627	0.1051
rs1028184	Chr2	0.461	0.464	0.9	34.2	39	3.53	0.0603
rs2056773	Chr3	0.492	0.473	0.32	39.5	37.1	0.808	0.3687
rs1479067	Chr5	0.398	0.395	0.972	25.9	28.4	1.099	0.2945
rs327747	Chr7	0.374	0.399	0.111	25.8	29.2	2.084	0.1488
rs12665321	Chr6	0.239	0.234	0.693	14.2	12.9	0.462	0.4969
rs1566838	Chr9	0.475	0.497	0.265	46.5	45.8	0.081	0.7762
rs12785524	Chr11	0.452	0.483	0.106	39	42.4	1.708	0.1912
rs975423	Chr13	0.429	0.464	0.056	35.1	37.8	1.134	0.2868
rs914904	Chr13	0.397	0.409	0.462	29.2	28.2	0.179	0.6723
rs876287	Chr14	0.459	0.484	0.182	41.3	40.9	0.024	0.8765
rs1582598	Chr16	0.378	0.394	0.321	27.5	26.5	0.164	0.6852
rs1364198	Chr16	0.361	0.364	0.908	25.2	22.7	1.242	0.2652
rs739259	Chr22	0.469	0.469	1	36.1	39	1.217	0.27
rs169479	Chr21	0.217	0.217	1	13.3	11.5	1.064	0.3024

°Observed Heterozygosity ΨPredicted Heterozygosity
HWP Hardy Weinberg Equilibrium Probability MAF, Minor allele frequency

Chapter 3

Candidate polymorphisms in genes *RANTES*, *CCR5*, *CCR2* and *SDF1*.

3.1 Introduction

Chemokine receptor 5 (CCR5) is expressed by T cells and macrophages. CCR5 is a receptor for macrophage inflammatory protein 1 (MIP-1) and regulated upon activation, normal T cell expressed and secreted (RANTES) which play an important role in the inflammatory response [137, 226]. CCR5 is a key entry point for human immuno-deficiency virus (HIV) into macrophages. It is a major co-receptor for the most commonly transmitted strains of HIV-1 [156] and this has been widely investigated. Toossi *et al* found that, in HIV patients, exposure to *Mycobacterium tuberculosis* (*M. tuberculosis*) *in vivo* may upregulate the expression of CCR5 on mononuclear cells [350]. As uninfected mononuclear cells are recruited to sites of active *M. tuberculosis* infection and exposed to *M. tuberculosis*, expression of CCR5 may be maintained at high levels allowing transmission of HIV-1 to these newly arrived cells [350]. There is evidence that an increased number of CCR5-positive monocytes accumulate in the lungs of tuberculosis patients [95]. There is increased expression of CCR5 on CD4+ T cells in response to tuberculosis infection, when whole blood is incubated with the *M. tuberculosis* cell wall component LAM [95, 145]. The mechanism by which CCR5 expression is increased in *in vitro* infection with low doses of *M. tuberculosis* and in active tuberculosis is still unknown [95]. However, *CCR5* knockout mice were capable of recruiting immune cells to the lung to form granulomas and successfully induced a Th1 response and controlled *M. tuberculosis* infection [7]. Although several reports show the importance of CCR5 in attracting Th1 cells to sites of inflammation [179], the mechanistic basis for this effect is not clear. Granuloma formation and the immune response are regulated by host and mycobacterial factors, of which one is mycobacterial heat shock protein 70 (myHsp-70), which stimulates dendritic cells to release pro-inflammatory mediators. Recently CCR5 has been identified as the crucial receptor for myHsp-70 mediated dendritic cell stimulation [93].

CCR5 has two functional promoters, with the downstream promoter stronger in monocytic and lymphocytic cell lines, as well as in CD4+ T cells [226]. The downstream promoter is responsible for the majority of transcriptional activity in activated cells and contains several potential transcription factor binding sites, including those for TFIID, AP-1, NFκB, CD28RC and STAT [57, 226]. Polymorphisms in the *CCR5* regulatory region result in changes in the level of *CCR5* gene transcription and mRNA expression. The structure and hypothesised function of CCR5 has been reviewed [224, 226].

The *CCR5* A59029G (also known as -2459G [137], rs1799987) single nucleotide polymorphism (SNP) is in the promoter of the *cis*-regulatory region of the *CCR5* gene [114]. The *CCR5*-59029G polymorphism results in decreased expression of CCR5 on the cell wall and therefore fewer potential entry points for HIV into the cell. The wildtype 'AA' genotype has higher promoter activity [48, 206] leading to increased expression of CCR5 on the cell wall. The 'G' allele has a marked effect on delaying HIV disease progression (up to a couple of years) [57, 206]. Although this association is not always found [85], this protective effect is independent of any protection conferred by *CCR2*-64I and *CCR5*Δ32 in relation to HIV infection [206]. The mutation *CCR5*Δ32 (rs333) is a 32 base pair (bp) deletion in the ORF in exon 4 of the *CCR5* gene, resulting in the expression of a truncated protein [179, 297]. Homozygosity for the *CCR5*Δ32 is

associated with resistance to HIV-1 infection and heterozygosity with decreased HIV-1 disease progression to AIDS [179, 297].

Chemokine receptor 2 (CCR2) is a receptor for monocyte chemo-attractant protein-1 (MCP-1) which is involved in the inflammatory response. MCP-1, a β chemokine, is upregulated by *M. tuberculosis* infection in macrophages [298, 350]. The *CCR2*-64I (rs1799864) [155, 206] promoter polymorphism is a 'G' to 'A' bp change leading to a conservative substitution, valine to isoleucine, in the first transmembrane domain [206], which has been found to have an effect on HIV-1 disease progression [11, 85, 210], but not transmission [210]. The mechanism of action of this polymorphism is unknown [167] but it may modulate CCR5 expression by binding to CCR5 in the cytoplasm [206]. A statistically significant decrease in the frequency of this polymorphism was seen between HIV seropositive and HIV seronegative controls in a study on the same ethnic group (SAC) [253].

The Stromal Derived Factor (*SDF1*) on chromosome 10, also known as pre-B cell growth stimulating factor (*CXCL12*), is a powerful chemo-attractant cytokine and is a CXCR4 ligand gene variant [394]. There is increasing evidence for SDF1 being involved in inflammatory diseases [173]. A single bp change, 'G' to 'A', occurs at position 801 in the 3' untranslated region, *SDF1* 3'A (rs1801157). This polymorphism is represented by the SDF1β transcript but not the SDF1α transcript [394]. The *SDF1* 3' heterozygote has been shown to have a protective effect on HIV infection in high risk individuals [394], where the frequency of *SDF1* 3'A homozygotes was low but had a strong protective effect against HIV progression in a large multi-centre study [394].

RANTES, also known as CCL5 and SCYA5, recruits several types of inflammatory cells including eosinophils, monocytes and T lymphocytes to sites of inflammation. It is a ligand for CCR1, CCR3, and CCR5. RANTES also induces monocyte haptotaxis (directed cell migration) [389], activation and proliferation of macrophages and T-cells and is found to have a significant role in granuloma formation [54, 294]. Mice deficient in RANTES have impaired monocyte and T-cell recruitment to sites of inflammation [191]. In particular, in the absence of the HIV protective *CCR5*Δ32 polymorphism, the *RANTES* promoter polymorphisms -403 and -28 altered HIV transmission and progression [115, 176, 205].

RANTES is located on chromosome 17 in the chemokine cluster, a region which has been shown to contribute to susceptibility to tuberculosis and leprosy in Brazilians [141] and has been associated with tuberculosis to a greater extent in the last few years. In tuberculosis patients, an increased level of RANTES has been found in bronchoalveolar lavage fluid and in lung alveoli during active pulmonary tuberculosis [162, 294]. The same study showed that *M. tuberculosis* infection of monocytes and alveolar macrophages stimulated release but not production of RANTES [294]. The recruitment of cells is vital for granuloma formation and containment of the infection. RANTES recruited infected macrophages more effectively than uninfected macrophages, but failed to exert direct antibacterial activity [327]. RANTES efficiently recruited *M. tuberculosis* infected macrophages to CD8+ T cells but failed to exert direct antibacterial activity [327], although a previous report showed that RANTES and MIP-1β suppressed intracellular growth of *M. tuberculosis* by two- to three-fold [298]. The chemokine RANTES plays

a significant role in inducing activation and proliferation of macrophages and T cells as well as lymphocytes and monocytes, which are essential for granuloma formation [294].

The *RANTES* -403A/G (rs2107538), -109T/C (rs1800825), and -28C/G (rs2280788) are promoter polymorphisms and the +1092 polymorphism (rs1065341) is found in the 3' untranslated region of the gene. The -403A variation has been shown to have 8-fold increased expression of RANTES [237]. The polymorphism creates a new consensus element for the GATA transcription family. This suggests a role in regulation of gene expression. The recruitment of cells is vital for granuloma formation and containment of the infection.

Because of the role of chemokines and their receptors in the immune and inflammatory response [63], we hypothesized that the *RANTES*, *CCR5*, *CCR2* and *SDF1* polymorphisms might be associated with tuberculosis susceptibility.

3.2 Results

3.2.1 Testing for an association with tuberculosis

The genotypes for all polymorphisms investigated in a South African Coloured (SAC) cohort showed no departure from Hardy Weinberg equilibrium. The *CCR5*Δ32 polymorphism was at too low a frequency (2%) in our population to provide enough power to draw any conclusions about an association with tuberculosis (Table 3.1). Only one tuberculosis case in the 331 samples tested was a homozygote for the deletion. Because of the low frequency of the deletion in this population, further analysis with a very large sample set would be required to determine whether there is an association with tuberculosis. With 80% power to detect an odds ratio of 1.86, it is unlikely that there is an association with the *CCR5*-59029 polymorphism and tuberculosis in our population (Table 3.1). The CCR5-59029A allele has a frequency of 0.47 in the control group. Results were submitted to NCBI (ss35528975). No association was found between the *CCR2*-64I polymorphism (frequency of the 'G' allele = 0.12) and tuberculosis (Table 3.1), SNP results were submitted to NCBI (ss37042809). There is no statistical difference between cases and controls with respect to *SDF1* 3'A (frequency of the A allele = 0.07) and tuberculosis.

In our study population, only the wild type *RANTES* -109 'TT' and -28 'CC' were found after testing 30 and 63 individuals, respectively. Therefore the frequency of these polymorphisms was too low to have statistical power for the number of samples available. The *RANTES* +1092 UTR polymorphism, genotyped in over 280 samples was not found to be associated with tuberculosis susceptibility in this population.

Table 3.1: Case-control analysis of polymorphisms in *RANTES*, *CCR5*, *CCR2*, and *SDF1* and tuberculosis in the SAC population. A: Genotype frequencies. B: Allele frequencies.

A:			**Genotype frequencies (%)**								
	Allele		**Cases**				**Controls**				**p value (Chi**
Polymorphism	**1**	**2**	**n**	**11**	**12**	**22**	**n**	**11**	**12**	**22**	**square)**
RANTES -403A	A	G	(n=433)	30	45	25	(n=461)	24.5	49.9	25.6	0.12
RANTES +1092	C	T	(n=163)	45	40	15	(n=119)	45	42	13	0.87
CCR5-59029	A	G	(n=185)	19	51	29	(n=179)	19	57	24	0.48
*CCR5*Δ32	wt	del	(n=158)	94	5	1	(n=173)	97	3	0	0.44
CCR2-64I	G	A	(n=146)	82	16	2	(n=160)	78	21	2	0.64
SDF1 3'A	A	G	(n=151)	<1	13	87	(n=150)	1	13	86	0.98

B:			**Allele frequencies (%)**						
	Allele		**Cases**			**Controls**			**P value (Fishers**
Polymorphism	**1**	**2**	**n**	**1**	**2**	**n**	**1**	**2**	**Exact)**
RANTES -403A	A	G	(n=433)	53	47	(n=461)	49	51	0.13
RANTES +1092	C	T	(n=163)	65	35	(n=119)	66	34	0.79
CCR5-59029	A	G	(n=185)	45	55	(n=179)	47	53	0.55
*CCR5*Δ32	wt	del	(n=158)	97	3	(n=173)	98	2	0.31
CCR2-64I	G	A	(n=146)	90	10	(n=160)	88	12	0.52
SDF1 3'A	A	G	(n=151)	7	93	(n=150)	7	93	0.88

After genotyping over 600 samples, a weak association was found with the *RANTES* -403 AA genotype, which was more frequent in cases and was significantly associated with risk of developing tuberculosis (p=0.026) (Table 3.2). With this sample set we did have 80% power to detect an odds ratio (OR) of 1.62. The *RANTES* -403 genotype containing a 'G' allele was found to be protective against developing tuberculosis (p=0.0122) with an OR of 1.622 (95% CI of 1.113 to 2.365). To ensure that this was a true association, an additional set of 281 samples were genotyped, namely 212 controls and 69 cases from the same population. This gave a final total of 461 controls and 407 cases, which had 80% power to detect an OR of 1.47. However, the association was lost as it went from p=0.026 to p=0.134 with the testing of additional samples (Table 3.1).

Table 3.2: Genotyping of *RANTES* -403 in the initial 613 samples of the SAC cohort.

	Controls		Cases		P value*,^
Genotype	n	%	n	%	
AA	53	21	111	30	
AG	136	55	165	45	
GG	60	24	88	24	0.0262
	249		364		
Allele					
A	242	49	387	53	
G	256	51	341	47	0.1304
Total	498		728		

*Genotype associations were analyzed using Chi Square.

^ Allele associations were analyzed by a Fisher exact Test.

Within the 461 tuberculosis cases genotyped, 48 were tuberculous meningitis (TBM) patients. Analysis of these cases on their own showed no association between susceptibility to TBM and the *RANTES* -403 polymorphism (Table 3.3). We had 80% power to detect an OR of 2.45 with a 95% CI.

Table 3.3: Genotyping of *RANTES* -403 in TBM patients in the SAC cohort.

	Controls		TBM		P value*
Genotype	n	%	n	%	
AA	113	24.5	18	37.5	
AG	230	49.9	22	45.8	
GG	118	25.6	8	16.7	0.11
Total	461		48		

* Chi-square analysis was done

The polymorphisms *CCR5*-59029 and *CCR2*-64I were found to have a linkage disequilibrium (LD) of D'=0.825, r2=0.09 in this population. The frequency of the *CCR5*Δ32 (2%) was too low to calculate LD. There was strong LD between *RANTES* -403 and *RANTES* +1092 with D'=0.972 and r2=0.469.

CCR2 and *CCR5* genes are separated by 19 kb on chromosome 3, therefore haplotype analysis was done between the polymorphisms investigated in these two genes [67]. The effects of *CCR2* and *CCR5* polymorphisms have been found to be additive in HIV cohorts [394]. There is increasing evidence that the haplotypes formed between *CCR2* and *CCR5* are important for susceptibility to disease. *CCR5* haplotype diversity has been found to be highest in Africans, and only a subset of these haplotypes were found in non-African populations [114]. As the haplotypes were inferred we also looked at manually constructed diplotypes. No significant difference in haplotype or diplotype frequencies was found between case and controls (p=0.59) (Table 3.4 and 3.5).

Table 3.4: Haplotype frequencies inferred by the programs Phase and Haploview for the *CCR5* and *CCR2* gene polymorphisms *CCR5*-59029, *CCR5*Δ32 and *CCR2*-64I in tuberculosis cases and healthy controls from the SAC population.

***CCR5*Δ32, *CCR5*-59029, *CCR2*-64I**	**All (n=459)**	**Tuberculosis Cases (n=218)**	**Controls (n=241)**
wt/G/G	0.530	0.539	0.518
wt/A/G	0.336	0.327	0.345
wt/A/A	0.103	0.096	0.108
Δ32/A/G	0.022	0.028	0.018
wt/G/A	0.011	0.010	0.011
Δ32/A/A	<0.01	0.000	<0.001
Δ32/G/G	<0.01	<0.001	<0.001

Global p value =0.59

Table 3.5: Diplotypes of the *CCR5*-59029, *CCR5*Δ32 and *CCR2*-64I gene in tuberculosis cases and healthy controls from the SAC population.

***CCR5*Δ32, *CCR5*-59029, *CCR2*-64I**	**Tuberculosis Cases**	**freq**	**Controls**	**freq**
hetero/AA/GG	0	0	2	0.03
wt/AA/GA	2	0.03	3	0.04
wt/AA/AA	1	0.01	2	0.03
homo/AA/GG	1	0.01	0	0
wt/AA/GG	8	0.1	10	0.13
hetero/AG/GG	4	0.05	2	0.03
wt/AG/GA	9	0.11	8	0.1
wt/AG/GG	28	0.35	35	0.44
wt/GG/GA	1	0.01	1	0.01
wt/GG/GG	25	0.32	17	0.21

Global p value =0.6604

No association was found between the *RANTES* haplotypes or diplotypes and tuberculosis susceptibility (Table 3.6 and 3.7). No *RANTES* -403G/ +1092T haplotype was identified in this population. The *RANTES* -403 AA/+1092 CT haplotype is more frequent in cases than

controls, but not significantly. With the use of GAIA [188] no interactions were observed between the various polymorphisms.

Table 3.6: Haplotypes inferred by Phase and Haploview for the *RANTES* gene polymorphisms -403 and +1092 in tuberculosis cases and controls from the SAC population.

***RANTES* -403, +1092**	**All (n=637)**	**Tuberculosis Cases (n=389)**	**Controls (n=257)**
A/C	0.180	0.172	0.183
A/T	0.338	0.330	0.342
G/C	0.477	0.490	0.468

Global p value = 0.28

Table 3.7: Diplotypes of the *RANTES* gene polymorphisms -403 and +1092 in tuberculosis cases and healthy controls from the SAC population.

***RANTES* -403, +1092**	**Tuberculosis Cases**	**Freq**	**Controls**	**Freq**
AA/ CC	4	0.03	5	0.05
AA/ CT	17	0.12	4	0.04
AA/ TT	22	0.15	14	0.13
AG/ CC	27	0.19	22	0.20
AG/ CT	40	0.28	39	0.35
GG/ CC	33	0.23	25	0.23
GG/ CT	0	0.00	2	0.02

Chi square p value = 0.2130

3.3 Discussion

Susceptibility to tuberculosis was investigated by means of a case-control study in a SAC population, to determine whether there is an association with polymorphisms in the genes *RANTES*, *CCR5*, *CCR2* and *SDF1*. All these genes are important in the inflammatory response and have been found to be associated with susceptibility and progression of a number of diseases, primarily HIV. We found no statistically significant associations between tuberculosis susceptibility in the SAC population and the chemokine *RANTES* -403 promoter polymorphism, or with the polymorphisms *CCR5*-59029, *CCR2*-64I and *SDF1* 3'A. The *CCR5*Δ32 and *RANTES* -109, -28, +1092 polymorphisms were all at too low a frequency in this population for meaningful analysis. We determined the frequencies of these polymorphisms that are known to influence HIV susceptibility and disease progression in this South African population, which has a relatively low prevalence of HIV at present [252, 253].

As stated above, the polymorphisms *CCR5*-59029, *CCR2*-64I and *SDF1* 3'A were not associated with tuberculosis in the SAC population. Both *CCR5*-59029 AA and *CCR2*-64I are promoter polymorphisms which increase expression of their respective genes. With the *CCR5*-59029 G polymorphism there is decreased expression of CCR5 on the cell wall and therefore fewer entry points for HIV into the cell. These mechanisms do not appear to be important in tuberculosis disease. The *CCR5*Δ32, *RANTES* -109, -28 and +1092 polymorphisms were at low frequencies in this population.

The *CCR5*Δ32 mutation appears to have an European origin with a heterozygote frequency of about 20% in Caucasians [48, 95, 179]. It is rare to almost absent in African and some Asian populations [297], and was not found in the San population of Southern Africa [67]. This deletion appeared to have been subjected to positive selection pressure in the European population and recent speculation on the causative agent has included bubonic plague, smallpox and hemorrhagic fever [67, 102].. However Sabeti *et al* (2005) found no evidence for selection of *CCR5*Δ32 and concluded that neutral evolution could be responsible for the frequencies seen in today's populations [293]. Homozygosity for the 32 bp deletion confers resistance against HIV infection and heterozygosity is associated with decreased HIV disease progression to AIDS [68, 179, 209, 297].

The 2% allele frequency of *CCR5*Δ32 found in the SAC population is probably due to admixture with the European population, which is also hypothesised to be the origin of the 2% frequency seen in African American populations [48, 156] (Table 3.8).

Table 3.8: Allele frequencies of the polymorphisms *CCR5*-59029, *CCR5*Δ32, S*DF1* 3'A, *CCR2*-64I, *RANTES* -403 and +1092 in the literature and the SAC controls (in bold).

	SA Caucasian	**US Caucasian**	**SA Black**	**African American**	**US Hispanic**	**SAC**
CCR5 59029 (A) rs1799987	0.51	0.53 [310] 0.44 [67] 0.57 [206]	0.48	0.46 [310] 0.56 [67] 0.43 [206]	0.51 [311] 0.68 [206]	**0.48**
*CCR5*Δ32 rs333	0.09 [392] 0.09 [67]	0.10 [48] 0.06 [156]	0 [127, 392]	0.02 [48, 67, 156]	0.02 [156] 0.03 [67]	0.03 [127] **0.02**
SDF1 3'A rs1801157	0.21 [392]	0.19 [310], 0.21[394]	0.01[392]	0.077 [310], 0.06 [394]	0.16 [394] , 0.19 [310]	**0.07**
CCR2 64I rs1799864	0.07 [392] 0.10 [253]	0.09 [310]	0.13 [392] 0.17 [253]	0.16 [310]	0.13 [310]	**0.12** 0.17 [253]
RANTES -403A rs2107538	-	0.19	-	0.48	0.24	**0.51**
RANTES +1092T rs1065341	-	0.06	-	0.32	0.05	**0.34**

An association between susceptibility to tuberculosis and the *RANTES* -403 polymorphism was initially found in our South African cohort, with the AA genotype being more frequent in cases than controls, but this did not hold up to validation. It was initially characterised by the AA genotype being more frequent in tuberculosis cases than controls. However when additional samples were genotyped, the significant difference between cases and controls was lost. This is an indication of the importance of replicating association studies, as they might not be true associations. The initial p value in this study was not strong (p=0.03) and after multiple testing correction using Bonferroni, the association would have been lost. No association was found with TBM and the *RANTES* -403 polymorphisms either.

Haplotypes were inferred for *CCR5*-59029, *CCR5*Δ32 and *CCR2*-64I polymorphisms but no significant associations were found with any specific haplotype. However a larger percentage of cases had the *RANTES* -403 AA/ +1092 CT diplotype compared to controls (Table 3.7) but this was not significant.

Previous studies have shown that the *RANTES* -403A allele is in strong LD with the *RANTES* -28G polymorphism and the haplotype A/G is associated with increased production of RANTES [176, 237]. Therefore it has been predicted that the *RANTES* -403A allele would be associated with increased production of the chemokine [237]. It is observed that the AA genotype is a risk factor, and the G allele-containing genotypes are protective.

A study by Stegelmann *et al* [327] has shown that RANTES is important in the adaptive immune response for *M. tuberculosis* infection and that previous exposure is important for macrophages to trigger RANTES expression in CD8+ T cells. BCG vaccination is widespread in the population studied here and there is a tremendous infection pressure in this community [372] so the probability that controls have been exposed to *M. tuberculosis* is extremely high. In a randomised survey within our study community, of 2401 adults analysed 76% had a positive TST result [71, 393]. In this community the adaptive immune response is more than likely activated, so CD8+ T cells will express RANTES.

Although there has been a fair amount of speculation on the similarities of pathogen entry into the cells in HIV and tuberculosis, it would appear from these results that the method of pathogenesis is sufficiently different that SNPs in *RANTES*, *CCR2*, *CCR5* and *SDF1*, that have an effect on HIV infection and/or progression to disease, do not affect the development of tuberculosis disease.

■

Chapter 4

Re-sequencing of the DC-SIGN (*CD209*) gene and genotyping of polymorphisms, including a length variation and a L-SIGN length variation.

These results were published:

Barreiro LB, Neyrolles O, **Babb CL**, Tailleux L, Quach H, McElreavey K, Helden PD, Hoal EG, Gicquel B, Quintana-Murci L (2006) Promoter Variation in the DC-SIGN Encoding Gene CD209 Is Associated with Tuberculosis.
PLoS Medicine 3:e20

Barreiro LB, Neyrolles O, **Babb CL**, van Helden PD, Gicquel B, Hoal EG, Quintana-Murci L (2007) Length Variation of DC-SIGN and L-SIGN Neck-Region has no Impact on Tuberculosis Susceptibility.
Human Immunology 68:106-12

4.1 Introduction

Phagocytic cells express a range of cellular receptors, known as pattern recognition receptors (PRR), involved in sensing micro-organisms [142]. One of the ways macrophages and dendritic cells (DC) perform their function in the innate immune system is through PRR. DCs bear a range of PRR, such as C-type lectins and Toll-like receptors, involved both in recognition of conserved microbial molecules and in the induction of adaptive immunity [142, 208, 341]. In particular, C-type lectins detect pathogens by their characteristic carbohydrate structures and internalise them for further antigen processing and presentation [207]. PRR bind to conserved microbial ligands, promoting phagocytosis and antigen presentation, and trigger intracellular signalling and cytokine secretion [264]. The quality of this initial pathogen recognition can have important consequences for both the outcome of infection and the pathogenesis of infectious disease. Two particular PRR of the C-type lectin receptor family, dendritic cell-specific intercellular adhesion molecule (ICAM)-3 grabbing nonintegrin (DC-SIGN, encoded by *CD209*) and dendritic cell-specific ICAM-3 grabbing nonintegrin related (*L-SIGN*, also known as *DC-SIGNR*) have recently been the focus of considerable attention [65, 107, 108, 257, 319]. These two lectins, are encoded by two genes located on chromosome 19p13.2-3 within a ~26 kb segment [25, 318].

DC-SIGN is specifically, though not exclusively, expressed on DCs and functions both as a cell adhesion receptor and as a PRR [321]. As an adhesion receptor, it plays an important role in many DC functions, such as DC-T cell interaction and DC migration [106, 107]. Besides its cellular recognition role, DC-SIGN serves as pathogen uptake receptor and mediates interactions with a plethora of pathogens other than *M. tuberculosis* [108]. Indeed, it has been shown that DC-SIGN allows DCs to capture other bacteria such as *Helicobacter pylori* and certain *Klebsiella pneumonia* strains, but also viruses such as HIV-1, Ebola, cytomegalovirus, hepatitis C virus (HCV), dengue, and severe acute respiratory syndrome coronavirus (SARS-coV), and parasites like *Leishmania pifanoi* and *Schistosoma mansoni* [8, 12, 30, 60, 105, 118, 185, 203, 340, 366]. In addition, recent data suggest that DC-SIGN may mediate intracellular signalling events leading to cytokine secretion and, on this basis, it has been proposed that the lectin could be used by pathogens, including *M. tuberculosis*, as part of an immune evasion strategy to their own advantage [109, 336].

Interactions between the tubercle bacillus and host phagocytes are crucial for immunity to mycobacteria and for tuberculosis pathogenesis [149]. DC-SIGN has been shown to be an important *M. tuberculosis* receptor on the surface of human monocyte derived DCs [109, 338], and, more recently, it has been shown that L-SIGN can also interact with the tubercle bacillus [154]. Both DC-SIGN and L-SIGN binding to *M. tuberculosis* is mediated by the mycobacterial cell-wall component mannosylated lipoarabinomannan (ManLAM) [365, 366] but does not bind to AraLAM, which is effectively LAM without the mannose cap [189]. This is intriguing because AraLAM is abundant in fast-growing atypical, avirulent mycobacteria, such as *M. smegmatis* and *M. fortuitum* whereas ManLAM is abundant in slow-growing virulent mycobacteria such as *M. tuberculosis* and *M. leprae* [109, 338, 366]. Pathogens that interact with DC-SIGN appear to have the features of being able to cause chronic infections that can last a lifetime. There is evidence that they manipulate the Th1 versus the Th2 cell balance which is central to their persistence [366].

DC-SIGN and L-SIGN exhibit high nucleotide (73%) and amino-acid (77%) identity, and identical exon-intron organization [318]. An additional characteristic of both lectins is the presence of a neck region, made up of primarily 7 highly conserved 23 amino-acid repeats, that separates the carbohydrate recognition domain (CRD) involved in pathogen binding from the transmembrane region. With regard to expression profiles, DC-SIGN is expressed mainly on endocytic cells, such as DCs and macrophages, whereas L-SIGN is expressed on endothelial cells in liver and lymph nodes, and in cells lining placental capillaries [24, 320, 321]. DC-SIGN and L-SIGN share the ability to bind high–mannose oligosaccharides through their CRD.

The extent to which the length of the neck region of both *DC-SIGN* and *L-SIGN* might have an impact on the host susceptibility to tuberculosis is unclear at present. This tandem-repeat region, which shows a varying degree of length polymorphism [25, 227], is involved in assembling both lectins into a tetrameric protein conformation on the cell surface, and the length of this region can critically influence the pathogen-binding properties of the CRD of these proteins [87, 117, 316]. At the population level, the length of the *DC-SIGN* neck region is highly conserved (mainly 7 repeats), whereas the *L-SIGN* neck region exhibits an extraordinarily high level of heterozygosity [20]. Furthermore, several studies suggest that the number of *DC-SIGN* and/or *L-SIGN* repeat units can contribute to the risk of HIV-1 [175, 177] and SARS infections [51], as well as to HCV replication efficacy [231].

Considering the ability of both DC-SIGN and L-SIGN to bind *M. tuberculosis*, the fact that neck-region length-variation may determine the ligand-binding capacities of these lectins and the observation that variation in these regions is associated with a number of infectious diseases, we hypothesized that variations in the *DC-SIGN* and *L-SIGN* neck regions might affect individual susceptibility to tuberculosis. To test this hypothesis, we explored the *CD209* promoter region and the relationship between the *DC-SIGN* and *L-SIGN* tandem repeat variation in the neck region and susceptibility to tuberculosis.

4.2 Results

4.2.1 *DC-SIGN* gene variation and susceptibility to tuberculosis

The *CD209* gene was sequenced in over 20 random South African Coloured (SAC) samples. L. Barreiro (Pasteur Institute) conducted the analysis of these results in collaboration with ourselves and determined which single nucleotide polymorphisms (SNP) were possibly tagSNPs. In total, eight haplotype-tagging SNPs were genotyped in all samples. Two variants located in the *CD209* promoter region (-871 A/G and -336 A/G) exhibited a frequency distribution significantly distorted between tuberculosis patients and controls, as indicated by a Chi-square (χ^2) test (Table 4.1).

Table 4.1: *DC-SIGN* genotype distributions in patients with tuberculosis and in healthy controls.

			Genotypes frequencies (%)							
			Cases (N=351)			Controls (N=360)			Cases vs. Controls†	
htSNP*	**1**	**2**	**11**	**12**	**22**	**11**	**12**	**22**	**p**	**OR**
-939	G	A	52.1	39	8.8	45.3	46.9	7.8	0.07	0.76 (0.57-1.02)
-871	A	G	83.2	15.7	1.1	72.8	26.1	1.1	**$8.2x10^{-4}$**	**1.85 (1.29-2.66)**
-336	A	G	29.3	50.4	20.2	38.1	43.3	18.6	**0.01**	**1.48 (1.08-2.02)**
-139	A	G	8.3	37.6	54.1	7.2	43.1	49.7	0.24	0.84 (0.62-1.13)
2392	G	A	96.6	3.4	0	95	4.7	0.3	0.29	0.67 (0.32-1.42)
3220	T	C	74.6	23.4	2	75.8	22.8	1.4	0.71	1.07 (0.76-1.50)
3838	A	C	80.1	17.9	2	84.7	14.4	0.8	0.1	1.38 (0.94-2.04)
4235	G	C	51.6	41	7.4	57.2	37.8	5	0.13	1.26 (0.94-1.69)

* All htSNPs (haplotype tagging SNPs) were in Hardy Weinberg (HW) equilibrium

† The Homozygotes for the most frequent allele were compared with the sum of the homozygotes and heterozygotes for the rare allele

For the -871 variant, genotypes GG and GA were less frequently observed in cases (16.8%) compared to the control group (27.2%) ($p = 8.2 \times 10^{-4}$). For the -336 variant, genotypes GG and GA were more frequent in cases (70.6%) than in controls (61.9%) ($p = 0.01$). These observations suggest that the alleles -871A (odds ratio (OR): 1.85; 95% confidence interval (CI): 1.29–2.66) and -336G (OR: 1.48; 95% CI: 1.08–2.02) increase the risk of developing tuberculosis in our cohort.

At the haplotype level (Table 4.2), a χ^2 test showed that the global distribution of haplotype frequencies was significantly different between cases and controls ($p=1.2x10^{-3}$). One haplotype (H3) was found to be the haplotype responsible for most of the distorted frequency distribution (Table 4.2). This haplotype, which contains both -871G and -336A, was found to be strongly associated with the control group ($p=1.6x10^{-3}$; OR: 1.7; 95% CI: 1.22– 2.38). The associations with this haplotype, and with the -871 SNP, remained highly significant ($p = 1.3 \times 10^{-2}$ and 6.6×10^{-3} respectively), even after the conservative Bonferroni correction for multiple testing. Figure 4.1 shows the linkage disequilibrium (LD) plot of the eight SNPs genotyped in the *CD209* promoter region. Block 1 contains the -871 and -336 SNPs. Permutation tests were done in Haploview, which also found block 1 -871G/-336A haplotype to be associated with tuberculosis ($p=0.0022$).

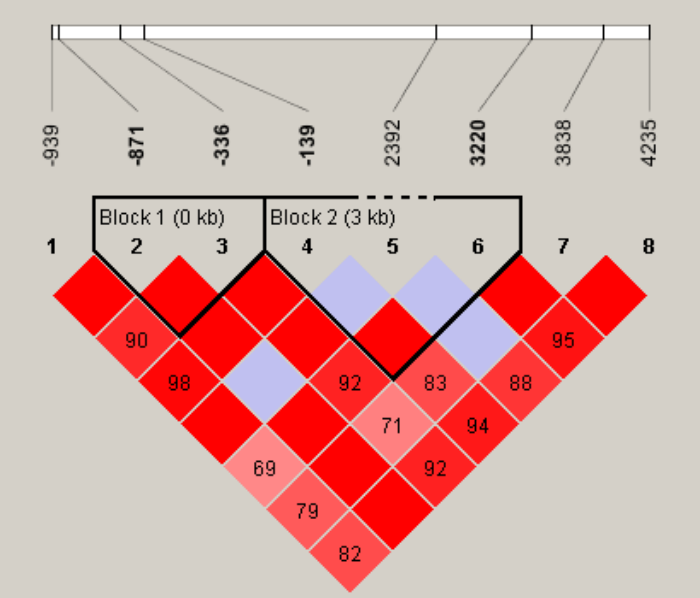
Table 4.2: *DC-SIGN* Haplotype distribution in patients with tuberculosis and in healthy controls.

Haplotype‡	Haplotype frequencies (%)		Cases vs. Controls§	
	Cases	Controls	p	OR
H1-AAAGGTCG	10	6.2	**0.01**	**1.70 (1.15-2.52)**
H2-AAAGGTAG	7.6	9.3	0.25	0.80 (0.55-1.16)
H3-AGAGGTAG	8.9	14.2	**$1.6x10^{-3}$**	**1.70 (1.22-2.38)**
H4-GAAAGTAG	26	27.9	0.41	0.91 (0.72-1.15)
H5-GAGGGTAG	17.3	15	0.25	1.17 (0.88-1.55)
H6-GAGGGTAC	12.2	8.5	**0.02**	**1.49 (1.05-2.11)**
H7-GAGGATAC	1.7	2.2	0.44	0.77 (0.36-1.63)
H8-GAGGGCAC	12.8	11.2	0.35	1.16 (0.84-1.60)

‡ Haplotypes with frequency greater than 1%. The alleles are ordered from SNP-939 until SNP4235

§ The frequency of each haplotype in cases and controls was compared with the sum of all the others

Figure 4.1: Linkage Disequilibrium plot for the eight polymorphisms genotyped in *DC-SIGN* gene of the cases and controls.



To correct for potential stratification, we divided the χ^2 values obtained for our candidate gene *CD209* by the level of stratification detected in this population sample set (1.25) (Chapter 2) [272]. Even after such a conservative correction, the associations observed with -336 and -871 as well as with H3 remained significant (*CD209*-336 p = 2.8 x 10^{-2}; *CD209* -871 p = 2.7x 10^{-3}; H3 p = 4.8 x 10^{-3}). These observations therefore support the idea that the -871G and -336A variants are indeed genuinely associated with a protective role against tuberculosis.

4.2.2 Repeat units in neck regions of *DC-SIGN* and *L-SIGN*

The allelic and genotype frequencies of the *DC-SIGN* and *L-SIGN* neck-region tandem repeats in the 351 tuberculosis patients and 360 healthy controls are summarized in Table 4.3 and 4.4, respectively. When examining the allelic frequencies of repeat units for *DC-SIGN* and *L-SIGN* neck regions, no statistical differences were observed between tuberculosis cases and healthy controls (Table 4.3).

Table 4.3: *DC-SIGN* and *L-SIGN* neck-region allelic frequencies among patients with tuberculosis and healthy controls.

	DC-SIGN				***L-SIGN***			
Alleles	**Cases (n =702)**	**Controls (n=720)**	**p**	**OR (95% CI)**	**Cases (n=702)**	**Controls (n=720)**	**p**	**OR (95% CI)**
4	0.14	0	0.49	nc	0.43	0.69	0.5	0.61 (0.15-2.58)
5	0.28	0	0.24	nc	9.40	11.81	0.14	0.78 (0.55-1.09)
6	1.28	0.97	0.58	1.32 (0.49-3.57)	28.49	24.17	0.06	1.25 (0.99-1.59)
7	98.15	98.89	0.25	0.60 (0.25-1.45)	58.40	59.72	0.61	0.95 (0.77-1.17)
8	0.14	0.14	1	1.02 (0.06-16.43	0.43	0.97	0.22	0.44 (0.11-1.70)
9	-	0	-	nc	2.85	2.64	0.81	1.08 (0.57-2.05)

n= Number of chromosomes analyzed

nc= not computable

Table 4.4: *DC-SIGN* and *L-SIGN* neck-region genotype frequencies among patients with tuberculosis and healthy controls.

	DC-SIGN				***L-SIGN***			
Genotype	**Patients (n=351)**	**Controls (n=360)**	**p**	**OR (95% CI)**	**Patients (n =351)**	**Controls (n=360)**	**p**	**OR (95% CI)**
4/4	0	0	-	nc	0	0.28	1.00	nc
5/4	0	0	-	nc	0.28	0	1.00	nc
5/5	0	0	-	nc	0.57	2.5	0.06	0.22 (0.04-1.04)
6/4	0	0	-	nc	0.28	0.56	1.00	0.51 (0.05-5.67)
6/5	0	0	-	nc	5.13	5.28	0.93	0.97 (0.50-1.88)
6/6	0	0	-	nc	7.12	5.83	0.48	1.24 (0.68-2.26)
7/4	0.28	0	1.00	nc	0.28	0.28	1.00	1.03 (0.06-16.46)
7/5	0.57	0	0.24	nc	12.25	13.33	0.67	0.91 (0.58-1.41)
7/6	2.56	1.94	0.58	1.33 (0.49-3.60)	36.18	29.72	0.07	1.34 (0.98-1.83)
7/7	96.30	97.78	0.24	0.59 (0.24-1.44)	31.34	35.28	0.27	0.84 (0.61-1.14)
8/6	0	0	-	nc	0	0.28	1.00	nc
8/7	0.28	0.28	1.00	1.03 (0.06-16.46)	0.85	1.67	0.51	0.51 (0.13-2.05)
9/6	0	0	-	nc	1.14	0.83	0.72	1.37 (0.30-6.17)
9/7	0	0	-	nc	4.56	3.89	0.66	1.18 (0.57-2.46)
9/9	0	0	-	nc	0	0.28	1.00	nc

n= number of individuals nc= not computable

In the case of *DC-SIGN*, the 7-repeat allele was by far the most frequently observed, with a frequency of more than 98%. As for *L-SIGN*, the 7-repeat and the 6-repeat alleles account for more than 83% of the overall diversity. We next examined whether the frequency distributions of *DC-SIGN* or *L-SIGN* neck-region genotypes were significantly different between cases and controls. Again, no significant differences were detected between diseased individuals and healthy controls for either *DC-SIGN* or *L-SIGN*. With regard to *DC-SIGN*, low genotypic variation was observed in accordance with the allelic data. The 7/7 genotype accounted for nearly all genetic variation (more than 96%), and genotypes 7/4, 7/5, 7/6, and 7/8 were observed at very low frequencies (Table 4.4). For *L-SIGN*, the genotypes 7/7 and 7/6 were present at similar frequencies (30 –36%), followed by 7/5 (~12%), 6/6 (~6%), 6/5 (~5%), and 9/7 (~4%).

4.2.3 Frequency distribution of *DC-SIGN* -871 and -336 polymorphisms

In order to gain insights into the frequency distribution, these two SNPs, were genotyped in 254 human chromosomes from sub-Saharan Africa, Europe, and East Asia as well as in eight chimpanzee chromosomes. Samples from African, Asian and European populations were provided by L. Quintana-Murci, Pasteur, France. We observed that the -871G and -336A forms, which we propose offer protection against tuberculosis, corresponded to the derived allele in humans; we also observed that these forms are present at higher frequencies in Eurasians as compared to Africans (Table 4.5).

Indeed, the -871G is absent in African populations, whereas it reaches high frequencies (20%–40%) in European and Asian populations. Given the absence of the haplotypic combination of -871G and -336A among sub-Saharan Africans, its presence among SAC suggests that it was introduced through the historically well known admixture with Europeans and Asians (Chapter 2) [361]. This observation highlights the power of using admixed populations to better understand historical issues associated with the geographic/ethnic origin of disease-affecting alleles, provided that their prevalence varies in the ancestors of the admixed population (i.e., different frequency of H3 in Africans versus non-Africans; Table 4.5) (Chapter 2). The SAC population represents a present-day homogeneous population which originated from the variable admixture of different populations, such as African Khoisan and Bantu-speakers, Malaysians, Indians, and Europeans (Chapter 2) [242, 361]. Consequently, the SAC population presents a large degree of genetic diversity, resulting in a high number of alleles or genotype combinations that can be used in association studies. For example, the *L-SIGN* genotypes 6/5, 7/5, and 9/7 are observed at relatively high frequencies among the South Africans (providing evidence for the genetic input received from European and Asian populations), whereas they are rare or even absent in other sub-Saharan African populations [20]. The presence of these genotypes in the SA population offers a unique opportunity for testing their association with disease in a single population, a hypothesis that would be difficult to test in other African populations because these genotypes are found at a very low frequency, or are even absent.

Table 4.5: Frequency distribution of the eight *CD209* polymorphisms genotyped in the multi-ethnic panel of 127 individuals (n=254 chromosomes) as well as in the 711 individuals of the SAC cohort (n=1422 chromosomes)

		Population frequencies (%)			
Polymorphism	Variant*	African (N=84)	Asian (N=86)	European (N=86)	SAC (N=1422)
-939	A	45.1	29.1	54.7	29.8
	G	54.9	70.9	45.3	70.2
-871	A	100	79.1	61.6	88.4
	G	0	20.9	38.4	11.6
-336	G	37.8	5.8	20.9	42.8
	A	62.2	94.2	79.1	57.2
-139	G	87.8	33.7	75.6	72.1
	A	12.2	66.3	24.4	27.9
2392	G	100	98.8	91.9	97.8
	A	0	1.2	8.1	2.2
3220	C	14.6	0	0	13.2
	T	85.4	100	100	86.8
3838	A	70.7	100	90.7	90.5
	C	29.3	0	9.3	9.5
4235	G	76.8	96.5	86	74.1
	C	23.2	3.5	14	25.9

*The allele in bold corresponds to the derived allele when compared with the sequence of the chimpanzee

4.3 Discussion

In the context of tuberculosis, it has been suggested that present-day susceptibility to tuberculosis is determined by previous history of exposure [243]. There is fairly convincing evidence that tuberculosis has been endemic in Europe for several hundred years, whereas in Africa it has probably been rare before contact was initiated with Europeans [66, 325, 326]. It is expected therefore that *M. tuberculosis* has exerted stronger selective pressures on European than African populations [325]. Our results lend support to this hypothesis and suggest that the protective alleles -871G and -336A increased in frequency in non-African populations as a result of genetic adaptation to a longer period of tuberculosis exposure. The potential impact of tuberculosis on the frequency of resistant alleles in European populations has recently been addressed using epidemiological data and statistical modelling [174]. They sought to evaluate the expected changes in resistant allele frequencies, during the 300 year period corresponding to the peak epidemics of tuberculosis in Europe. They concluded that if a given resistant allele was at a low frequency in the beginning of an epidemic, selection by *M. tuberculosis* alone would increase the frequency of this allele, but not significantly. In this context, since DC-SIGN is known to interact

with a vast range of pathogens, it is indeed likely that the increased frequencies observed today for both -871G and -336A in non-African populations (specially for -871G which is absent in sub-Saharan Africans) may have been driven, not only by the selective pressures imposed by *M. tuberculosis*, but also by other infectious agents. Indeed, two independent studies have reported a genetic association between the -336A variant and protection against parental HIV infection [200] and severity of dengue pathogenesis [295]. Although HIV infection, for example, is too recent to have left any signature of selection on *CD209*, these observations emphasize the possible action of other pathogens in shaping the patterns of variability of this gene.

From a functional point of view, the -336A allele has been shown to affect an Sp1-like binding site and to modulate transcriptional activity *in vitro* by increasing the levels of expression of DC-SIGN [295]. In the context of tuberculosis, increased DC-SIGN expression levels by DCs may result in better capture and processing of mycobacterial antigens, leading to a stronger and wider T-cell response. In addition, it has recently been shown that DC-SIGN expression is markedly induced in alveolar macrophages in active tuberculosis patients and that *M. tuberculosis* is preferentially phagocytosed by DC-SIGN-expressing macrophages in these individuals [337, 338]. Thus, the higher prevalence observed among healthy individuals of the -336A variant, which is associated with increased DC-SIGN expression, may underlie an increased efficiency of host phagocytes, such as DCs and macrophages, to control the infection. In addition to the -336A variant, our genetic data showed a strong association of the -871G allele with healthy controls, suggesting also a functional consequence of this variant that, either alone or in combination with -336A, remains to be defined.

The results also indicate that the number of repeats of the *DC-SIGN* and *L-SIGN* neck regions does not seem to influence the host susceptibility to develop tuberculosis. In the case of *DC-SIGN*, our results are in agreement with a case-control study in a cohort of north-western Colombian origin [113]. In this report, the authors analyzed *DC-SIGN* neck-region variation in a cohort of 110 tuberculosis patients and 299 matched controls, and observed no statistical differences between the two study groups. Thus, both studies support the notion that length variation of the *DC-SIGN* neck region does not influence the host susceptibility to develop tuberculosis.

In the context of other infectious diseases, the only positive association published so far between *DC-SIGN* neck-region variation and susceptibility to infectious disease is restricted to HIV-1 infection [177]. In this study, they observed an excess of heterozygous individuals for *DC-SIGN* tandem-repeats (repeat deletion 7/6) in a group of repeatedly exposed seronegative individuals, compared HIV-1 seronegative and HIV-1 seropositive individuals. These observations were interpreted to imply that heterozygosity in the *DC-SIGN* neck region can be associated with reduced susceptibility to HIV-1 infection [177]. At the level of the general population, it is of interest that the *DC-SIGN* neck region exhibits a very low level of the polymorphism [20, 24]. The entire Human Genome Diversity Panel (CEPH panel), which is composed of more than 1,000 control individuals from 52 different ethnic groups, was recently screened for repeat variation in the neck regions of both *DC-SIGN* and *L-SIGN* [20]. For *DC-SIGN*, it was observed that the 7-repeat allele accounts for nearly all genetic variation (~99%), and that the other alleles, which range from 2–10 repeats, are present at very low frequencies. In addition, the levels of

sequence variation in the entire *DC-SIGN* coding-region, particularly of those that affect amino-acid identity, were found to be extremely low [20]. The low levels of genetic variation observed in the *DC-SIGN* coding region are also reflected in the context of disease association studies. Indeed, the different associations published so far between *DC-SIGN* genetic variation and susceptibility to infectious disease always involve polymorphisms in the *DC-SIGN* promoter region, and not polymorphisms in the coding region [200, 295]. We have now also shown this with the -336 and -871 promoter region SNPs [21]. The genetic variation in *DC-SIGN* has been associated with protection against parental HIV-1 infection [200] and with the severity of dengue pathogenesis [295] and involves polymorphisms (i.e. -336A/G) restricted to the *DC-SIGN* promoter region. Taken together, all these studies support the view that it is the variation in the amount of DC-SIGN protein being produced that can influence infectious disease susceptibility, and not differences in the DC-SIGN protein itself or variation in its neck region.

This study presents the first investigation of the role of *L-SIGN* neck-region variation in susceptibility to tuberculosis. A number of studies have already explored possible correlations between *L-SIGN* neck-region variation and susceptibility to other infectious diseases [51, 171, 175, 231]. For example, the *L-SIGN* tandem-repeat 7/5 genotype has been associated with an increased protection against HIV-1 infection in high risk individuals [175]. However, this association remains controversial because a previous study failed to detect such an association [171]. In the context of HCV infection, a study comparing the frequency distribution of *L-SIGN* neck-region polymorphisms in a group of infected patients versus non-infected individuals failed to demonstrate any statistical difference between the two groups [231]. However, the same authors did observe an association between neck-region polymorphisms and individual HCV viral loads, and suggested that length variation in the *L-SIGN* neck region affects HCV replication efficacy. Finally, a study focusing on susceptibility to SARS infection has shown that individuals who are homozygous for *L-SIGN* neck region repeats are better protected against SARS infection [51]. Our results however, clearly indicate that the *L-SIGN* neck-region allele/genotype frequencies are not statistically different between tuberculosis patients and healthy controls in our cohort. However because of the low frequency of this polymorphism in our population, we cannot conclusively exclude this length variation in the *L-SIGN* neck region as a factor influencing tuberculosis susceptibility. A larger cohort would be required to confirm this.

In summary, the results show that the length of the neck regions of both *DC-SIGN* and *L-SIGN* are not associated with an increased or decreased host susceptibility to develop tuberculosis, at least in our cohort [21]. These data are in contrast with other disease association studies where the tandem-repeat polymorphisms of *DC-SIGN* and/or *L-SIGN* seem to contribute to different susceptibilities to HIV-1 and SARS infections, and to the HCV replication efficacy [51, 175, 177, 231]..

In conclusion, the significant association found for the *CD209* promoter variants together with their phylogenetic status and frequency distribution strongly suggests that the -871G and -336A alleles may reduce the risk of developing tuberculosis [21]. More generally, our results, together with those reporting association of *CD209* promoter variants with both HIV susceptibility and dengue pathogenesis [200, 295] suggest that variation in this lectin may be of crucial importance

in the outcome of a number of infections due to DC-SIGN interacting with pathogens. The low level of sequence variation indicates the conservation of an important gene product and the immune system is presumably more tolerant of fluctuation in transcription levels of DC-SIGN. Detailed *in vitro* and *in vivo* studies assessing the functional consequences of *CD209* variants on the quality of the host immune response against pathogens, including *M. tuberculosis*, are now required to eventually develop knowledge based and effective pathway targeted treatments.

■

Chapter 5

Re-sequencing of the SP110 gene and genotyping of polymorphisms.

The major part of these results was published:

Babb C, Keet EH, van Helden PD, Hoal EG (2007)
SP110 polymorphisms are not associated with pulmonary tuberculosis in a South African population.
Human Genetics 121:521-2

5.1 Introduction

As discussed in Chapter 1, the use of murine models has led to the identification of genes such as solute carrier family 11A member 1 (*SLC11A1*, formerly known as *NRAMP1*, natural resistance-associated macrophage protein 1), where association with human tuberculosis has been confirmed in a number of studies [132]. C3HeB/FeJ inbred mice are extremely susceptible to virulent *Mycobacterium tuberculosis* (*M. tuberculosis*) and develop marked lung pathology, which leads to their rapid death after infection [148, 158, 250]. The "super susceptibility to TB1" (*sst1*) locus in C3HeB/FeJ inbred mice has a specific effect on the progression of tuberculosis and was related to less efficient control of *M. tuberculosis* multiplication, primarily in the lungs, after infection [250]. It was recently reported that the intracellular pathogen resistance 1 (*Ipr1*) gene, within the *sst1* locus, conferred a measure of innate immunity to *M. tuberculosis* infection in mice [250]. Unlike *SLC11A1*, which affects resistance to bacillus Calmette-Guérin (BCG) infection in mice, *Ipr1* plays a major role in outcome of *M. tuberculosis* infection. *Ipr1* is therefore a good candidate gene to test human susceptibility to tuberculosis.

The human homologue of *Ipr1* is *SP110*, a nuclear body protein, which is believed to be involved in regulation of transcription by facilitating DNA-binding, as it contains a conserved SAND domain (Sp100, AIRE-1, NucP41/75, and DEAF-1) [38, 333]. The expression of both *Ipr1* and SP110 is regulated by interferons [146], indicating a role in immunity. The *SP110* gene is found on chromosome 2q, consists of 18 exons and has three transcripts (*a*, *b* and *c*). Transcript *SP110b* is the closest homologue to the predicted *Ipr1* protein [35]. The *SP110b* protein physically interacts with viral proteins such as Epstein-Barr virus SM protein and hepatitis C virus (HCV) core protein [236, 385]. In addition, it has recently been found that there is an association between mutations in the *SP110* gene and hepatic veno-occlusive disease with immunodeficiency, an autosomal recessive disorder, characterised by combined B and T cell immunodeficiency, which presents with multiple infections [283].

A transmission disequilibrium test (TDT) association study in families from the Gambia, Guinea Bissau and the Republic of Guinea, identified three SP110 gene polymorphisms that possibly influenced genetic susceptibility to tuberculosis [351]. The single nucleotide polymorphisms (SNPs) associated with tuberculosis susceptibility were rs2114592, rs3948464 and a novel SNP, *SP110*int10 (rs41547617). Association testing by TDT showed that the haplotype rs2114592 and rs3948464 (C/C) conferred susceptibility when all three populations were analyzed separately or together (all populations combined, 384 families, p=0.000005) [351]. Polymorphisms in this gene were tested in our South African Coloured (SAC) population (Chapter 2) in order to perform replication of the above study in a different population, which is always important in validating the association of a gene with a disease [15].

5.2 Results

5.2.1 Sequencing of the *SP110* gene

We sequenced the following regions of the SP110 gene in more than 11 random case-control samples; exons 4 to 8, 11, 14, 15, alternative 15, 17 and 18 including surrounding intronic regions. These regions contained either known polymorphisms or domains such as a SAND, zinc finger or bromo domain, or caused a splice variant. The results are summarised in Table 5.1 and the linkage disequilibrium (LD) plot is shown in Figure 5.1. Of the 46 SNPs investigated, 15 were novel and identified by re-sequencing of at least 11 individuals (22 chromosomes) bi-directionally (Appendix A1.4). All novel polymorphism were submitted to NCBI dbSNP and received rs numbers (Table 5.1). Of the known polymorphisms reported in previous studies and in public databases, 10 were not polymorphic in our population (Table 5.1). Exon 15 had no polymorphisms within the sequenced region. As can be seen in Figure 5.1 there are no definable blocks when using the parameter's confidence intervals, and no tagSNPs were identified. Heterozygosity was calculated with HaploView.

Although the D' was high (reported value in block), the LOD score was low (light shading of the block) (Figure 5.1). This is possibly because of the low polymorphism frequency and since few samples were sequenced. However, when LD for the eight SNPs investigated in over 750 samples was analysed, there was still low LD (Figure 5.2).

Table 5.1: Polymorphisms found through bidirectional sequencing of the *SP110* gene in the SAC population.

						Genotype Frequency				
Location in *SP110* gene	**dbSNP rs number**	**Base position**	**Major/ Minor allele**	**Novel**	**Number of Samples sequenced**	**MAF**	**Homo-zygote major allele**	**Hetero-zygote**	**Homo-zygote**	**Hetero-zygosity**
Exon 4	rs3177554	7529	C/T		15	0	1	0	0	0
	rs11556887	7578	C/T		23	0.217	0.696	0.174	0.13	0.34
	rs41552315#	7714	C/T	yes	23	0.13	0.826	0.087	0.087	0.227
Intron 4	rs6749386	7894	C/T		23	0	1	0	0	0
	rs41554115#	7993	T/C	yes	23	0.13	0.826	0.087	0.087	0.227
	rs41309096#	8101	C/T	yes	20	0.2	0.7	0.2	0.1	0.32
Exon 5	rs28930679	8142	C/T		20	0.025	0.95	0.05	0	0.049
	rs9061	8144	G/A		18	0.278	0.611	0.222	0.167	0.401
	rs1063154	8153	T/G		20	0	1	0	0	0
	rs1047254	8160	A/G		22	0	1	0	0	0
Intron 5	rs3769839	8629	A/G		5	0	1	0	0	0
	rs4973299	8637	T/C		11	0	1	0	0	0
	rs3769837	8679	A/G	yes	11	0.318	0.364	0.636	0	0.434
	rs3820974	8965	A/C		11	0.136	0.727	0.273	0	0.236

Table 5.1: Continued

Location in *SP110* gene	dbSNP rs number	Base position	Major/ Minor allele	Novel	Number of Samples sequenced	Genotype Frequency				
						MAF	Homo-zygote major allele	Hetero-zygote	Homo-zygote	Hetero-zygosity
Exon 6	rs3769838	9063	A/G		11	0	1	0	0	0
Intron 6	rs3769837	9222	A/C		11	0.364	0.273	0.727	0	0.463
	*rs2114592	10048	C/T		778	0.283	0.53	0.375	0.095	0.406
	rs41552220#	10246	T/C	yes	12	0.333	0.333	0.667	0	0.444
	rs41548221#	10350	C/T	yes	11	0.182	0.636	0.364	0	0.298
Exon 7	rs1129425	10579	A/G		12	0	1	0	0	0
	rs12467	10608	G/A		12	0	1	0	0	0
Intron 7	rs6719893	12013	A/G		20	0	1	0	0	0
	rs41553213#	12221	A/T	yes	20	0.325	0.4	0.55	0.05	0.439
	rs41543312#	12291	G/A	yes	20	0.1	0.8	0.2	0	0.18
Exon 8	rs1365776	12545	A/G		777	0.166	0.705	0.257	0.037	0.277
Intron10	rs41545715#		Indel	yes	798	0.085	0.83	0.17	0	0.156
	*rs41547617#	34384	A/G		30	0	1	0	0	0
Exon 11	*rs3948464	34539	C/T		763	0.267	0.547	0.374	0.08	0.391
Intron 11	rs41563112#	34660	A/C	yes	25	0.08	0.84	0.16	0	0.147
	rs41552512#	34667	G/C	yes	21	0.095	0.81	0.19	0	0.172
	rs2114591	34685	C/T		23	0.196	0.652	0.304	0.043	0.315
Exon 14	rs1804027	42978	T/C		742	0.302	0.495	0.407	0.098	0.422
Intron 14	rs41559116#	43130	T/C	yes	13	0.038	0.923	0.077	0	0.074
	rs41543719#	43315	C/G	yes	13	0.038	0.923	0.077	0	0.074
	rs1427285	43398	C/T		13	0.269	0.462	0.538	0	0.393
	rs10933327	43474	T/C		13	0.269	0.538	0.385	0.077	0.393
Exon 15 Alternative	rs35495464	47638	T/C		792	0.021	0.958	0.042	0	0.041
Intron 17	rs12995739	47899	A/C		12	0.083	0.833	0.167	0	0.153
Exon 17	rs13018234	48388	C/T		11	0.045	0.909	0.091	0	0.087
	rs3948463	48394	G/A		797	0.018	0.964	0.036	0	0.036
Intron 17	rs34034766	48486	A/G		11	0.091	0.818	0.182	0	0.165
	rs34897866	48500	T/C		11	0.091	0.818	0.182	0	0.165
	rs45567238	48722	A/T	yes	11	0.045	0.909	0.091	0	0.087
Intron 18	rs41541917#	48923	C/T	yes	792	0.064	0.878	0.117	0.005	0.119
	rs12988385	48943	A/G		11	0.091	0.818	0.182	0	0.165
	rs11690519	48957	G/A		11	0	1	0	0	0

# indicates new NCBI dbSNP rs numbers after submission by Babb *et al* [15]
* SNP found to be associated with tuberculosis susceptibility by Tosh *et al* [351]

Figure 5.1: LD plot for the *SP110* regions in all samples sequenced. The shading of a block indicates the LOD score (darker blocks represent a higher LOD score).

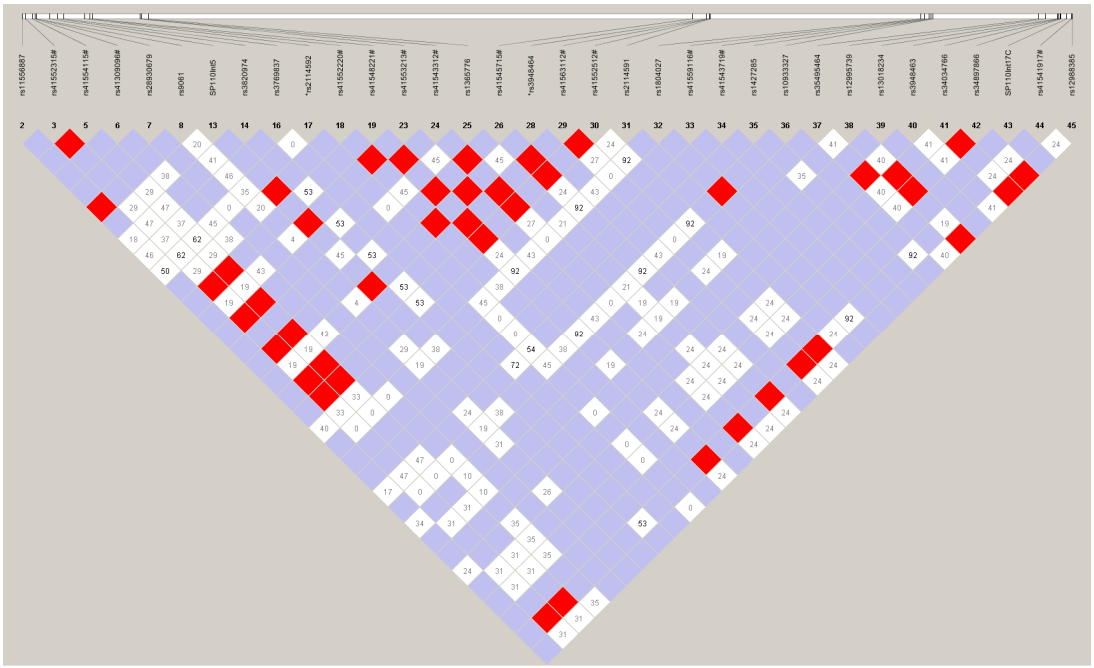
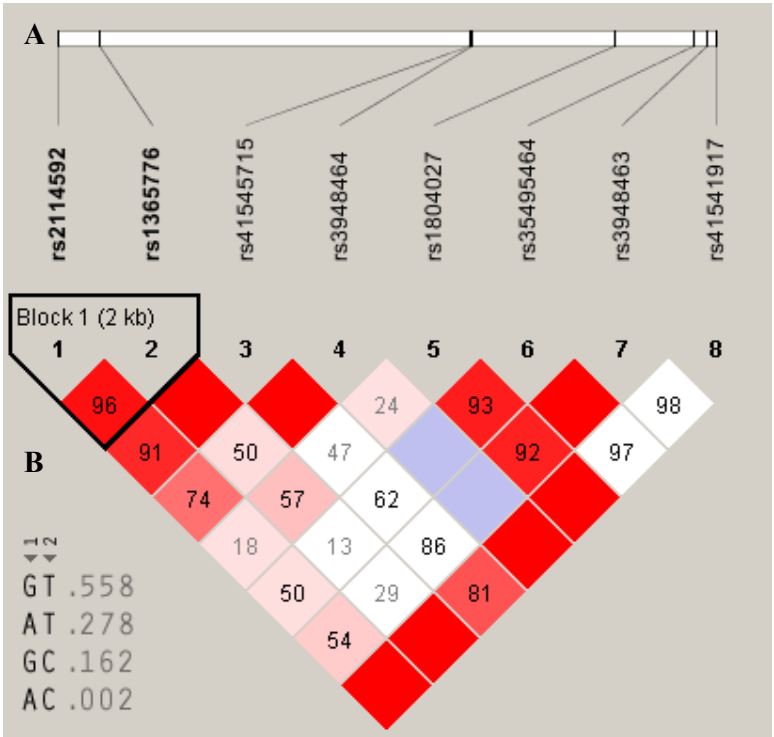


Figure 5.2: A: Linkage Disequilibrium plot for the eight polymorphisms investigated in gene *SP110* B: Block 1 (in A) haplotypes and their frequency, defined by confidence intervals in Haploview.



5.2.2 Genotyping of eight polymorphisms and case-control analysis

The SNPs that were chosen to be genotyped were SNPs that had previously been investigated and SNPs that spanned the gene. Eight SNPs were genotyped by SNaPshot ® Multiplex System (Appendix A1.5.1) in 381 tuberculosis cases and 417 healthy controls. The mean ages for the two groups were 37.1 and 33.1 years, respectively. The SNPs analysed included two novel intronic polymorphisms; one was an insertion/deletion (indel) of 4 base pairs (submitted to NCBI dbSNP, rs41545715) and the other was in Intron 18 (rs41541917) and had a heterozygosity value that indicated it would provide enough power with our sample size to detect a meaningful significant difference between cases and controls. In addition we genotyped two polymorphisms (rs2114592 and rs3948464) that were previously found to be associated with susceptibility to tuberculosis [351]. The third novel polymorphism, *SP110*int10 (rs41547617), that Tosh *et al* found to be associated with susceptibility to tuberculosis [351], was not found to be polymorphic in the SAC population after 30 samples were sequenced. The information on this polymorphism was submitted to the NCBI dbSNP and the SNP received the rs number rs41547617. As this SNP was found to be in LD with rs3948464 [351] it would not have provided additional information. The technique used to genotype the samples, SNaPshot, also restricts SNP selection because primers are designed to be adjacent to the polymorphism (Appendix A1.5.1). Interestingly, SNP rs1804027 in Exon 14 falls within the SAND domain and if it had been associated would have provided clues to a possible function. However, no significant associations were found between the *SP110* polymorphisms and susceptibility to tuberculosis (Table 5.2).

All polymorphisms were in Hardy-Weinberg equilibrium. This study had 95% confidence, 80% power to detect an OR of 1.6 for both the previously implicated SNPs.

Table 5.2: *SP110* genotype distributions in tuberculosis cases and healthy controls in the SAC population.

			Genotype frequencies								
	Allele		Cases			Controls			Cases vs Controls^		
Poly.	**1**	**2**	**11**	**12**	**22**	**11**	**12**	**22**	**p**	**OR (95% CI)**	
rs2114592*	C	T	0.526	0.379	0.095	0.533	0.372	0.095	0.886	0.971	(0.733-1.288)
rs1365776	A	G	0.710	0.253	0.037	0.701	0.262	0.037	0.813	1.046	(0.768-1.425)
rs41545715	GAAG	-	0.816	0.184	0.000	0.842	0.158	0.000	0.348	0.835	(0.577-1.209)
rs3948464*	C	T	0.551	0.373	0.077	0.543	0.374	0.083	0.828	1.033	(0.776-1.374)
rs1135791	T	C	0.500	0.407	0.093	0.490	0.407	0.103	0.825	1.041	(0.780-1.389)
rs3948463	G	A	0.950	0.050	0.000	0.976	0.024	0.000	0.060	0.472	(0.217-1.028)
rs41541917	C	T	0.869	0.123	0.008	0.886	0.112	0.002	0.516	0.855	(0.559-1.308)
rs35495464	T	C	0.947	0.053	0.000	0.969	0.031	0.000	0.583	0.156	(0.217-1.028)

^The homozygotes for the most frequent allele were compared with the sum of the heterozygote and homozygote for the rare allele. Comparing the homozygotes for the rare allele with the sum of the heterozygotes and homozygotes for the most frequent allele gave similar results.

* Polymorphism found to be associated with tuberculosis susceptibility by Tosh et al in West Africa [351]

5.2.3 Sequencing of the *SP110* gene in two disseminated BCG patients

In addition to the sequencing of case-control samples, the *SP110* gene (regions as outlined in section 5.2.1) was sequenced in two patients that suffered from disseminated BCG. Disseminated BCG is an extreme phenotype where shortly after BCG vaccination the child develops disease. It may be a rare event but the genetic, immunological and clinical study of the signalling defects can lead to invaluable insights in human control of intracellular pathogens. In addition, the identification of these unique defects should enable us to identify individuals at risk of an adverse, sometimes fatal, reaction to BCG vaccination.

The patients investigated are described in a recent publication [128]. Patient DNA sample ID 4338 is Patient ID 22 in the publication and Patient DNA sample ID 4339 is Patient ID 19 (Table 5.3). Neither was HIV positive, but Patient ID 4339 had a primary immunodeficiency (T-cell deficiency). Both were from the SAC population, had a BCG scar present, were unrelated and were adherent in terms of their therapy. Good adherence is defined as more than 80% of doses taken.

No novel *SP110* gene polymorphisms or mutations were identified in the disseminated BCG patients and there were no unique differences between them and the SAC samples sequenced (Table 5.4). Both the disseminated BCG samples are heterozygotes for a polymorphism in Exon 5, rs28930679, of which the minor allele frequency is 0.025 in the general population. The major alleles of three SP110 polymorphisms (rs2114592, rs3948464 and rs41547617) were associated with tuberculosis susceptibility in the West African populations. Of these, only one, rs2114592, was heterozygous in one of the BCG disseminated patients. The other two, rs41547616 and rs3948464 were both homozygous for the major allele in the BCG disseminated patients.

Table 5.3: Disseminated BCG patient information.

	4338	**4339**
Patient ID, reference [128]	22	19
Gender	M	F
Age (in months) at diagnosis (culture positive *M. tuberculosis* complex infection)	2	16
Mantoux TST finding (mm)	6	11
Revised paediatric disease classification	Regional and dual disease due to *M. tuberculosis*.	Distant disease.

Table 5.4: Comparison of the overall minor allele frequency (MAF) in *SP110* of the SAC case-control samples and the sequencing results for the disseminated BCG patients (4338 and 4339). * Polymorphisms found to be associated with tuberculosis susceptibility by Tosh *et al* (2006) [351] .

Location in SP110 gene	Polymorphism	Major/ Minor Allele	BCG-osis patient		MAF (SAC case-control)	Number of samples
			4338	4339		
Exon4	rs3177554	C/T	cc	cc	0	15
	rs11556887	C/T	cc	cc	0.217	23
	rs41552315	C/T	cc	cc	0.13	23
Intron 4	rs6749386	C/T	cc	cc	0	23
	rs41554115	T/C	tt	tt	0.13	23
	rs41309096	C/T	ct	tt	0.2	20
Exon 5	rs28930679	C/T	ct	ct	0.025	20
	rs9061	G/A	aa	aa	0.278	18
	rs1063154	T/G	tt	tt	0	20
	rs1047254	A/G	aa	aa	0	22
Intron 5	rs4973299	T/C	tt	tt	0	11
	SP110Int5	A/G	aa	ga	0.318	11
	rs3820974	A/C	cc	ca	0.136	11
Exon6	rs3769838	A/G	aa	aa	0	11
Intron6	rs3769837	A/C	aa	ca	0.364	11
	*rs2114592	C/T	cc	ct	0.283	778
	rs41552220	T/C	tt	ct	0.333	12
	rs41548221	C/T	cc	cc	0.182	11
Exon 7	rs1129425	A/G	aa	aa	0	12
	rs12467	G/A	gg	gg	0	12
Intron 7	rs6719893	A/G	aa	aa	0	20
	rs41553213	A/T	aa	at	0.325	20
	rs41543312	G/A	gg	gg	0.1	20
Exon8	rs1365776	A/G	aa	aa	0.166	777
Intron10	rs41545715	Indel	wt	wt	0.085	798
	*rs41547617	A/G	aa	aa	0	30
Exon11	*rs3948464	C/T	cc	cc	0.267	763
Intron 11	rs41563112	A/C	aa	aa	0.08	25
	rs41552512	G/C	gg	gg	0.095	21
	rs2114591	C/T	tt	ct	0.196	23
Exon 14	rs1804027	T/C	tt	ct	0.302	742
Intron 14	rs41559116	T/C	tt	tt	0.038	13
	rs41543719	C/G	gg	cg	0.038	13
	rs1427285	C/T	ct	ct	0.269	13
	rs10933327	T/C	tt	ct	0.269	13
Exon 15 Alt.	rs35495464	T/C	tt	tt	0.021	792
Intron 17	rs12995739	A/C	aa	aa	0.083	12
Exon 17	rs13018234	C/T	cc	cc	0.045	11
	rs3948463	G/A	gg	gg	0.018	797
Intron 17	rs34034766	A/G	aa	aa	0.091	11
	rs34897866	T/C	tt	tt	0.091	11
	SP110Int17C	A/T	aa	aa	0.045	11
Intron 18	rs41541917	C/T	cc	cc	0.064	792
	rs12988385	A/G	aa	aa	0.091	11
	rs11690519	G/A	gg	gg	0	11

5.3 Discussion

The *SP110* gene was investigated by sequencing exons and their surrounding intronic regions and determining the frequency of the polymorphisms found in tuberculosis cases and controls, as well as two patients with disseminated BCG. We intended to use haplotype tagging through Haploview to determine which SNPs would be most informative to genotype. However, for the number of samples sequenced, the frequency of the polymorphisms was too low to provide high enough confidence intervals to determine reliable haplotype tags. Allele frequencies were determined and used as a guide for the selection of the polymorphisms investigated.

In the SAC population, no significant association with tuberculosis susceptibility was found for any of the eight polymorphisms investigated (Table 5.2) [15]. Analysis of the haplotypes by Haploview also revealed no association with any of the inferred haplotypes. LD over the whole gene (Figure 5.1) appears to be low but this might be because of the low frequency of the polymorphisms and sequencing of too few samples. However, when the LD for the eight SNPs investigated in over 750 samples was analysed there was still low LD (Figure 5.2). A low LOD score was a result of the low frequency of the polymorphism.

Replication of the association with tuberculosis susceptibility reported by Tosh *et al* was not found in the SAC case-control study. Each of the three polymorphisms found to be associated with susceptibility to tuberculosis were genotyped by Tosh *et al* in a TDT study involving 298 to 373 families from three different West African nations, namely Guinea-Bissau, Republic of Guinea and the Gambia. While this investigation was underway, two large studies in a Ghanaian population [348] (over 990 cases, and 1000 controls) and a Russian cohort [335] (1912 cases and 2104 controls) found no association with *SP110* polymorphisms, also by means of a case-control analysis. Since a number of independent populations [15, 335, 348] and *SP110* variations spanning most of the gene, have now been investigated (Table 5.5), with only the West African TDT study indicating an association with tuberculosis [351], it is likely that this gene does not contribute significantly, if at all, to genetic susceptibility to tuberculosis in all human populations.

Table 5.5: Frequencies of the *SP110* polymorphisms in different population groups.

				Frequency of the major allele (number of samples tested) in different populations						
	Polymorphism	**Base position**	**Major Allele**	**SAC** [15]	**Ghana** [348]	**Gambian** [351]	**Republic of Guinea** [351]	**Guinea-Bissau** [351]	**Gambian, Guinea and Guinea-Bissua** [351]	**Russian** [335]
Exon4	rs3177554	7529	C	np (15)		0.99(390)				0.91(3894)
Exon4	rs11556887	7578	C	0.78(23)	0.99(2421)					0.87(3970)
Exon 5	rs28930679	8142	C	0.98(20)	0.80(2640)					
Exon 5	rs9061	8144	G	0.72(18)	0.94(2640)	0.96(358)				
Exon 5	rs1047254	8160	A	np (22)		np		np		
Intron 5	rs3820974	8965	A	0.86(11)	0.58(2119)	0.49(384)				
Exon6	rs3769838	9063	A	np (11)		np		np		
Intron6	*rs2114592	10048	C	0.72(778)		0.86(278)	0.81(144)	0.82(174)	0.84(596)	
Exon 7	rs1129425	10579	A	np (12)		np		np		
Exon 7	rs12467	10608	G	np (12)		np		np		
Exon8	rs1365776	12545	A	0.83(777)	0.98(2464)	0.95(378)				0.60(3975)
Intron10	*rs41547617	34384	A	np (30)		0.97(422)	0.99(150)	0.96(174)	0.97(746)	
Exon11	*rs3948464	34539	C	0.73(763)	0.86(2320)	0.84(398)	0.86(166)	0.84(166)	0.85(730)	0.90(3960)
Intron 11	rs2114591	34685	C	0.80(23)	0.54(2153)	0.58(410)				
Exon 14	rs1804027	42978	T	0.70(742)	0.81(2304)	0.80(360)				
Exon 15 Alternative	rs35495464	47638	T	0.98(792)	0.96(24)					
Exon 17	rs13018234	48388	C	0.95(11)	0.94(>2000)					
Exon 17	rs3948463	48394	G	0.98(797)		freq>99%		freq>99%		0.92(3951)
Intron 18	rs12988385	48943	A	0.91(11)						0.91(3950)
Intron 18	rs11690519	48957	G	np (11)						0.97(3774)

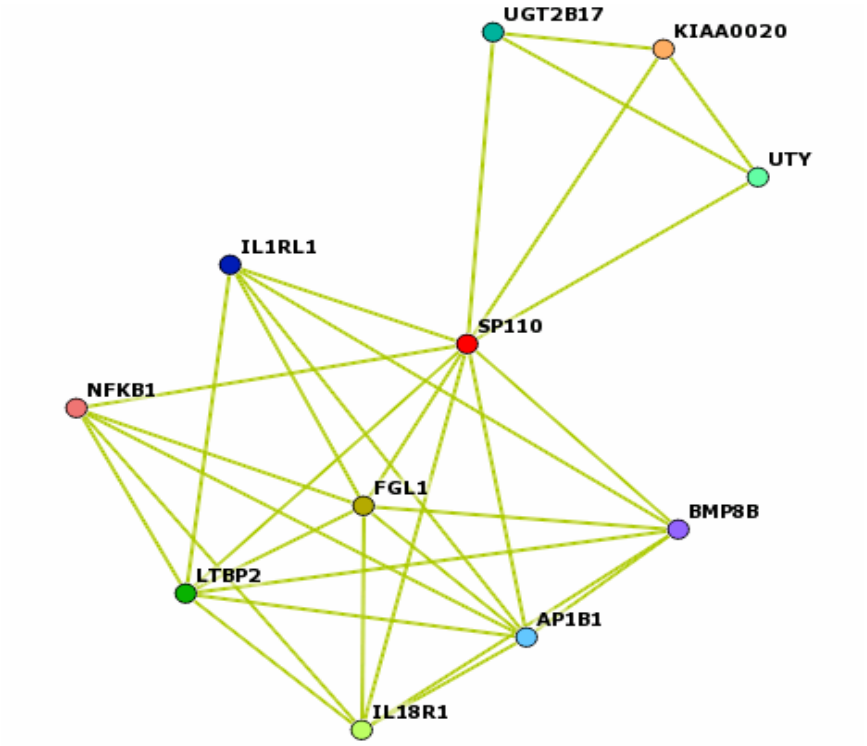
* polymorphism found to be associated with tuberculosis susceptibility by Tosh et al [351]
np, not polymorphic

In the West African TDT cohort, the major allele of the SNPs rs2114592, rs41547617 (*SP110*int10) and rs3948464 conferred susceptibility to tuberculosis. The two BCG-osis patients were nearly always homozygous for these major alleles, but this was to be expected given the allele frequencies of the three SNPs in the SAC population. This, unfortunately, does not provide useful information for BCG-osis susceptibility prediction. If the minor alleles had been associated with tuberculosis susceptibility and the BCG-osis patients were homozygous for the minor allele (i.e. the frequency of the allele was higher than expected) then this would have provided useful information for further investigation. The BCG-osis patients display an extreme phenotype and identifying a genetic association will lead to a better understanding of tuberculosis pathology. Further studies to identify polymorphisms in the IL-12/IL-23/IFNγ axis previously associated with BCG-osis patients [193, 254] should be done in these two patients.

Further investigation of the *SP110* gene and promoter polymorphisms could be done by association studies. It would also be interesting to investigate whether there is an association between these SNPs and time to sputum conversion in first time pulmonary tuberculosis patients, as was done with the *VDR* gene (Chapter 6). Finally, it is possible that *SP110* may confer protection against infection, which was not tested in this study (Chapter 7), as the majority of the SAC population is latently infected with *M. tuberculosis*.

The *SP110* gene was a promising candidate because of studies done in a mouse model, but was subsequently not confirmed in human case-control studies. This situation is unlike that with the *SLC11A1* gene [29], which was also identified through mouse studies of control of susceptibility to BCG inoculation [159, 375]. As mentioned earlier *SP110* was a more promising candidate, since in mice, its homologue *Ipr1* conferred innate immunity to *M. tuberculosis* infection. That this was not found in humans raises the question of whether mouse models are useful in identifying human genes involved in resistance to *M. tuberculosis*. Although *M. tuberculosis* is not a natural pathogen of mice, the mouse model has proved very helpful in understanding the immune response in tuberculosis [247]. Proteins of the signalling pathway upstream and downstream of the *SP110b* protein should not be excluded as candidates for susceptibility to tuberculosis, as different stages in the pathway may be affected and in turn may affect susceptibility to tuberculosis (Figure 5.3). The influence of the *SP110* gene might be minor and not detectable in an outbred human population, although it was detected in the inbred mouse strains. Alternatively there might be different pathways between the two species that confer resistance or susceptibility to tuberculosis.

Figure 5.3: Predicted Functional Partners with the *SP110b* protein according to the STRING website (http://string.embl.de//).
Note these are only text mining results, none are experimental.
http://string.embl.de/newstring_userdata/net_image_XxsXn8jrshUA.png accessed June 2007.



Your Input:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	Homology	Score
SP110	Sp110 nuclear body protein (Speckled 110 kDa) (Transcriptional coactivator Sp110) (Interferon-induced protein 41/75) (713 aa) *(Homo sapiens)*									

Predicted Functional Partners:

KIAA0020	Protein KIAA0020 (HBV X-transactivated gene 5 protein) (Minor histocompatibility antigen HA-8) (H							•		0.636
FGL1	Fibrinogen-like protein 1 precursor (Hepatocyte-derived fibrinogen- related protein 1) (HFREP-1) (H							•		0.615
IL18R1	Interleukin-18 receptor 1 precursor (IL1 receptor-related protein) (IL-1Rrp) (541 aa)							•		0.595
LTBP2	Latent transforming growth factor-beta-binding protein 2 precursor (LTBP-2) (1821 aa)							•		0.584
UTY	Ubiquitously transcribed Y chromosome tetratricopeptide repeat protein (Ubiquitously transcribed T							•		0.582
UGT2B17	UDP-glucuronosyltransferase 2B17 precursor (EC 2.4.1.17) (UDPGT) (C19- steroid specific UDP-glu							•		0.579
AP1B1	Adapter-related protein complex 1 beta 1 subunit (Beta-adaptin 1) (Adaptor protein complex AP-1 I							•		0.575
IL1RL1	Interleukin-1 receptor-like 1 precursor (ST2 protein) (556 aa)							•		0.572
BMP8B	Bone morphogenetic protein 8B precursor (BMP-8B) (BMP-8) (Osteogenic protein 2) (OP-2) (402 aa							•		0.540
NFKB1	Nuclear factor NF-kappa-B p105 subunit (DNA-binding factor KBF1) (EBP- 1) [Contains- Nuclear fac							•		0.534

■

Chapter 6

Vitamin D Receptor gene polymorphisms and sputum conversion time in first time pulmonary tuberculosis patients.

The major part of these results was published in:

Babb C, van der Merwe L, Beyers N, Pheiffer C, Walzl G, Duncan K, van Helden P, Hoal EG (2007)
Vitamin D receptor gene polymorphisms and sputum conversion time in pulmonary tuberculosis patients.
Tuberculosis 87, 295-302

6.1 Introduction

The role of Vitamin D and its receptor (VDR) in innate immunity, and specifically tuberculosis, is a subject of intense debate [37, 92, 169, 301, 391]. The treatment of tuberculosis in the earlier part of the 20th century frequently consisted of consuming cod liver oil, or heliotherapy (sun bathing), an important step in the synthesis of Vitamin D in the skin. Vitamin D is subsequently metabolised in the liver to its active steroid hormone 1,25 dihydroxyvitaminD_3 (1,25$(OH)_2D_3$) or calcitriol [83, 86, 169]. *In vivo* studies have shown that calcitriol suppresses growth of *M. tuberculosis* [73, 278] and is produced within granulomas by activated monocytes and macrophages [303]. The growth-suppressive effect may be facilitated by Toll-Like Receptors *in vivo* [178]. The nuclear hormone receptor for calcitriol is encoded by the *VDR* gene, which belongs to the family of trans-acting transcriptional regulator factors. The receptor regulates a variety of metabolic pathways such as those involved in the immune response, but the principle downstream targets are involved in mineral metabolism. The receptor is widely expressed in all cells of the immune system such as monocytes, macrophages, and dendritic cells (all antigen presenting cells), natural killer cells, T cells and B cells [204, 230, 282]. In addition, when appropriately stimulated by interferon gamma, calcitriol can be locally produced in monocytes [282, 331]. The importance of Vitamin D and VDR has been investigated in a variety of conditions such as diabetes [244, 251], cancer [44, 402], bone mass density [221, 344], Vitamin D dependent rickets type I [136, 152] and mycobacterial infection [37, 92, 110, 169, 301, 391].

Case-control studies to assess the importance of *VDR* polymorphisms in tuberculosis have produced varying results in different populations [180, 391]. Roth *et al* (2004) reported an association of *VDR TaqI* and *FokI* single nucleotide polymorphisms (SNP) with the time to sputum conversion in 78 pulmonary and extra-pulmonary tuberculosis patients [285], although no association with susceptibility to tuberculosis was found in a case-control analysis. They found that sputum culture conversion was significantly faster if cases had the *FokI* 'FF', or the *TaqI* 'Tt' genotype. This was the first report of tuberculosis treatment response being investigated in relation to a genetic component. Validation of this preliminary report of tuberculosis treatment response being investigated in relation to a genetic component is vital and should preferably be done in a different population.

We analysed three *VDR* polymorphisms, *FokI*, *ApaI* and *TaqI,* and a number of other clinical and demographic factors in a large, well characterised group of first time pulmonary tuberculosis cases from what is known as the Surrogate Marker (SM) study that was conducted over a 30 month period in the population discussed in Chapter 2, the South African Coloured (SAC). The SM cohort is independent of all the previous case-control studies discussed in this thesis and is a unique cohort of patients. A large number of patients were followed up with regular clinic visits [39, 378] and blood samples were available for DNA analysis of the cohort.

VDR Polymorphisms

The *VDR* gene spans 79kb, has 11 exons and is found on chromosome 12q13.11. Four *VDR* gene polymorphisms have been extensively studied in a variety of diseases. Three polymorphisms

were analysed in this cohort because two of the four are in strong linkage disequilibrium (LD) in a number of studies and in a variety of populations; in other words the SNP is tagged.

The *FokI* (rs10735810) polymorphism is a functional polymorphism [379] and forms a new start site / initiation codon (ATG to ACG) at the junction of intron 1 and exon 2. Although its role still remains unclear, it has been shown that the *FokI* 'ff' polymorphism has increased expression [232, 357]. *FokI* is not known to be in LD with any other known *VDR* polymorphism [232, 356].

ApaI (rs7975232) a G (*ApaI* 'a') to T (*ApaI* 'A') change in intron VIII, 3' region was analysed because it appears to form an important haplotype [232], in particular with *FokI*. Bornman *et al* [37], showed significantly more transmission of the haplotype combination 'FA' to tuberculosis affected offspring (χ^2 =11.621, P=0.0007, 1 degree of freedom).

TaqI (rs731236) is a silent polymorphism with a base T (*TaqI* T) to base C (*TaqI* t) change, located in exon IX, in the 3' region. Roth *et al* [285] found an association with the *TaqI* Tt genotype and response to treatment of tuberculosis. There is strong linkage between *ApaI* and *TaqI*, yet they are not predictive of each other. They are separated by 80 base pairs and are in LD with a polyA variable number of tandem repeats (VNTR) in exon 9 which results in either a long or short allele, where the former is more transcriptionally active [126, 387].

BsmI (rs1544410) is a T/C SNP in intron VIII at the 3' end and has frequently been analysed. This polymorphism is in strong LD with *TaqI* (the 'T' and 'b' genotypes) and was not analysed in this study [285, 344, 391].

These *VDR* gene polymorphisms have been suggested by Roth *et al* as having potentially for being involved in the response of tuberculosis patients to treatment, which requires validation in a different population. In this study, we have extended the number of variables and examined their effect on response during tuberculosis chemotherapy in a large, well-characterised cohort of our SAC population.

6.2 Methods

6.2.1 Surrogate Marker Patient enrolment and surveillance

The tuberculosis cases used in this chapter were distinct from those in the rest of the thesis. All tuberculosis cases were enrolled in a longitudinal study to evaluate pulmonary tuberculosis cases during and after treatment; known as the SM study [39, 378]. All information and data was accessed from version 4 of the SM database. Informed consent was obtained from all subjects. Supplementary documentation for the SM database is in Appendix 2. Cases were all first-time pulmonary tuberculosis patients, not non-tuberculous mycobacteria (NTM) infected, HIV negative, 18-65 years old, not pregnant, had no chronic diseases, not multi-drug resistant tuberculosis (MDR-TB) and received direct observed treatment short-course (DOTS). All cases were diagnosed by having a positive sputum sample on auramine staining, a second positive

sputum sample or a posterior anterior and lateral chest radiograph typical of pulmonary tuberculosis. Blood was collected at diagnosis and total peripheral white blood cell (WBC) counts, absolute neutrophil, monocyte, and lymphocyte numbers were measured.

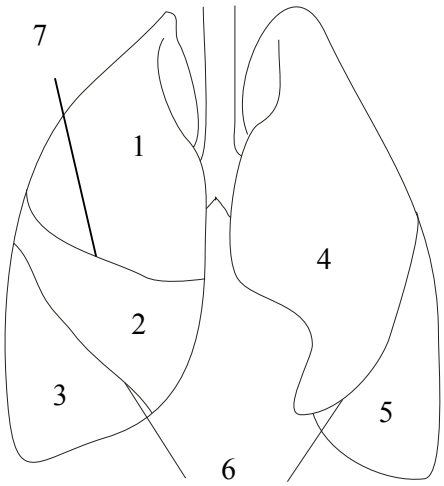
Cases were followed up with regular clinic visits from diagnosis (day 0), including day 1 and day 2, then weekly visits until month 2 (day 55) followed by monthly visits until month 6 (day 182) and subsequent visits in month 9 (day 273) and month 12 (day 364). Patients were followed for a total of 30 months, to detect possible second episodes of disease. However, for our purposes, information for the first episode of tuberculosis disease only was used.

For research purposes, sputum was collected at all time points for smear microscopy using Ziehl-Neelsen (ZN) staining and for *M. tuberculosis* culture using the BACTEC method (Beckton-Dickinson, USA). Restriction fragment length polymorphism (RFLP) of the *M. tuberculosis* strain and isoniazid and rifampicin susceptibility testing was performed. Serum was not available for the majority of samples, therefore the Vitamin D levels were not measured.

At enrolment to the study, a questionnaire was used to collect data on education level (completion of high school, yes/no), smoking of tobacco or cigarettes (yes/no), cannabis use (yes/no), alcohol consumption (yes/no) and limited data on income, measured as either above or below ZAR1000 / ± £80/ ±$150 per month. The following data was collected for all cases: age, gender, number of self reported tuberculosis symptoms (Cough, Coughing Blood, Night sweats, Chest pain, Back Pain, Deep breathing, Chest Tight, Fever, Weight Loss) at first clinical examination, weight, height, body mass index, and Mantoux skin test size at diagnosis and month 2. The chest radiographs were interpreted based on a previous method [309], by a single clinician who used a standardized reading form to evaluate the number of cavities visible on the chest radiograph, and the extent of pulmonary involvement (<right upper lobe (RUL), =RUL, >RUL and > 1 lung) as an indication of disease severity (Figure 6.1). Data on the number of cavities present was also available but was not as informative as the extent measured by RUL, as it does not take the actual size of the cavity into consideration.

Figure 6.1: Diagram of the lungs showing the lobes. The right upper lobe (RUL) was used as a reference for the extent of pulmonary involvement for the cases used in the Surrogate Marker study.

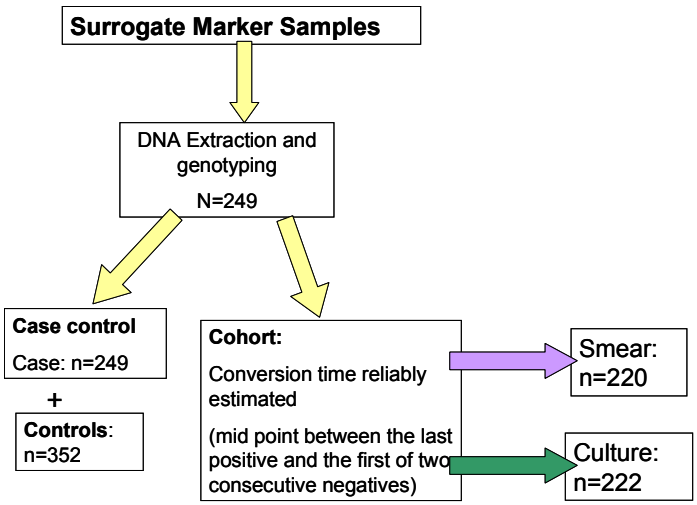
1. Right upper lobe (RUL)
2. Right middle lobe
3. Right lower lobe
4. Left upper lobe
5. Left lower lobe
6. Oblique fissure
7. Horizontal fissure



6.2.2 Case-control samples

The 249 tuberculosis cases were those enrolled in the longitudinal SM study as first time pulmonary tuberculosis patients. The 352 healthy controls were taken from the "EileenTB" database as discussed in Appendix 1 (Figure 6.2). Controls were older than 17 years. The average age of males in the cases was 37 years, and the female cases 33 years. In controls the average age of the males was 30 years and in female controls 31 years. Males formed 58% of the cases and 15% of the controls.

Figure 6.2: Samples that were genotyped for the *VDR* polymorphisms *FokI*, *ApaI* and *TaqI*.



6.2.3 Cohort with known sputum conversion times

A cohort from the longitudinal SM study (249 cases, section 6.2.1) was selected for conversion time analysis. Inclusion required the case to have a conversion time for smear and/or culture. Defaulters (patients who did not take the prescribed medication for at least 2 consecutive months) and cases that transferred outside of the study area were excluded if there was no conversion time available before this event. A conversion time in days (from positive to negative *M. tuberculosis* in sputum) was estimated from the day of diagnosis (day 0) and a date midway between the last positive result and the first of two successive negative results [285]. This resulted in a total of 220 cases with smear conversion times and 222 cases with culture conversion times (Figure 6.2). When more than 91 days had elapsed between the last positive and the first of two successive negative results, samples were censored for analysis purposes and the last positive day was used as the day of conversion.

The conversion time was estimated from information provided from the SM database. The smear and culture sputum results were available in the 'resSputum' table in the SM database. This table listed the results according to the 'WhenID' (Table 6.1). The 'WhenID' was based on the number of days of an expected clinical visit from the diagnosis. The clinical visit days were predetermined

and part of the SM study design. However, use of the WhenIDs for the construction of survival curves and Cox regression analysis the resulted in clustering of results on the particular 'WhenID' days. To have more accurate and reliable results and the actual visit date to the clinic was therefore used.

Table 6.1: The look up table: 'TestInterval Table' in the SM database which lists the WhenID and has a description of all the clinical visits when samples were collected.

WhenID	When (days)	Interval Description
1	0	Diagnosis
20	1	Day 1
21	3	Day 3
2	7	Week 1
3	14	Week 2
4	21	Week 3
5	28	Week 4
6	35	Week 5
7	42	Week 6
8	49	Week 7
9	56	Week 8 (month 2)
23	59	Week 8 + 3 days
11	91	Month 3 (week 13)
12	126	Month 4 (week 18)
13	154	Month 5
14	182	Month 6 (week 26)*
24	245	Month 8
15	273	Month 9
16	364	Month 12
17	546	Month 18
18	728	Month 24
19	910	Month 30
22	0	Other

*Following standard DOTS, medication was stopped at Month 6.

The 'WhenID' of the last positive sample for a patient was identified in the 'resSputum' table, which also has the field 'VisitDate'. The visit date lists the calendar date of a clinic visit. The visit date of the diagnosis ('WhenID' 1), the last positive and the first date of two consecutive negative samples were extracted. This was done for both smear and culture outcomes. The actual number of days from diagnosis was calculated with the use of Microsoft Excel. This resulted in the conversion time not falling exactly on a particular 'WhenID' day, giving a more accurate conversion day. The true date was on average one or two days from the expected 'WhenID' visit day.

The SM database was also carefully checked for samples that were possibly NTM samples (which was frequently mentioned in the comments column and not correctly coded with 999/888 in the result field). This was corrected in the next version of the SM database.

Once the conversion day was established, patients could be divided into slow and fast responders for both smear and culture conversion time. Fast responders were defined as cases that converted to negativity before day 55 (month 2) and slow responders on or after day 55. Month 2 was chosen as the cutoff as this corresponds to the International Union Against Tuberculosis and Lung Disease recommendations for the time at which sputum conversion is determined in order to establish treatment effectiveness.

6.3 Statistical analysis

A number of statistical analyses were done with the data from the study using SM samples. This was because this was not only an association study but an investigation of variables affecting the conversion time from a positive *M. tuberculosis* sputum sample to negative in both culture and smear. There were a large number of factors available that could be risk predictors or possible confounders for time-to-sputum conversion. Each analysis will be discussed in the relevant order: Fast versus slow, survival curve, univariate and then multivariate analysis.

Case-control

Genotype, allele and diplotype frequencies of the polymorphisms were calculated. As unrelated cases and controls were analysed, haplotypes could only be inferred using R (done by L. van der Merwe) and Haploview (done by C. Babb) (Appendix A1.6.3). However, the diplotype is a true representation of the genotype and was constructed by combining the observed genotypes in the order in which they occur in the gene: *FokI*, *ApaI* and *TaqI* and their association with tuberculosis was analysed by logistic regression. Hardy-Weinberg equilibrium and LD were determined. A total of 249 cases and 352 controls were analysed to investigate associations with susceptibility to tuberculosis using Fisher's exact test.

Cohort

Analyses were done in R: A language and environment for statistical computing, Base R [266], as well as packages survival [345], genetics [381], and haplo.stats [312] by Dr. L. van der Merwe, a bio-statistician, specialised in genetic statistics, who conducted the analysis of the univariate and multivariate analysis and assisted with the interpretation of results. The genotypes of the fast and slow responders to treatment were compared by a χ^2 test using GraphPad Prism (version 5). In the construction of the Kaplan-Meier survival curves, the cases are subgrouped according to their genotype (or whichever factor is investigated). The survival curve demonstrates the decreasing proportion of positive sputum samples in the cohort, where Day 0 is the day of diagnosis and initiation of treatment. Analysis is done by (Mantel-Cox) log rank, where differences in the curves are assessed. Kaplan-Meier survival curves were constructed in Prism, where a significant p value (<0.05) indicates significant differences between the curves. The use of Kaplan-Meier survival curves possibly indicated to Roth *et al* that there was an association with the *VDR* polymorphism and time to sputum conversion in their tuberculosis patients. However, Kaplan-Meier survival curves are not conclusive and further statistical analyses were required, including univariate analyses to identify which factors could be possible confounders. Variables with significant p-

values could then be used in the multivariate analysis, which provides a more statistically accurate outcome of any association, by taking the confounders into consideration.

An individual, age and gender adjusted, Cox regression analysis (univariate) was done for the variables separately for both smear and culture conversion times using R.

A multivariate analysis was done using R by means of the selection of an optimal Cox model for each of the methods (smear and culture) for measuring conversion times. A set of candidate variables were identified from which the optimal selection was made. They included those that have previously been found to be important for conversion in tuberculosis by Roth *et al* [285] i.e. high school completion, *TaqI* and *FokI*. Isoniazid mono-resistance was investigated instead, as no rifampicin mono-resistance was seen in this cohort (MDR-TB, a previously noted predictor, was one of the exclusion criteria for our cases). We also added variables that were significantly associated ($p<0.050$) with conversion time in smear or culture according to the age and gender adjusted individual Cox analyses (univariate). A basic set of variables consisting of age, gender and the genotypes were included in the optimal model irrespective of their effect. The variables included in the final models were selected with a backwards procedure. With the backwards procedure, one starts with a Cox model containing all the candidate factors. Factors are discarded one at a time, until the model with the lowest possible Akaike information criterion (AIC) is reached. The AIC is a measurement that describes how well the model fits the data: the lower the AIC value the better the fit of the model. The procedure discards factors until the lowest possible AIC is reached. The factors remaining indicate the best combination of factors that describes the observed conversion rates.

Some individuals (12%) had the value for a single variable missing and a few individuals had two or a maximum of three data points missing, which were mainly blood measurement variables. Following advice from the statistician, missing values were replaced with the mean (numerical values) or median (categorical variables) of all the values for the individual's diplotype, as it was assumed that the missing values were missing at random for that diplotype.

6.4 Results

6.4.1 Case-control

For the *VDR* polymorphisms a lower case ('f', 'a' or 't') was used to indicate the presence of an endonuclease site (Appendix A1.5.2). In the controls, there was strong LD between *ApaI* and *TaqI* (D' = 0.999). *FokI* was not linked with either *TaqI* or *ApaI*, (D'=0.048 and D'=0.005, respectively). All polymorphisms were in Hardy Weinberg equilibrium.

No association was found between genotype or alleles for the individual *VDR* polymorphisms and the presence or absence of pulmonary tuberculosis (Table 6.2). A non-significant association was seen between the inferred *FokI-ApaI-TaqI* haplotype and tuberculosis (global p value=0.078), with the *FokI-ApaI-TaqI* 'FaT' haplotype overrepresented in controls (p=0.063) and the 'FAT' haplotype more frequent in cases (p=0.062) (Table 6.3). Of the 27 possible diplotypes only 18

were observed. There were no significant differences between the diplotype frequencies in cases and controls after adjusting for age and gender using logistic regression (p=0.24) (Table 6.4).

Table 6.2: Case-control analysis of *VDR* genotype and pulmonary tuberculosis in the SAC population.

	Allele		Genotype frequencies								
			Control (n=352)			Cases (n=249)			Cases vs Controls ^		
	1	2	11	12	22	11	12	22	p	OR	(95% CI)
FokI	f	F	6	37	58	5	42	53	0.279	1.21	(0.8711-1.674)
ApaI	a	A	18	49	33	16	43	41	0.058	0.72	(0.5143-1.009)
TaqI	t	T	6	40	54	8	38	55	0.934	0.97	(0.7034-1.350)

^The homozygotes for the most frequent allele were compared with the sum of the heterozygote and homozygote for the rare allele.

Table 6.3: Two and three locus inferred haplotypes for the *VDR* gene polymorphisms *FokI*, *ApaI* and *TaqI* in pulmonary tuberculosis cases and controls in the SAC population.

Haplotype:	% of Controls (n=352)	% of Cases (n=249)	P value	Global p value
FokI, ApaI, TaqI				
faT	10	11	0.913	
fAt	6	9	0.117	
fAT	8	6	0.826	
FaT	32	27	0.063	
FAt	20	17	0.386	
FAT	24	30	0.062	0.078
ApaI, TaqI				
aT	43	38	0.099	
At	26	27	0.896	
AT	31	36	0.111	0.206
FokI, ApaI				
fa	10	11	0.866	
fA	14	15	0.327	
Fa	32	27	0.069	
FA	45	44	0.285	0.279
FokI, TaqI				
ft	6	10	0.107	
fT	18	17	0.95	
Ft	20	17	0.388	
FT	56	57	0.947	0.247

Note: Haplotypes for the polymorphisms were inferred and tested for association with pulmonary tuberculosis using the R package: haplo.stats.

Table 6.4: Observed diplotypes of *VDR* gene polymorphisms *FokI, ApaI* and *TaqI*, in pulmonary tuberculosis cases and healthy controls from the SAC population.

Diplotype	**Controls n (%)**	**Cases n (%)**	**Total n (%)**
ffaaTT	3 (1)	2 (1)	5 (1)
ffAaTt	5 (1)	5 (1)	10 (2)
ffAaTT	5 (1)	0 (0)	5 (1)
ffAAtt	2 (1)	3 (1)	5 (1)
ffAATt	4 (1)	2 (1)	6 (1)
ffAATT	1 (<1)	1 (<1)	2(<1)
FfaaTT	25 (7)	17 (7)	42 (7)
FfAaTt	23 (7)	25 (10)	48 (8)
FfAaTT	39 (11)	27 (11)	66 (11)
FfAAtt	8 (2)	7 (3)	15 (2)
FfAATt	21 (6)	17 (7)	38 (6)
FfAATT	13 (4)	11 (4)	24 (4)
FFaaTT	35 (10)	21 (8)	56 (9)
FFAaTt	50 (14)	20 (8)	70 (12)
FFAaTT	51 (14)	31 (12)	82 (14)
FFAAtt	12 (3)	9 (4)	21 (3)
FFAATt	37 (11)	25 (10)	62 (10)
FFAATT	18 (5)	26 (10)	44 (7)
Total n	352	249	601

Note: Logistic regression analysis of cases and controls, adjusting for age and gender, gave a p value =0.24.

6.4.2 Cohort and Sputum Conversion times

A variety of basic demographic information was available for the SM cohort. Some is summarized in Table 6.5.

Table 6.5: Basic demographic information from the Surrogate Marker patient cohort.

		Count	Frequency
Age (years)	18-20	8	4
	21-30	84	37
	31-40	64	28
	41-50	45	20
	>51	25	11
Gender	Males	132	58
	Females	94	42
Drug sensitivity to isoniazid	Yes	23	10
	No	199	88
	Unknown	4	2
Education	None	5	2
	<Std 3	53	23
	Std 4-7	105	46
	Std 8-10	50	22
	Diploma	3	1
	Unknown	10	4
Income Self In SA Rands per month (pm)	None	16	7
	<R250 pm	22	10
	R251-500 pm	10	4
	R501-1000 pm	18	8
	R 1001-1500 pm	10	4
	R1501-2000 pm	12	5
	>2001 pm	6	3
	Unknown	132	58
Drink Alcohol	Yes	174	77
	No	36	16
	Unknown	16	7
Smoke	Yes	202	89
	No	12	5
	Unknown	12	5
Died (none of tuberculosis)		5	2
Age (Years)	Average Age	35.3	
	Average Age Males	37	
	Average Age Females	33.0	
BMI	Average male	18.1	
	Average female	18.63	
Height (m)	Average male	1.7	
	Average female	1.6	

Time to sputum conversion was used as a measure of time to recovery. The median conversion times for both smear and culture are in Table 6.6.

Table 6.6: Median and average sputum conversion days for smear and culture results in the SAC cohort of first time tuberculosis patients.

	Smear		Culture	
	Median days	**Average days**	**Median days**	**Average days**
Overall	41	44.1	71.5	67.5
Males	41	40.2	44	59.7
Females	41	47.0	72	73.0
FokI FF	41	41.2	61	63.4
FokI Ff	42	48.6	71.5	72.3
FokI ff	41	36.4	71	67.5
ApaI AA	42	45.9	71	68.7
ApaI Aa	41	40.0	48.5	64.2
ApaI aa	41	50.8	72.5	74.2
TaqI TT	41	44.3	71.5	70.1
TaqI Tt	41	40.8	61	63.1
TaqI tt	43	61.1	69.5	70.6

6.4.2.1 Fast/slow sputum conversion time analysis of the VDR genotype

When the cases were divided into either fast (n=160) or slow (n=60) responders according to smear conversion before/after day 55, i.e. month 2 (Table 6.7), there was a strong trend to slower conversion in the 'F'-containing genotypes of *FokI* (p=0.055). Dividing cases into fast (n=96) and slow (n=126) responders with respect to culture conversion before or after and including day 55 showed a significant association with the *ApaI* genotypes (p=0.029). Fewer fast responders (10%) had an *ApaI* 'aa' genotype than slow responders (23%). None of the diplotypes were significantly associated with a fast or slow response.

Table 6.7: *VDR* genotypes of fast versus slow responders for tuberculosis treatment (at day 55) for the SAC cohort.

	Smear (n=220)		Culture (n=222)	
Polymorphism, genotype	**Fast responders§ n (%)**	**Slow responders† n (%)**	**Fast responders§ n (%)**	**Slow responders† n (%)**
FokI				
ff	10 (6)	1 (2)	5 (5)	6 (5)
Ff	61 (38)	33 (55)	38 (40)	58 (46)
FF	89 (56)	26 (43)	53 (55)	62 (49)
p value‡		p=0.055		p=0.649
ApaI				
aa	28 (18)	11 (18)	10 (10)	29 (23)
Aa	67 (42)	24 (40)	47 (49)	46 (37)
AA	65 (41)	25 (42)	39 (41)	51 (40)
p value‡		p= 0.959		p= 0.029
TaqI				
tt	9 (6)	6 (10)	7 (7)	8 (6)
Tt	63 (39)	21 (35)	40 (42)	45 (36)
TT	88 (55)	33 (55)	49 (51)	73 (58)
p value‡		p= 0.490		p= 0.573

§ Pulmonary tuberculosis cases that responded to treatment before day 55 after treatment initiation

† Pulmonary tuberculosis cases that responded to treatment on or after day 55 after treatment initiation

‡ A Chi-square analysis was done.

In the Peruvian study, Roth *et al* used the much earlier time point of month 1. It is unclear why they chose this particular time point but, by using it, associations with the *VDR* polymorphism *FokI* (p=0.02) and *TaqI* (p=0.012) were found. Analysing our results at the time point of 30 days did not show an association between *VDR* genotype and time to conversion (Table 6.8). The difference between the studies might be due to the culture method, namely MODS versus BACTEC.

Table 6.8: *VDR* genotype and fast response to tuberculosis treatment at 30 days (1 month) in the Peruvian and SAC populations.

		Sputum conversion* at month 1 (30 days after treatment initiation)							
Population:		**Peruvian**		**SAC**		**Peruvian**		**SAC**	
		Culture MODS		**Culture BACTEC**		**Smear Auramine stained**		**Smear Ziehl Neelsen**	
		n=78		n=222		n=78		n=220	
FokI	FF	0.78	(7/9) [a]	0.14	(16/115)	1	(9/9) [b]	0.43	(50/115)
	Ff	0.41	(9/22)	0.10	(10/96)	0.64	(14/22)	0.36	(34/94)
	ff	0.30	(14/47)	0.09	(1/11)	0.62	(29/47)	0.36	(4/11)
TaqI	TT	0.34	(24/71) [c]	0.11	(13/122)	0.66	(47/71)	0.43	(52/121)
	Tt	0.86	(6/7)	0.13	(11/85)	0.71	(5/7)	0.39	(33/84)
	tt	0	0	0.20	(3/15)	0	0	0.20	(3/15)
ApaI	AA	-	-	0.12	(11/90)	-	-	0.33	(30/90)
	Aa	-	-	0.13	(12/93)	-	-	0.46	(42/91)
	aa	-	-	0.10	(4/39)	-	-	0.41	(16/39)

*Sputum conversion is expressed as percentage of patients with a negative sputum result per genotype, where the numbers that are negative are divided by the total number with the given genotype.

[a] Significant difference between FF and either Ff or ff (p=0.023)

[b] Significant difference between FF and either Ff or ff (p=0.025)

[c] Significant difference between Tt and TT (p=0.012)

In addition, we investigated cases that had a fast or slow conversion time at day 30 (n=150) or day 55 (n=149) according to both culture and smear conversion time (Table 6.9). In other words, both their smear and culture results were negative at day 30 or day 55 for a fast conversion time. There was no significant association in these cases between their genotypes or haplotypes and time to sputum conversion.

Table 6.9: Pulmonary tuberculosis cases having a concordant culture and smear conversion time at month 1 (day 30) or month 2 (day 55).

		Smear and culture conversion day					
***VDR* Polymorphism**		**<30**	**≥30**	**p-value**	**<55**	**≥55**	**p-value**
FokI	FF	12	61		50	24	
	Ff	9	60		36	33	
	ff	1	7	0.836	5	1	0.088
ApaI	AA	8	57		37	24	
	Aa	10	48		44	23	
	aa	4	23	0.742	10	11	0.333
TaqI	TT	10	67		47	32	
	Tt	10	50		38	21	
	tt	2	11	0.831	6	5	0.757

6.4.2.2 Kaplan-Meier survival curves and Log rank analysis

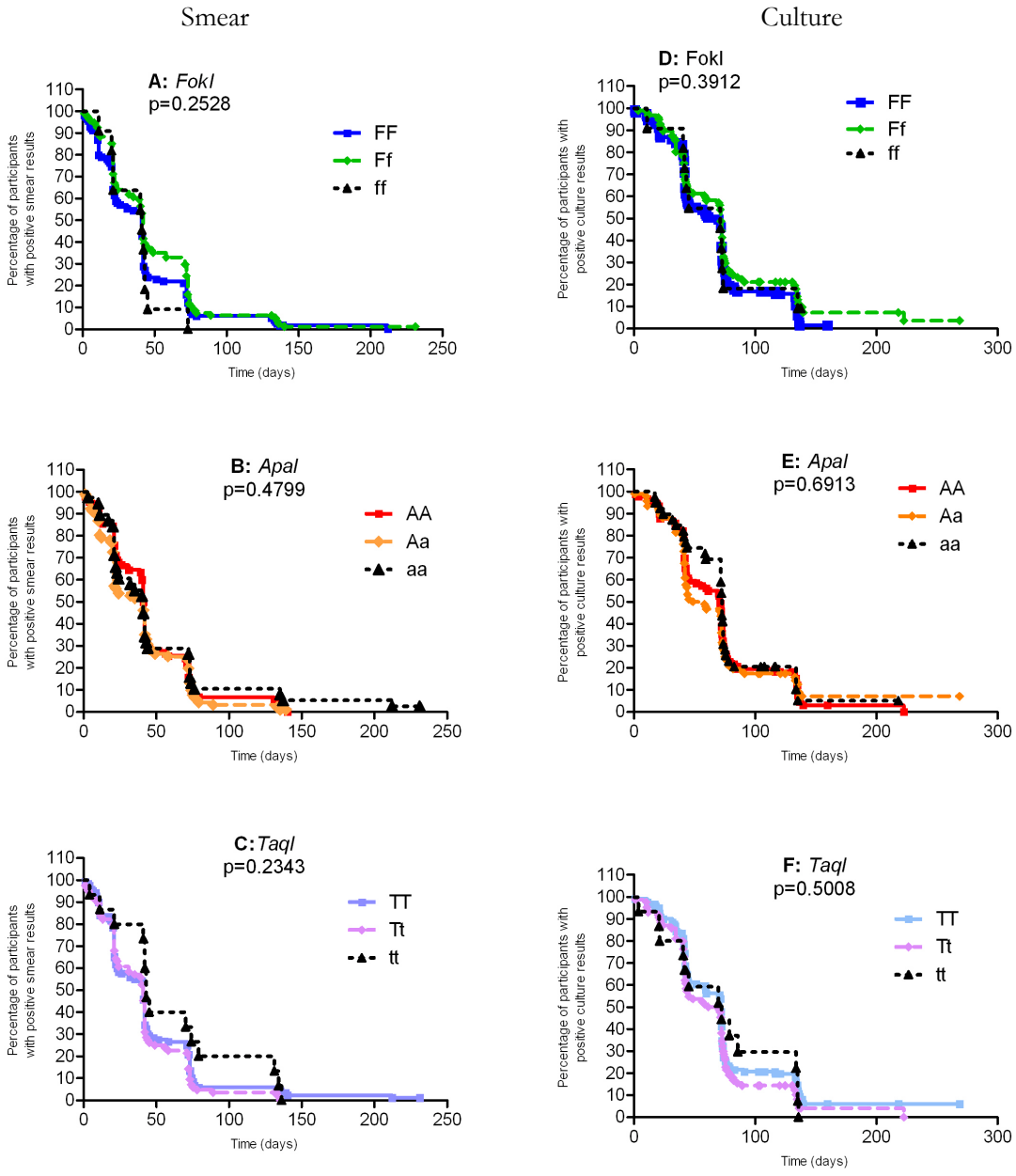
Construction of Kaplan-Meier survival curves allows for visualization of the conversion time and differences in the curves are assessed by log rank (Mantel-Cox) analysis. All variables had a Kaplan-Meier curve constructed and were assessed by log rank (Table 6.10). No significant differences were detected with the Kaplan-Meier survival curves for conversion time with the different genotypes (Figure 6.3).

Table 6.10: Kaplan-Meier (Mantel-Cox) log rank test results for *VDR* polymorphisms genotyped in a SAC population assessing smear and culture conversion times to negative smear.

	p values*	
Variables	**Smear**	**Culture**
Gender	0.148	**0.008**
Smokers: yes versus no	**0.018**	0.202
Isoniazid sensitive only, smokers versus non smokers	**0.003**	0.598
FokI	0.193	0.333
ApaI	0.501	0.652
TaqI	0.194	0.516
FokI; Combined (ff versus Ff and FF)	0.169	0.142
TaqI; Combined (tt versus Tt and TT)	0.831	0.342
FokI; Smokers only	0.284	0.273
ApaI; Smokers only	0.491	0.811
TaqI; Smokers only	**0.030**	0.306
FokI; Isoniazid sensitive cases only	0.255	0.308
ApaI; Isoniazid sensitive cases only	0.427	0.545
TaqI; Isoniazid sensitive cases only	0.213	0.492
FokI; Isoniazid sensitive and smokers	0.278	0.216
ApaI; Isoniazid sensitive and smokers	0.413	0.720
TaqI; Isoniazid sensitive and smokers	**0.045**	0.325
Number of cavities (1,2,3,4,>4)	**0.036**	0.279
Extent of TB at Diagnosis (<RUL,RUL, >RUL, >1 lung)	**0.003**	**0.005**
Absolute neutrophil number (40-60, 60.1-70, 70.1-80, 80.1-92)	**0.046**	**0.043**
Absolute lymphocyte number (4-9, 9.1-12.0, 12.1-15, 15.1-18, 18.1-21, 21.1-24, >24.1)	0.164	**0.037**
High school completion	**0.014**	0.120

* All p values are uncorrected for multiple comparisons.

Figure 6.3: Kaplan-Meier survival curves for the time to sputum conversion (as measured by smear and culture) of individuals with the *VDR* genotypes.



In the log rank analyses there were significant associations with faster smear conversion time and no smoking, low number of cavities at diagnosis, low extent of tuberculosis at diagnosis, low absolute neutrophil numbers and completion of high school (Table 6.10). Non-smokers (n=12) converted sooner than smokers. This association was significant with smear conversion time and not with culture conversion time, but there was a similar trend in both (Figure 6.4). In smokers (n=202), an association between smear conversion time and *VDR* genotype was observed (p= 0.031), where individuals with a *TaqI* 'tt' genotype took longer to become sputum negative (Figure 6.5).

Faster culture conversion time was associated with females, low extent of tuberculosis at diagnosis, high absolute lymphocyte numbers and low absolute neutrophil numbers (Table 6.10).

Figure 6.4: Kaplan-Meier survival curves of A: smear (p=0.0162) and B: culture (p=0.1066) conversion times for smokers and non-smokers.

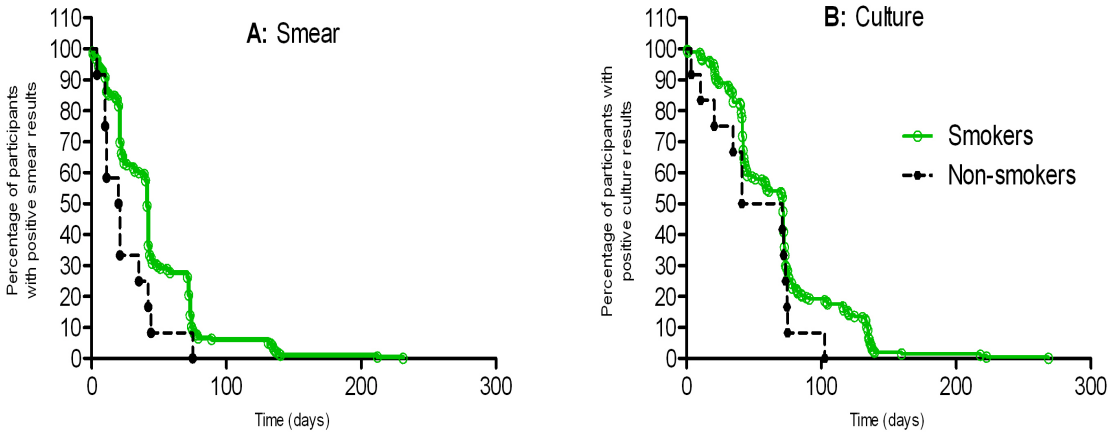
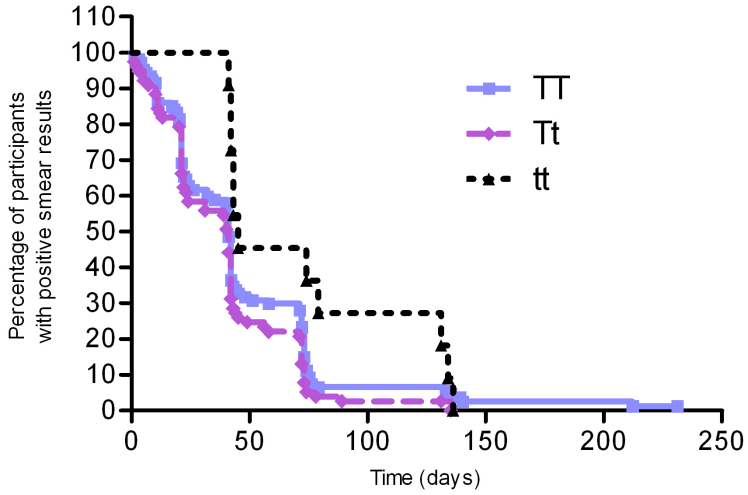


Figure 6.5: Kaplan-Meier survival curves for smear conversion time to negativity, of only smoking tuberculosis patients, sub-grouped according to *TaqI* genotype (p=0.0314).



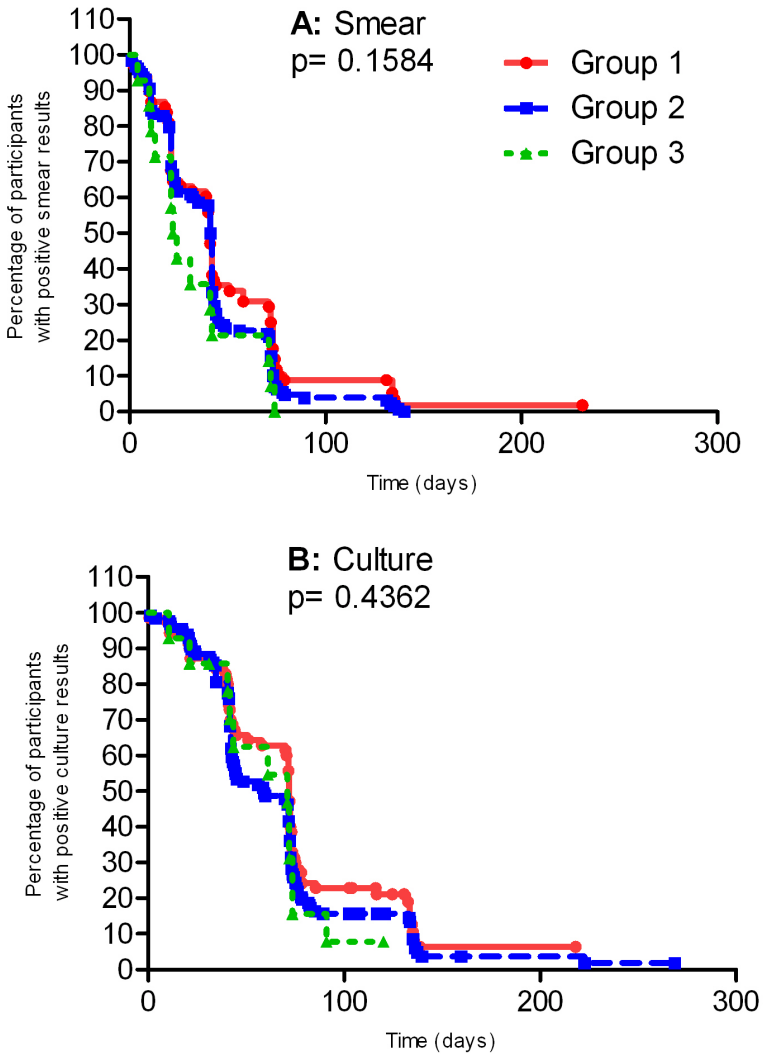
Kaplan-Meier survival curves were constructed for all variables available. Of particular interest in this department was the mycobacterial strains and the conversion time in patients. It has been hypothesized that different *M. tuberculosis* strains may give different responses to infection and treatment by the host. The SM cohort gave us the opportunity to investigate whether different strains have an effect on time to sputum conversion. The *M. tuberculosis* strain families and the median conversion times found in this cohort are summarized in Table 6.11. Kaplan-Meier survival curves were constructed for survival of all the RFLP families and also the genotypic or

virulence groups [324] (Figure 6.6) of the families. No significant associations were observed. The most frequent *M. tuberculosis* families in this cohort were Family 11 and Beijing and these made up more than 50% of the cohort (Table 6.11).

Table 6.11: *M. tuberculosis* RFLP families in the SM cohort and the median conversion time for each family.

RFLP Family	Count	Frequency	Median Culture conversion day	Median Smear Conversion day	Virulence Group
0	6	2.7	67.5	21.5	0
2	8	3.6	60	42	2
4	4	1.8	65.5	72.5	2
6	6	2.7	42.8	21	2
7	7	3.1	58.5	23	2
8	1	0.4	41.5	21	3
9	5	2.2	56	41	2
11	47	20.9	72	42	2
12	1	0.4	42.5	139	0
13	3	1.3	42	21	2
14	6	2.7	31.5	23	2
16	4	1.8	66.5	31	3
17	1	0.4	73.5	10	3
18	4	1.8	57.8	27.5	3
19	1	0.4	59	73	2
21	2	0.9	47.2	23	3
23	2	0.9	21	41	3
24	2	0.9	42	42.5	2
25	1	0.4	20.5	2	1
27	2	0.9	57.3	57.5	1
28	6	2.7	41.3	20.5	2
29	69	30.7	72	41	1
30	1	0.4	116	212	0
31	1	0.4	72	72	1
110	5	2.2	43.5	22.5	2
120	8	3.6	56.8	42	2
130	2	0.9	72	57	2
140	18	8.0	43	38	2
150	2	0.9	42.3	16	2

Figure 6.6: Kaplan-Meier survival curves for the sputum conversion time of the *M. tuberculosis* genotypic groups represented in the SM cohort.



6.4.2.3 Individual Cox regression analysis

In the age and gender adjusted individual Cox analyses of smear conversion time (Table 6.12), only the following factors were found to have a significant ($p < 0.050$) effect on faster conversion: not smoking, high school completion, low extent of tuberculosis at diagnosis, low total WBC and absolute neutrophil numbers, and high absolute lymphocyte numbers. The number of cavities present at diagnosis was weakly associated with smear conversion time ($p=0.054$). In the individual, age and gender adjusted, Cox analyses of culture conversion time (Table 6.12), only the following factors significantly affected faster conversion: low extent of tuberculosis at diagnosis, short height, high absolute lymphocyte numbers and low absolute neutrophil numbers.

Table 6.12: Results of the age and gender adjusted individual Cox regression analysis of the SAC cohort for smear and culture conversion times.

	Smear		Culture	
	estimated coefficient	**p value**	**estimated coefficient**	**p value**
FokI	0.116	0.330	0.220	0.074
ApaI	0.053	0.570	0.057	0.560
TaqI	0.062	0.550	-0.050	0.650
Smoker Yes	-0.064	**0.0430**	-0.274	0.390
Isoniazed sensitive No	-0.159	0.490	0.078	0.740
High school completion Yes	0.427	**0.0140**	0.270	0.120
Self reported symptoms	-0.037	0.540	-0.043	0.470
Extent at diagnosis	-0.226	**0.0005**	-0.208	**0.0017**
Income level	0.077	0.180	0.013	0.820
White blood cell numbers	-0.051	**0.0077**	-0.024	0.260
Absolute Monocyte numbers	-0.013	0.700	-0.010	0.790
Absolute Lymphocyte numbers	0.027	**0.0160**	0.035	**0.0063**
Absolute Neutrophil numbers	-0.021	**0.026**	-0.029	**0.0047**
Mantoux size at diagnosis	0.001	0.940	-0.015	0.230
Height (m)	0.073	0.930	-1.963	**0.0400**
Number of cavities >4	-0.293	0.054	-0.147	0.350
High Income (>R1000 per month)	0.044	0.830	-0.262	0.210
BMI	0.039	0.084	0.021	0.340

6.4.2.4 Optimal Cox model

Smear conversion time: Variables that were independently predictive of faster smear conversion time in the final, optimal Cox model (Table 6.13), were the genotypes *ApaI* 'AA' versus 'aa' and *TaqI* 'Tt' and 'TT' versus 'tt'. Low extent of tuberculosis at diagnosis, not smoking and low total WBC count at diagnosis were also predictive.

Culture conversion time: Variables that were significantly associated and independent predictors in the optimal Cox model for faster culture conversion time were low extent of tuberculosis at diagnosis, short height and low absolute neutrophil numbers. The magnitude of the effect of the variables can be seen by referring to the estimated coefficient in Table 6.13.

The final optimal Cox models constructed for time to smear and culture conversion, were both highly significant (p=0.0004) (Table 6.13). These models obviate the need to adjust for multiple testing.

Table 6.13: Optimal Cox regression analysis model of factors influencing time to sputum conversion in tuberculosis patients from the SAC population.

	Smear (n=220)		**Culture** (n=222)	
Variables:	**p value**	**Est. coef.**	**p value**	**Est. coef.**
Age	0.910	-0.001	0.560	-0.004
Gender	0.940	0.011	0.440	-0.145
Smoking Yes/No	**0.047**	-0.675	-	-
High school completion Yes/No	0.150	0.272	0.078	0.318
Extent of TB at diagnosis from chest radiograph (<RUL, =RUL, >RUL, > 1lung)	**0.010**	-0.186	**0.024**	-0.164
Total White blood cell count at diagnosis	**0.039**	-0.043	-	-
Absolute Neutrophil numbers at diagnosis	-	-	**0.023**	-0.026
Height (cm)	-	-	**0.003**	-0.029
FokI Ff versus *ff*	0.094	-0.576	0.470	0.262
FokI FF versus *ff*	0.480	-0.235	0.180	0.476
ApaI Aa versus *aa*	0.130	0.338	0.960	-0.011
ApaI AA versus *aa*	**0.040**	0.480	0.920	0.024
TaqI Tt versus *tt*	**0.047**	0.600	0.740	0.105
TaqI TT versus *tt*	**0.026**	0.679	0.960	-0.014
Final model Global p value	**0.0004**		**0.0004**	

RUL Right Upper Lobe

A significant p value (in bold) indicates that a variable in the model contributes independently to conversion time.

Est. coef. = Estimated coefficient for modelled conversion time in days.

6.5 Discussion

Host genetic susceptibility to the development of tuberculosis after infection has been the subject of many case-control studies comparing genotypes between the groups. More subtle effects of genotype are those impacting on the severity of disease, or the ability of the treated tuberculosis cases to recover i.e. undergo sputum conversion to negative culture or smear stain. We investigated both of the above aspects with respect to the VDR gene in a large study of a South African population, as *VDR* polymorphisms have been implicated both in susceptibility to pulmonary tuberculosis disease [28, 37, 169, 301], and recently, in the time to mycobacterial resolution of tuberculosis in a Peruvian population [285].

No significant association was found in our case-control analysis between pulmonary tuberculosis and the *VDR* polymorphisms. Diplotypes and haplotypes were also analysed because they are believed to be more informative than individual polymorphisms [37], but we found only weak associations between the *VDR* haplotype and tuberculosis susceptibility. In our study the *FokI-ApaI-TaqI* 'FAT' haplotype tended to be associated with tuberculosis and may be a risk factor, whereas the 'FaT' haplotype is possibly protective. A recent case-control analysis of a South African Venda population found that the haplotype 'FbAT', but not the genotypes, significantly protected from tuberculosis [183]. In a TDT study of West African families, Bornman *et al* [37] found an association between tuberculosis and the *FokI-ApaI* 'FA' haplotype where the *ApaI* 'A' allele showed increased transmission to affected offspring. However no association between the *ApaI* genotype and tuberculosis susceptibility was found in a Gambian case-control study [28]. In our case-control study 41% of cases had the *ApaI* 'AA' genotype compared with 33% of controls, but this difference was statistically non-significant, as was the association of the *FokI-ApaI* 'Fa' haplotype with controls (p=0.070). There was no association between the *VDR* diplotype and tuberculosis susceptibility. A larger sample set might provide more information although the present case-control study provided 80% power to detect an OR of 1.7 or higher.

The polymorphisms investigated here may be in LD with another polymorphism which determines susceptibility, possibly the polyA repeat region which affects transcription [139, 387]. Patterns of LD vary between populations, often accounting for the inconsistency in results from different investigations [69, 140, 169]. The population frequencies for the *VDR* genotypes differ between populations. Particularly in the Peruvian population there appears to be a major change in population frequency of the *FokI VDR* polymorphism. The most frequent allele in the Peruvian population for SNP *FokI* (f) is the least frequent in all other populations (Table 6.14).

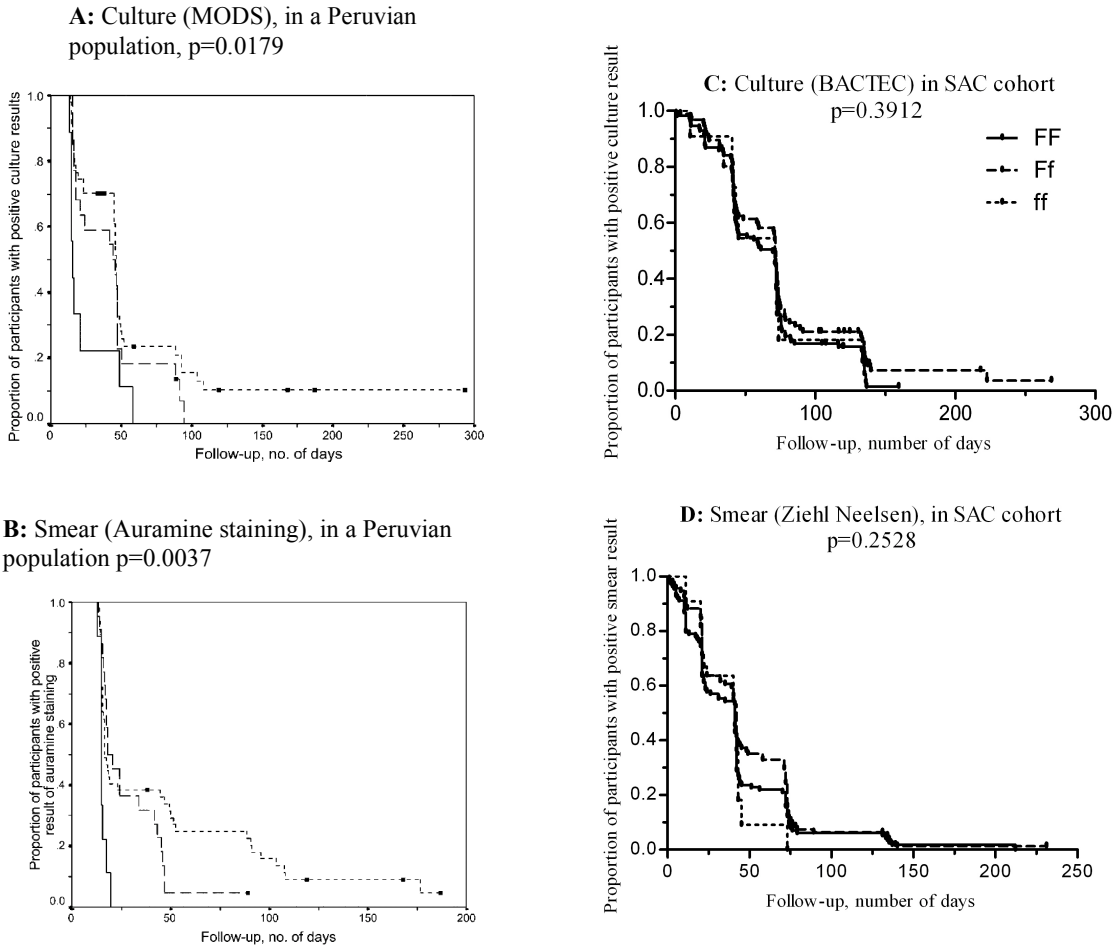
Table 6.14: Frequency of *VDR* genotypes in various populations (adapted from Wilbur *et al* 2007) [390].

	Genotype					
	FokI			***TaqI***		
Population	**FF**	**Ff**	**ff**	**TT**	**Tt**	**tt**
US, Mostly white [314]	0.36	0.49	0.15	0.36	0.46	0.18
African (Gambian) [28]				0.48	0.43	0.09
Gujarati [391]	0.61	0.34	0.05	0.41	0.50	0.09
Indian [301]				0.42	0.43	0.15
Indian [288]				0.43	0.45	0.12
Korean females [151]	0.39	0.44	0.17			
Chinese [332]				0.95	0.05	0.00
Han Chinese [52]				0.90	0.10	0.00
Taiwenese Females [53, 135]	0.42	0.34	0.24	0.90	0.10	0.00
Japanese [239]				0.66	0.31	0.03
Peruvian [285]	0.08	0.35	0.57	0.87	0.12	0.00
Ache, S. America [390]	0.67	0.33	0.00	0.44	0.51	0.05
Ava, S. America [390]	0.78	0.16	0.06	0.69	0.31	0.00
South African Coloured [16]	0.56	0.40	0.28	0.55	0.39	0.07
South African Venda [183]	0.77	0.21	0.02	0.57	0.42	0.01
West African (The Gambia, Guinea and Guinea-Bissau) [37]	0.62	0.34	0.05	0.52	0.39	0.08

Comparison between the fast and slow responders to therapy indicated that individuals with an *ApaI* 'A'-containing genotype had a faster response to treatment (culture conversion time p=0.029), and tended to convert before day 55 (month 2) when on standard DOTS treatment. *FokI* 'F'-containing genotypes tended to show delayed conversion, but the frequency of *FokI* 'ff' was too low in our population to show a significant effect (smear conversion time p=0.055). Roth *et al* [285] (n=78) found a significant difference at an earlier time point of 30 days (month 1), where the homozygous *FokI* genotype of the less frequent allele was associated with faster conversion time by culture or smear. Their study also found an association between *TaqI* genotype and culture conversion time at 30 days although no *TaqI* 'tt' genotypes were present in the Peruvian cases. In the SAC, the frequency of the *TaqI* 'tt' genotype is 7%. Investigating individuals that had a negative sputum result for both their culture and smear conversion times at days 30 or day 55 revealed no associations with the genotypes (Table 6.9). We found no association with *TaqI* and the response to treatment at day 30 or day 55, but *TaqI* was an independently significant predictor of smear conversion time in our final, optimal Cox model where *TaqI* 'tt' genotype was associated with delayed smear conversion time. Such associations could be used to identify people who would benefit from more intensive treatment or follow up.

The use of Kaplan-Meier survival curves allows for the visualization of the conversion time. The log rank analysis assesses differences between conversion times of the different subgroups but cannot take confounding factors into consideration. Roth *et al* found an association with *FokI* and both culture and smear conversion time, where the *FokI* 'FF' genotype (the least frequent genotype in the Peruvian population) converted fastest (Figure 6.7). We did not find an association in our Kaplan-Meier survival curves and the genotype *FokI* time to smear conversion (Figure 6.7).

Figure 6.7: Kaplan-Meier survival curves assessing the *VDR FokI* genotypes in A: Culture (MODS) and B: Smear (auramine staining) in a Peruvian cohort (n=78) compared to the conversion time in the SAC population C: Culture (BACTEC) and D: Smear (Ziehl Neelsen). The solid line is *FokI* 'FF' genotype, the dashed line is *FokI* 'Ff' genotype, and the dotted line is the *FokI* 'ff' genotype. The *FokI* genotype was associated with conversion time in the Peruvian cohort [285] as measured by culture, and by smear, where 'FF' resulted in faster conversion (p=0.0179).



There were numerous *M. tuberculosis* strains represented in this cohort, however the frequency of the rarer families was low and therefore did not provide enough information to make an assessment of their impact on time to sputum conversion. Families 29 and 11 were the most frequent in this cohort. This corresponds with previous studies done in the same population [275]. More samples of the rarer strains would be required to determine if they affect time to sputum conversion. Assessing the families according to their genotypic groups [324] did not reveal an association with time to sputum conversion.

In the cohort study of pulmonary tuberculosis cases, the only variable to contribute independently to the final optimal model of both smear and culture conversion time during treatment, was the extent of tuberculosis at diagnosis as determined by chest radiograph. It could

be expected that the extent of tuberculosis at diagnosis would be associated with the length of time to sputum conversion, with severely affected cases taking longer, as has been shown to occur with MDR-TB disease [134]. However, this is the first report, to the best of our knowledge, of a correlation between the severity of tuberculosis disease and the rate of treatment response.

In the final, optimal model of variables contributing to smear conversion time, *VDR* genotypes contributed independently to the model. Both the *TaqI* and *ApaI* genotypes were independently significant, as were smoking status and total WBC count at diagnosis. In the culture conversion optimal model, additional independently significant variables were absolute neutrophil numbers at diagnosis and height. Although not independently predictive of faster conversion, high school completion contributed to both optimal models, and lower educational level is a known risk factor for tuberculosis [306, 343]. The final models were both highly significant (p=0.0004).

The final model of the Cox regression analysis therefore indicates that the following variables are independent predictors of a faster conversion time in smear and/or culture: not smoking, low extent of tuberculosis at diagnosis, low total WBC count and low absolute neutrophil number at diagnosis, shorter height, an *ApaI* 'AA' genotype, and a *TaqI* 'T' containing *VDR* genotype. In the optimal model for smear conversion (Table 6.13), it can be seen from the estimated coefficient that *VDR* genotype and smoking had a greater clinical effect than the other variables. All effects were fairly modest.

The function of the *TaqI* polymorphism is not clear. Previously, the *TaqI* 'tt' genotype was found to protect against tuberculosis in a Gambian population [28] and the combination of the *TaqI* 'TT/Tt' genotype and 25-hydroxycholecalciferol deficiency was associated with tuberculosis in Gujarati Asians [391]. Contrary to these indications that *TaqI* 't' may be advantageous, the *TaqI* 'tt' genotype is associated with decreased levels of *VDR* mRNA and protein levels in peripheral blood mononuclear cells [244]. Roy *et al* [289] found the *TaqI* 'tt' genotype to be associated with susceptibility to leprosy *per se*, and *TaqI* 'tt' predisposed Tamil-speaking females to tuberculosis [301]. Our results correspond with the latter views of the function of *TaqI* in that cases with a 'tt' genotype take longer to respond to tuberculosis chemotherapy, and therefore may have a poorer immune response. Even though a patient is undergoing chemotherapy for tuberculosis, an efficiently functioning immune system is still necessary for optimum response and early sputum conversion.

The negative correlation between total WBC count at diagnosis or absolute neutrophil number at diagnosis and time to conversion in the final models for smear or culture respectively, may reflect the involvement of both of these cell types in the acute response to inflammation. Our results extend to the human model, the evidence recently presented by Keller *et al* [150], that an early influx of granulocytes contributes to susceptibility to *M. tuberculosis* in mice.

An association between smoking and tuberculosis infection status (a positive tuberculin skin test) has been reported in our population [71] and in India smokers have a higher rate of progression from infection to clinical tuberculosis [101]. We have shown here that smoking is also a significant factor in delaying sputum conversion time while on DOTS, as measured by smear.

This was seen in all three analyses done; log rank, univariate, and multivariate, optimal model. However, the number of non-smokers in this cohort was low and a further study would be needed to confirm this.

The association between height and culture conversion time could be an artifact due to the fact that there were a small number of people of small stature and high mass in whom culture conversion time was unusually fast. As can be seen from the estimated coefficient in Table 6.13, the actual effect of height was negligible.

The statistical model that is the most informative is the optimal cox regression (multivariate) model, as it takes all factors into consideration. The two analyses done before this model provide clues and assists in the interpretation of the optimal model. The association between smoking was consistently associated with time to sputum conversion. The same applies to extent of disease according to chest radiograph at diagnosis, which was associated with conversion time in both smear and culture. The absolute neutrophil number was also a significant factor in all three analyses. WBC count was not significant in the Kaplan-Meier survival curves but this might be because the subgroups were incorrect and the association was missed. Grouping could also explain why height was not found associated in the Kaplan-Meier survival curves.

The lack of concordance in the results between smear (ZN) and culture (BACTEC) conversion illustrates the complexity of this disease and the subtly different end points measured by the two techniques. At month 2, 31% of patients were smear negative but still culture positive. Roth *et al* found an association between *VDR* genotypes and culture conversion times (as measured by MODS) [285] whereas in our large study of patients with strict inclusion criteria and intensive follow-up, associations were found with smear conversion times. Culture is more sensitive, but as a very low number of bacteria can give rise to a positive culture, it may be a less robust tool for measuring an endpoint of sputum conversion time.

Further investigations into the effects of Vitamin D metabolism in response to the medication used in DOTS, would also be of benefit to better understand the role of Vitamin D and its receptor on response to tuberculosis treatment [41, 64]. There have been some reports on the benefits of vitamin D supplementation during tuberculosis treatment but they have been methodologically flawed [201]. A double blind trial placebo-controlled study was recently reported which showed that a single oral dose of 2.5mg vitamin D significantly enhanced the ability of participants' whole blood to restrict BCG-*lux in vitro* 6 weeks later [202].

Currently, in most tuberculosis programmes, the first measure of effective therapy is sputum conversion at 2 months. The discovery of a genetic or immunological marker indicating how efficiently a patient responds to treatment could have a major impact on clinical trials and shorten the time period necessary for drug testing. In the coming era of personalized medicine, genetic markers have the potential to be developed into simple, affordable tests applicable even in high-burden countries. Estimates of disease transmission in the community studied here are extremely high, and the percentage of tuberculosis due to transmission could be in excess of 70% [362]. The extended time to microbial resolution of tuberculosis may be a contributory factor in the continuation of the tuberculosis epidemic.

Chapter 7

Concluding remarks

In the South African Coloured (SAC) population, susceptibility to tuberculosis was investigated using a case-control study design. With the number of samples available, there was sufficient power to determine whether there was an association with tuberculosis susceptibility in a variety of candidate genes (Table 7.1). Polymorphisms in these candidate genes were assessed for an association with susceptibility to active tuberculosis and not susceptibility to *Mycobacterium tuberculosis* (*M. tuberculosis)* infection (otherwise healthy individuals that are latently infected).

Table 7.1: Summary of all the polymorphisms investigated via a case-control method using the SAC population.

Gene	SNP	Results in table:	significant p values	Comment
RANTES	-403(rs2107538)	3.1		
RANTES	+1092(rs1065341)	3.1		
CCR5	59029(rs1799987)	3.1		
CCR5	Δ32(rs333)	3.1		
CCR2	64I(rs1799864)	3.1		
SDF1	3'A(rs1801157)	3.1		
DC-SIGN	-939	4.1		
DC-SIGN	-871	4.1	$8.2x10^{-4}$	
DC-SIGN	-336	4.1	0.01	
DC-SIGN	-139	4.1		
DC-SIGN	2392	4.1		
DC-SIGN	3220	4.1		
DC-SIGN	3838	4.1		
DC-SIGN	4235	4.1		
DC-SIGN	neck length variation	4.3		
L-SIGN	neck length variation	4.3		
SP110	rs2114592	5.2		
SP110	rs1365776	5.2		
SP110	rs41545715	5.2		
SP110	rs3948464	5.2		
SP110	rs1135791	5.2		
SP110	rs3948463	5.2		
SP110	rs41541917	5.2		
SP110	rs35495464	5.2		
VDR	*FokI* (rs10735810)	6.3		
VDR	*ApaI* (rs7975232)	6.3		Appears to weakly contribute to smear conversion time (Table 6.13)
VDR	*TaqI* (rs731236)	6.3		Appears to weakly contribute to smear conversion time (Table 6.13)

In addition, this admixed population was formally assessed for population stratification and the Ravensmead/Uitsig SACs were found not to be stratified. However, a possible criticism is that the 25 single nucleotide polymorphisms (SNPs) used were not enough to confidently assess the stratification within an admixed population. Further analysis of more SNPs or microsatellites could be done [258] to strengthen the evidence that the SAC population is not stratified. The impact that possible population stratification could have on case-control association studies is debated and the extent and implications have not been established [47, 96, 346, 377].

Candidate polymorphisms in RANTES, CCR5, CCR2 and SDF1, which had previously been associated with a number of diseases, were not associated with tuberculosis susceptibility. The polymorphisms occurred at high enough frequencies to provide sufficient power to detect an OR of at least 1.6. The RANTES polymorphism -403 initially showed an association, which was lost when additional samples were genotyped. This illustrates one of the challenges with association studies.

The *DC-SIGN* gene was re-sequenced using samples from the SAC population, and the genotyping of promoter polymorphisms -336 and -871 found both to be associated with tuberculosis susceptibility. The association was strong and was still present after (non-significant) stratification adjustment. The *DC-SIGN* -336A variant has been implicated in increased DC-SIGN expression which could explain its role in tuberculosis susceptibility. The increased expression of DC-SIGN may be the underlying reason for an increased efficiency of host phagocytes. As both *DC-SIGN* -336 and -871 are promoter polymorphisms, further studies assessing their functional consequences, for example gene expression studies where luciferase reporter vectors are utilised, are now required to eventually develop knowledge-based and effective pathway-targeted treatments. These functional studies are being done at the Pasteur Institute, Paris, France. Additional *DC-SIGN* haplotype tagging SNPs (htSNPs) were also investigated but not associated with tuberculosis susceptibility.

The *SP110* gene was identified as a potential susceptibility gene through mouse studies, as was *SLC11A1* (NRAMP1). Re-sequencing sections of the *SP110* gene identified fourteen novel polymorphisms and in the case-control study eight polymorphisms in the *SP110* gene were genotyped. No SP110 genotypes or haplotypes were associated with tuberculosis susceptibility. Although, *SLC11A1* was identified through mouse studies and subsequently proved to be a susceptibility gene for tuberculosis in humans, the same does not appear to be true for the *SP110* gene. The mouse model has its uses but it is not designed to identify human candidate genes, as the mouse does not always reflect the human disease model sufficiently to allow findings to be translated to the human patient.

Both the *SP110* and *DC-SIGN* genes were re-sequenced to determine the frequencies of the polymorphisms, identify novel SNPs and determine linkage disequilibrium (LD) and infer haplotypes present in the SAC population. When testing for an association with tuberculosis, haplotype tags (htSNPs) and tagSNPs were desired, as one polymorphism could provide information on a whole haplotype. In particular, in an admixed population it is useful to sequence and identify LD and haplotypes within that population. Re-sequencing *SP110* gene sections to determine the haplotype and LD showed low LOD values and no tagSNPs could be determined. The re-sequencing of the *DC-SIGN* gene did enable us to determine htSNPs.

There are indications that 60 independent samples would provide optimal performance of tagSNPs, when considering a common SNP (>5% frequency) [220]. In a resource-poor setting the sequencing of 60 samples might not be feasible, in particular when the gene is fairly large. The use of htSNPs and tagSNPs does reduce the cost of genotyping as all SNPs spanning the gene do not need to be genotyped. Information on tagSNPs and htSNPs is available from the HapMap database (http://www.hapmap.org/). The SAC population is not represented in this

database, but the information has been found to be translatable to other populations. Using online resources will provide a wealth of information from reference populations, but they seldom reflect admixed populations, a term that can refer to many distinct populations. As the SAC population is admixed with contributions from Asian, Caucasian and African populations, all the tagSNPs spanning the gene of interest from all the reference populations should be used, as this will provide the most reliable information. This includes polymorphisms that are unique to one of the reference populations together with the other tagSNPs identified, and common to all three reference populations (M. Möller PhD thesis 2007). However, this could result in a large number of polymorphisms that would need to be genotyped. Thus the initial expenditure of re-sequencing genes in samples from the SAC population might be beneficial in providing valuable information on LD, tagSNPs and novel SNPs in this population. Re-sequencing or using online resources both have their advantages and disadvantages and do not have to be mutually exclusive. Combining the information from both sources provides the most information.

Care must be taken when assessing published positive associations with disease. There are many underpowered, poorly-defined studies published and they should be used with caution. Specific guidelines for the publication of association studies have been released to ensure that there is no bias towards positive associations, and studies are published on the basis of correct design, not a positive result [281].

There are a plethora of positive association studies that have not been replicated and there are numerous reasons why some studies cannot be replicated [58]. As seen in Chapter 3, there was initially a positive association with the *RANTES* -403 polymorphism but after analysis of more samples there was no association. The positive association could be spurious and also indicates why replication in an independent study and with an alternative population is important. The initial study should be scrutinised to assess whether it has enough power to justify its claims. A simple check is to assess if the cases and controls are reasonably closely matched for age, ethnicity, gender and, if feasible, similar exposure in environment. Care must be taken that reporting of subgroups is not done in an attempt to put a positive spin on negative results.

Polymorphisms in the *SP110* gene were associated with tuberculosis susceptibility in West African populations but not in SAC, Russian and Ghanaian populations. This could indicate a spurious result in the original West African populations, or the polymorphisms could be indirectly associated with disease through LD. The investigations in Ghana and Russia used thousands of samples in their case-control association studies [335, 348]. Only if the polymorphism occurs at a low frequency are these numbers required. A large number of samples is not always necessary if there is sufficient power in the statistical analysis [43], but it does provide convincing evidence for negative results. However, when a polymorphism with a low frequency is associated with susceptibility for a common disease, the practical implications are limited, as the population attributable fraction of the susceptibility genotype will be low. In other words, the findings will have theoretical value but have little practical public health value [279].

Finally, *VDR* gene polymorphisms *FokI*, *ApaI* and *TaqI* were investigated in a case-control study and found not to be associated with susceptibility to tuberculosis. The 249 cases were from a longitudinal study, known as the Surrogate Marker (SM) cohort, of first time pulmonary

tuberculosis patients, which were followed during and after their treatment. With the regular sputum sample testing which had been done, it was possible to reliably determine when sputum conversion, from a positive to negative result, took place. Sputum results were available for both smear and culture. Apart from Kaplan-Meier survival curves and univariate analysis, an optimal Cox regression (multivariate) analysis was done. The optimal Cox regression model is a useful analytical method as it takes all variables into consideration and no statistical correction for multiple testing is required. The final optimal model lists all factors that may contribute to the outcome being investigated, i.e. time to sputum conversion. The optimal model identified that the *VDR* polymorphisms, *ApaI* and *TaqI* contributed to smear conversion time. The extent of tuberculosis at diagnosis, according to the chest radiograph, was the only variable that contributed to both smear and culture conversion time. This model provides a wealth of information that takes factors from the environment, host and pathogen into a holistic consideration, and not merely as individual factors. This provides a more nuanced investigation of gene effects than the conventional case-control study.

The SM cohort will be useful for further studies investigating time to sputum conversion in tuberculosis patients and the contributions of various genetic loci to treatment outcome. This type of investigation has only been done with the *VDR* gene in SAC and Peruvian populations. Knowledge of which gene polymorphisms contribute to sputum conversion would provide valuable information for antibiotic development and trials, and possibly provide a surrogate marker to identify poor responders to treatment. Selection of candidate genes for association with time to sputum conversion would be different from those assessing associations with susceptibility to tuberculosis disease, as sputum conversion will involve genes important in recovery from tuberculosis disease while on treatment and not necessarily those involved in progression to disease.

The ability of the *M. tuberculosis* strains to elicit a heterogenous response from the human host and the role that genetic variants play needs further investigation. The interaction between *M. tuberculosis* strains and *VDR* genotype was briefly assessed in the SM cohort, with respect to influence on time to sputum conversion, but there were no significant findings. In genetic association studies it would require the cases to be phenotyped and catergorised by the strain type of the patient and the clinical form of disease. There are clearly interacting roles for the host, pathogen and the environment and we require a better understanding of the intricate interplay taking place.

Tuberculosis susceptibility has been investigated in a number of genes worldwide and a number of positive associations (some unpublished) have been found. The combined effects, either additive or epistatic, of these polymorphisms needs to be determined [229]. As the response to *M. tuberculosis* infection is complex, it is possible that individual susceptibility genes may occur in complementary pathways, or gene function may be compensated for by another redundant gene. In addition, epigenetic effects might also be playing a key role in susceptibility to tuberculosis disease and these have not been explored yet [144]. Comprehensive studies considering the genomics, transcriptomics, proteomics and metabolomics would provide a complete view of the hosts' response to *M. tuberculosis* infection.

These results, and other association studies, could have implications for vaccine and drug development [114, 392]. The design of drug and vaccine trials might need to take the effect of these polymorphisms into consideration, either by stratification of study populations or correction in statistical analysis of the efficacy of the drug or vaccine. Failure to consider inter-population genetic differences could have implications on the efficiency and the outcome of trials when tested in different populations. This important point has already been raised in HIV vaccine trials [114, 392] but would apply equally to tuberculosis. This may already be demonstrated by the variable efficacy of BCG vaccination in different populations [9]. Host genetic variability might not only affect the efficacy of a trial between sites but in addition disease progression and the outcome of treatment. Knowledge of the genetic background of the susceptibility to a disease is required to implement effective treatment. Positive genetic associations in the population that we studied would need to be replicated in other population groups, including those from other areas in South Africa, before we could contemplate using this knowledge to improve tuberculosis treatment and public health.

Tuberculosis is a major burden to the world, especially developing countries. With no major developments in vaccines and drug therapy for over 40 years, something drastic needs to happen. There are new drugs in the clinical trial pipeline but there seem to be no 'golden bullets' within sight. There is still a lot we do not understand and we require better knowledge on what contributes to tuberculosis disease, which will hopefully help develop better vaccines or therapeutics. Therapeutics that reduces the treatment period of 6 months and the development of rapid, low cost diagnostics are required. A vaccine that would protect both children and adults and particularly one that could target the genetically susceptible groups would help to reverse the epidemic.

■

Appendix 1

Methods and Materials

These Methods were used in the following publications:

Babb C, Keet EH, van Helden PD, Hoal EG (2007)
SP110 polymorphisms are not associated with pulmonary tuberculosis in a South African population.
Hum. Genet. 121:521-2

Babb C, van der Merwe L., Beyers N, Pheiffer C, Walzl G, Duncan K, van Helden P, Hoal EG. Vitamin D receptor gene polymorphisms and sputum conversion time in pulmonary tuberculosis patients.
Tuberculosis 2007;87:295-302.

Barreiro LB, Neyrolles O, **Babb CL**, Tailleux L, Quach H, McElreavey K, Helden PD, Hoal EG, Gicquel B, Quintana-Murci L (2006)
Promoter Variation in the DC-SIGN Encoding Gene CD209 Is Associated with Tuberculosis.
PLoS Medicine 3:e20

Barreiro LB, Neyrolles O, **Babb CL**, van Helden PD, Gicquel B, Hoal EG, Quintana-Murci L (2007)
Length Variation of DC-SIGN and L-SIGN Neck-Region has no Impact on Tuberculosis Susceptibility.
Human Immunology 68:106-12

A1.1 Patients and Controls

Samples have been collected over the last 10 years from the Ravensmead and Uitsig communities and some adjacent suburbs in the proximity of Tygerberg Hospital, Parow, Western Cape. Individuals that were enrolled for the studies are from the population discussed in Chapter 2, namely the South African Coloured.

All projects received ethical approval from the ethics committee of the Faculty of Health Sciences, University of Stellenbosch, South Africa. The case-control study received project registration number: 95/072, the Surrogate Marker has project registration number 99/039 with amendments and the Sequella/ Aeras study has registration number 95/072 with amendments.

Tuberculosis patients are diagnosed and bacteriologically confirmed through the local health system and placed in the direct observed therapy, short course (DOTS) treatment as prescribed by the South African National Tuberculosis Programme, based on World Health Organization (WHO) guidelines discussed in Chapter 1.1.4. For the human susceptibility to tuberculosis studies described, only HIV negative patients were included.

Controls were enrolled from a variety of sources such as clinics, households and places of work in the same area as the tuberculosis cases. They were all from high TB incidence suburbs, similar socio-economic status and the SAC population group. They reported having no clinical symptoms of tuberculosis and have no record or recollection ever having had tuberculosis. In addition their medical records (if any) were checked for a history of tuberculosis. No PPD results were known, but as discussed in Chapter 1, they are very likely to be positive because of the high tuberculosis exposure in this community [371, 372, 383]. Familial information was collected for over 225 families, to allow for family-based studies, such as transmission disequilibrium test (TDT) studies. I checked and constructed all pedigrees, for future TDT studies, in Cyrillic version 3.0.3. However no familial studies were done for this thesis.

Venous blood samples of between 2 and 10 ml were collected from subjects in one to two EDTA lavender vacutainers (BD Vacutainer Systems, Plymouth, UK, provided by Scientific Group Cat no. BD367655). The date of sampling, date of birth, medical history, consent forms and, in some cases questionnaires, were all collected at the same time. Records were kept in the "EileenTB" database which is stored on the server "Zeus" in the Division of Molecular Biology and Human Genetics, Faculty of Health Sciences, Tygerberg Campus, University of Stellenbosch. Confidentiality is strictly adhered to with numerical Patient IDs being assigned on the first visit. All clinical and epidemiological information is linked to the Patient ID and only clinicians, the primary investigator and database manager were permitted access to patient names.

A1.2 Genomic DNA Isolation

On the day of phlebotomy samples were kept upright at room temperature. Once the plasma had settled one millilitre was pipetted from each of the vacutainers into a 2ml tube. This was then centrifuged at 14 000 rpm for 1 minute, so that any remaining red blood cells pelleted. Two ml of

plasma was then pipetted into two one millilitre eppendorfs to be stored at minus eighty degrees. The collection tube containing the pellet was discarded. The two aliquots of remaining whole blood were then pooled in a 50ml sterilon collection tube and stored at minus 20°C until genomic DNA was extracted using the Nucleon BACC3 Kit for blood (Amersham Biosciences, UK Catalogue number: BACC RPN-8512). Extractions were done according to manufacturers instructions, which are briefly described below.

Outline of Nucleon DNA extraction case-control Samples

1. Dilute Reagent A 1 to 4 in distilled water: (12 samples = 94.5ml Reagent A + 283.5ml water).

Cell lysis

2. Collect the defrosted 50ml collection tubes, containing the pooled blood, from cold room. Two millilitres of plasma has been removed and should contain 5-10 ml of whole blood. Fill with diluted Reagent A to 30 ml.
3. Vortex each tube for +- 30 seconds
4. Centrifuge at 1300g for 15 minutes (Switch the waterbath on and dilute some bleach).
5. Pour off the supernatant into the diluted bleach. Be careful not decant the pellet.
6. Add 2ml of diluted Reagent A to the remaining pellet in the 50ml collection tube (double lysis). Shake by hand to release the pellet from the bottom of the tube.
7. Centrifuge the 50ml tubes at 1300g for 15 minutes. Collect 15 ml sterilon tubes, two for each sample. Add 500µl of sodium perchlorate to one set of empty 15ml tubes.
8. Pour off the supernatant in the 50ml tubes into the diluted bleach, again being careful not to loss the pellet.
9. Add 2ml Reagent B to the 50ml tube.
10. Place the samples in the water bath (37°C) until the pellets are dissolved in reagent B. Check and swirl by hand every few minutes.

Deproteinisation

11. Transfer the diluted pellet to the 15ml sterilon tube, containing the 500µl of sodium perchlorate solution, by pouring. Mix by inverting ± 20 by hand.

DNA extraction

12. Add 2ml of chloroform to the 15ml tube containing the sample. Mix by inverting ± 20 by hand.
13. Add 300 µl of resin that is supplied with the kit. Add slowly and close to surface so it does not splatter.
14. Centrifuge the sample at 1300g for 8 minutes

DNA precipitation

15. Remove the top layer, ~2.5ml (NOT RESIN), to an empty 15ml sterilon tube.
16. Add 5ml of ice cold ethanol. The DNA will precipitate and become visible.
17. Remove the strands out of the ethanol by pipetting, to a 1.5 ml eppendorf tube. +- 70 µl of ethanol can be pipetted with the DNA. Invert the 15ml tube as more DNA lumps may appear, collect any precipitate.

DNA washing

18. Centrifuge the 1.5ml tube at top speed (4000g) for 5 minutes to pellet the DNA.
19. Pour off the ethanol supernatant.
20. Add ~ 1.5ml of 70% ethanol.

21. Centrifuge for 5 minutes at top speed.
22. Pour the ethanol off completely. Carefully as not to loose the pellet of DNA.
23. Reconstitute the DNA pellet with TE. The volume of TE required depends on the size of the pellet, usually between 200 and 1000 µl when starting with 5-10ml of whole blood. Flick the 1.5ml tube to release the pellet. The pellet must dissolve fully before reading the concentration on the Nanodrop (NanoDrop Technologies Inc., USA). Advisable to leave on a rotary wheel, at least, overnight.

A1.3 Polymerase Chain Reaction

All techniques in this thesis used PCR as the starting point of all genotyping. The different elements of PCR will be discussed in this section. Unless stated otherwise all PCRs were done on an Eppendorf Mastercycler Gradient PCR Machine (Eppendorf, Hamburg, Germany).

A1.3.1 Primer Design

Correct primer design is important to achieve accurate results. Using online databases is essential but should be used with caution as mislabelling and incorrect alignments can sometimes occur. Online databases store globally produced data and make them publicly available. The most commonly used are: NCBI (http://www.ncbi.nlm.nih.gov/), Ensembl (http://www.ensembl.org) and EMBL (http://www.ebi.ac.uk/embl/). A comparison of the desired target nucleotide sequence between these different databases is desirable because, even though they have overlap in data input, they also have unique submissions of data from various sources around the world. When designing a primer, and consideration for pseudogenes, incorrect alignment and homologues must be taken.

If possible the target DNA sequence should be sequenced. Not only will sequencing ensure that the correct sequence is being used, but will also allow for the detection of novel polymorphisms that may not be present in populations that currently dominate genome projects, such as the CEPH group, white American or African American. Once the correct target nucleotide sequence template has been identified, primer design can take place.

There are a number of computer programs that can assist with primer design. We used DNAman Version 4.1 (http://www.lynnon.com/), but another that is commonly used is Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Information on the primer includes the GC content, melting temperature (Tm), molecular weight and more. Primer design must follow these basic guidelines: avoidance of substantial tandem repeats of one or more nucleotides, avoid sequences prone to secondary structures which could form hairpin loops and dimerization, and the Tm should be between 50 and 60 °C.

The program Oligocalculator (http://trishul.sci.gu.edu.au/tools/OligoCalculator.html) was used as an independent check for the Tm and the GC content. Testing for the presence of hairpins and dimers was done by using Scitools Oligoanalyzer 3.0

(http://scitools.idtdna.com/Analyzer/oligocalc.asp). This program tests for auto- and hetero-dimers, and hairpin loops. Any dimers that were more than 4 bp and overlapped the 3' end were rejected, as were primers where a hairpin bound at the 3'end and had a Tm above 39.9 °C. The program AutoDimer was used as an additional check [359].

To check that the primers will anneal only to the desired target, a BLAST needs to be conducted. This is done through the NCBI website (http://www.ncbi.nlm.nih.gov/blast/) and choosing the "search for short nearly exact matches" or "nucleotide-nucleotide BLAST" option to ensure specificity. Selection of "chromosome" as the database, "Homo sapiens" as the organism, and checking the "Mask for lookup table only" box, which then gives a real idea of the hits on the genome without all the confusion of different genes, mRNA etc. It must be noted that the BLAST results are ranked by "expect value", and this calculation means that a hit of 16 continuous bases in a 20bp primer will rank higher than a 19bp hit with one gap in it. Therefore all hits must be manually checked to find the ones of biological importance.

A virtual PCR on the website (http://grup.cribi.unipd.it/cgi-bin/mateo/vpcr2.cgi) was done to check if amplification would occur with the chosen primer pair. However this is just a check and cannot replace actual experimentation and optimisation of the PCR, as it is not always accurate.

A1.3.2 PCR amplification

Super-Therm Gold DNA polymerase (JMR holdings, USA, Catalogue Number JMR 850 supplied by Southern Cross SA) was the Taq used throughout all the experiments. Deoxynucleotide triphosphates (dNTPs) were all supplied from Bioline (catalogue number BIO-39025) and a master mix with 25mM of each dNTP, with a working concentration of 2.5mM was used throughout. Primers were manufactured by Integrated DNA Technologies Inc (IDT, USA; supplied by Whitehead Scientific, SA). The stock was diluted with distilled water to a concentration of 100µM and a working concentration of 10µM. Genomic DNA was used at a concentration between 10 and 50ng/µl.

PCR programs were done with the general program: 1 cycle at 94°C for 15 minutes followed by 35 cycles at 94°C for 30 seconds, annealing temperature °C for 30 seconds and 72°C for 45 seconds, then 1 cycle at 72 °C for 10 minutes and finally a 15°C hold step.

Agarose gels were either Tris-borate-ethylenediaminetetra acetic acid (TBE) or Sodium Boric (SB) acid [42] buffer gels. SB buffer allows gels to be run at higher voltage without causing a high temperature increase, which decreases the run time of a gel.

A1.4 Sequencing

Instead of the Sanger/ chain terminator sequencing method, labelled terminators were used, commonly called 'dye terminator sequencing'. The major advantage of this approach is that the complete sequencing set can be performed in a single reaction. This is accomplished by labelling each of the dideoxynucleotide chain-terminators with a separate fluorescent dye, each with a different wavelength. Software allows for automated calling of the bases.

A1.4.1 PCR Clean-Up

To get good quality sequencing, cleaning/purification of PCR products is required to remove primers, buffers and dNTPs. Initially, all purification for genomic DNA sequencing was done using the Wizard ® SV Gel and PCR Clean-Up system (Promega, Catalogue number: A9282), briefly described below.

Outline of Wizard ® SV Gel and PCR Clean-Up system

This is a DNA purification done by centrifugation protocol, as described by the manufacturer:

Processing PCR reactions

1. Added equal volumes of the membrane binding solution to the PCR reaction.

Binding of DNA

2. The SV minicolumn was inserted into the collection tube.
3. The prepared PCR sample was transferred to the minicolumn assembly and incubated at room temperature for 1 minute.
4. Then was centrifuged at 16 000g for 1 minute. The through-flow was discarded and the minicolumn reinserted into the collection tube.

Washing

5. 700µl of Membrane Wash Solution (ethanol already added, as required by manufacturer) was added to the minicolumn. Then was centrifuged at 16 000x g for 1 minute. Through flow was discarded and the minicolumn reinserted into the collection tube.
6. The above step was repeated with 500µl of membrane wash solution, and centrifuged at 16 000g for 5 minutes.

Elution

7. Carefully transferred the minicolumn to a clean 1.5ml microcentrifuge tube.
8. 50µl of nuclease free water was added to the minicolumn and incubated for 1 minute at room temperature, then was centrifuged at 16 000g for 1 minute.
9. Discarded the minicolumn and stored the cleaned PCR product at 4°C or -20°C until needed.

The above method worked but we found that the ExoSAP-IT (USB Corporation, Catalogue Number: US78201) clean-up method produced a definite improvement in sequencing results.

The ExoSAP-IT protocol, briefly described below, was easier and simpler to follow and despite being more expensive, the benefits of having clean, clear sequencing outweighed the costs.

Outline of ExoSAP-IT Clean-Up

1. ExoSAP-IT was to be stored on ice through the whole set-up.
2. 8µl of post-PCR reaction product was mixed with 4µl of ExoSAP-IT. Giving a total of 14µl. This gave excess volume for bi-directional sequencing (one sequencing reaction required 6 µl). Compared to the manufacturers' advice we added a slight excess of ExoSAP-IT to guarantee Clean-Up.
3. The mix was incubated at 37°C for 15 minutes to degrade the remaining PCR primers and nucleotides.
4. A further incubation step at 80°C for 15 minutes was done to inactivate the ExoSAP-IT enzymes.
5. The cleaned PCR product was then stored at 4°C or -20°C until needed.

A1.4.2 Sequencing reactions

Sequencing was done by using the ABI PRISM BigDye Terminator v3.1 Ready Reaction Cycle Sequencing Kit (100 Reactions) (Applied Biosystems, CA, USA, Part Number: 4337455), from here on referred to as the sequencing kit. It contains all the required reagents in a premixed, ready to use format, including the fluorescently labelled dideoxynucleotides ([F]ddNTPs), which are added sequentially to the primer through the sequencing reaction. Sequencing templates should generally be shorter than 800bp, but there are other kits on the market which will sequence over 1000bp successfully.

Per sample, the reaction required 0.5µl of DMSO, 1.5µl of sequencing kit (these were premixed in a 1:3 ratio), 5µl of primer at a concentration of 1.1µM, and 5µl of cleaned template PCR product. The additive DMSO is required to open secondary structures that may form. For bidirectional sequencing, 10µl of PCR product is required and 5µl of each primer, per sample. Sequencing reactions were done in a 96-well plate format. Each plate contained a positive control which is provided with the kit. The positive control requires twice the volume of the sequencing premixed reagents (3µl), 1µl of DMSO and 1.5µl of pGEM template with 5µl of T13 primer. The sequencing reaction for the samples was scaled down from manufacturers' recommendations to save consumables.

After setting up the sequencing reaction, the plate was run on a 3700xl thermocycler (Applied Biosystems) with the following cycles: 95°C for 5 minutes, followed by 35 cycles of 95°C for 45 seconds followed by 58°C for 4 minutes and then a hold step of 4°C.

A further cleaning step to remove the [F]ddNTPS and primers was required using the Centri-SEP 96 protocol, briefly described below, was also in plate format. The samples were then ready to be run on the ABI-Prism 3130xl DNA analyzer platform.

Outline of Centri-SEP 96 Protocol

Clean-up of the sequencing reaction was done according to the Centri-Sep 96 protocol (Princeton Separations, USA Catalogue number CS-965). This is a simple and easy method to follow.

1. The Centri-Sep 96 plate must be at room temperature before use. Removed the adhesive from the bottom and then from the top of the Centri-Sep 96 plate.
2. The Centri-Sep 96 plate was stacked on top of a 96-well wash plate and centrifuged at 1500xg for 2 minutes. (The use of an external timer was advised.) Discarded the liquid from the wash plate. The gel matrix in the wells appeared opaque now.
3. The samples (20µl or less) were transferred to the Centri-Sep 96 plate, taking care to place the samples in the centre of the gel beds.
4. The Centri-Sep 96 plate was stacked on top of a 96 well collection plate (96-well optical reaction plate, Applied Biosystems, Part number: 430673) and centrifuged at 1500g for 2 minutes, again using an external timer. The samples were ready to be analysed.

In the Data Collection software Version 3.0 (Applied Biosystems), the layout of the plate was entered, and linked to the plate loaded on the platform of the sequencer. The assignment of sample names was done so that the sample name and the primer used were easily identified. Prior to loading the plate on the analyzer, it was ensured that the running buffer (stock solution of running buffer with EDTA was a 10x concentrate (Applied Biosystems: Part number 402824 25ml)) was fresh and replaced if necessary.

The sequencing reaction was run on a 36 cm 16 capillary array (3130 Cap Array 36cm, Applied Biosystems, part number: 4315931) with Performance Optimum Polymer (POP) 7 (Applied Biosystems, part number 4352759) in the Division of Molecular Biology and Human Genetics using the Sequencing analysis software Version 5.2 (Applied Biosystems) with the Instrument protocol 'Seq_Z_36_POP7' and Analysis Protocol '3130POP7_BDTv3-KB-Denovo_v5.2'. Sample results were then analysed on the Sequencing Analysis software. As part of quality control, all plates contained pGEM with primer T13 as a positive control and all results were analysed first by the sequencer operator and subsequently and independently by the researcher.

The *SP110* gene was sequenced to identify novel polymorphisms and to search for the presence and frequency of known polymorphisms in our population. We sequenced *SP110* exons 4 to 8, 11, 14, 15, alternative 15, 17 and 18, including surrounding intronic regions. Per sample, amplification was done using 0.13µl of Taq, with 2.5µl of 10x PCR buffer, 2µl of dNTPs, 13µl of distilled water and 0.35µl of each primer (forward and reverse), giving a total volume of 19µl before 1µl of DNA template was added. The final 20µl reaction mix was amplified using the primers listed in Table A1.1. The samples were then cleaned with ExoSAP-IT before undergoing the sequencing reaction, the Centri-Sep 96 purification and finally being run on the ABI Prism 3130xl Genetic Analyzer, as described previously.

Table A1.1: Amplification and Sequencing Primers for the *SP110* gene with the fragment lengths and annealing temperatures.

Primer Name	Orientation	Sequence 5' → 3'	Fragment Length	Annealing temp. (°C)
*SP110*Exon4,5	Sense	TCT GAG TCT CCA GGG TAC TGA TG	796	62
	Antisense	GGT TGG CAG ACG CAT GTT		
*SP110*Exon6	Sense	AGC TTC TCA AAT TGT ATC ACT TGT CC	737	62
	Antisense	CTA TCC AAG TCT ACC TTT TCC AGA CT		
*SP110*Exon7	Sense	GCA GAG CTT TAT ATG TCT TTG CTG	821	62
	Antisense	GTC ACA TAG TGG TGC TCT TGC CA		
*SP110*Exon8	Sense	CCT TTC AAA CCT CAA GCC CT	781	64.5
	Antisense	TGG CTT CCC ATT GCA TTT A		
*SP110*Exon11	Sense	TGA GCA AGA CAG ACA CAA ATC C	765	62
	Antisense	CAA TCC TGC AAA TGT GTC CA		
*SP110*Exon14	Sense	TGG AAA GGT AGA AGG CCA CA	797	62
	Antisense	CCA TTC CTT TCT CCT TCC ATT T		
*SP110*Exon15	Sense	CCA CAG GGT CAG CAT TCG TG	397	59
	Antisense	CTG TCC AGG GAA TAG CAT ACT AG		
*SP110*Exon15	Sense	CAG TGT GAT GTC TGG GTA CAG	460	61
Alternative	Antisense	GGT CTT GCT ATG TTG CCT AGG		
*SP110*Exon17,18	Sense	CTT GGA CAC TCT GGG AGG TG	789	66.7
	Antisense	ACC AGG TAG TCC CTC TCC AGA		

A1.5 Genotyping of Polymorphisms

There were a variety of genotyping methods that were used to assess the various genes. The methods used include; Restriction enzyme digestion (for genotyping *CCR5*, *CCR2*, *SDF1* and *VDR*), Amplification refractory mutation system (ARMS; for genotyping RANTES polymorphisms), Template-directed Dye-terminator Incorporation with Fluorescence Polarization (TDI-FP; for genotyping of *DC-SIGN*) and SNaPshot (for genotyping *SP110*). Taqman, a hybridisation method which requires no post-PCR processing or label-separation steps was also used in the analysis of the *DC-SIGN* gene but was not conducted as part of this thesis, so will not be discussed. The ARMS method is an oligonucleotide ligation technique; TDI-FP and SNaPshot are both primer extension techniques, which is also known as mini-sequencing. All are frequently used for genotyping of single nucleotide changes in the genome [334]. In addition two insertion-deletion (indel) polymorphisms (*CCR5*Δ32 and *SP110*indel) were also genotyped by band separation on agarose gels. The repeat length variation in the neck region of *DC-SIGN* and *L-SIGN* were also genotyped in the same manner at the Pasteur Institute, France.

The principle of primer extension for the analysis of PCR products, is that a primer is designed so that the 3'end hybridises one base prior to the 5'end of the polymorphism. This does have its limitations as the design of a primer is confined to the sequence surrounding the polymorphism, which does not always allow for annealing of a primer.

A1.5.1 SNaPshot ® Multiplex System

SNaPshot ® multiplex system (Applied Biosystems) is a primer extension technique that can interrogate up to ten polymorphisms between one and ten templates in a single reaction [40, 265, 355]. SNaPshot was done on the ABI-Prism® 3130xl DNA analyzer platform for the genotyping of polymorphisms in the *SP110* gene (Chapter 5).

A1.5.1.1 Primer design for SNaPshot (primer extension method)

The design of SNaPshot primers requires them to be designed so that they are immediately adjacent to the polymorphism of interest. The correct primer design can save expense, time and improve quality of data. Particularly when a multiplex SNaPshot reaction is used, correct design is vital as there is always the risk of unforeseen cross reactions, binding of primers and hybridisation with the incorrect sequence. To reduce the chance of this happening there are a number of steps that were taken when designing SNaPshot primers for the analysis of the *SP110* gene.

The SNaPshot primers were initially designed as standard primers (section 2.3.1), and checks for hairpins, dimers and BLAST analyses were done. Not only must the primers pass these checks individually but also when combined, particularly avoiding dimerisation with each other and binding to the incorrect fragment. Dimers where the bases were complementary at the extreme 3' end of two primers were avoided as they could reduce amplification efficiency and this could result in spurious or incorrect results. If a primer failed any checks, the complementary strand was used as a template for redesign. Primers that were intended to run in a multiplex were designed to have equal or similar annealing temperatures.

Following completion of primer design for the polymorphisms of interest, compatibility of the primers for a SNaPshot multiplex was assessed by conducting a Local BLAST in Bioedit version 7.0.5.3 [120]. This required the construction of a nucleotide database file with FASTA reports for all the fragments amplified. (Open the program 'Bioedit', 'Accessory application' in the toolbar, then selection of 'BLAST' and subsequently 'Create a local nucleotide database file'). A Local BLAST is then conducted on this database (Open the program 'Bioedit', 'Accessory application' in the toolbar, then selection of 'BLAST' and subsequently 'Local BLAST'), with a query done on the SNaPshot primers designed for the multiplex. Each primer should only match the fragment where the polymorphism of interest is situated.

Analysis of the polymorphisms by SNaPshot was done by two assessments; the electrophoretic mobility (run length) of the primers, and detection of the dye present. To ensure that the primers have varying lengths and therefore different electrophoretic mobility, 5' non-homologous tails of neutral sequence (dC) or (dGACT) were added to the 3' end in order to act as mobility modifiers (Table A1.2). This is recommended by the manufacturers of the SNaPshot assays (Applied Biosystems) and has been verified in various studies [40, 265].

Table A1.2: Extension primers used in the genotyping by SNaPshot of seven polymorphisms in the *SP110* gene.

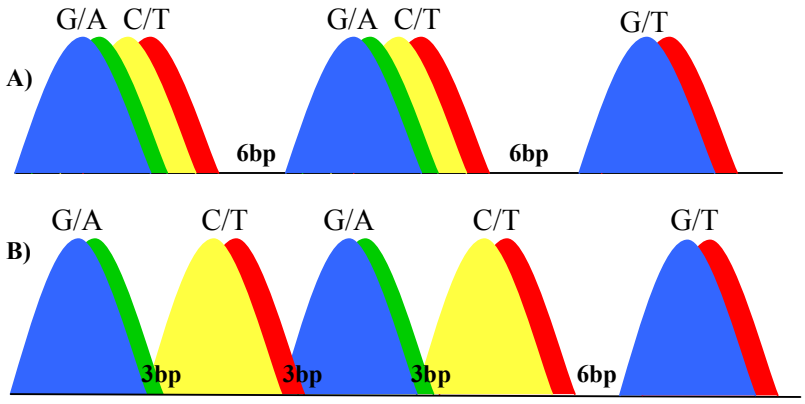
***SP110* SNP**	**Orientation**	**SNaPshot Primer Sequence 5' → 3' [expected allele call]**	**Primer conc. in the master mix for SNaPshot (µM)**
rs3948463	Sense	AGTTGCACCTTCTGCAGGAT[g/a]	0.2
rs3948464	Sense	ATGTGCCCGAAAGTCCAGAT[c/t]	0.5
rs1135791	Sense	CCCCCCGGAATATACGTTGTGAGGAA[c/t]	0.5
rs1365776	Antisense	CCCCCCCCCCCCTTGTACTCTCATCTTACCTC[c/t]	0.5
rs2114592	Antisense	CTGACTGACTGACTGACTGACTTCTATTGCTTTGACGTCTATGT[g/a]	0.5
rs35495464	Sense	ACTGACTGACTGACTGACTGACAACTTCTCTGCTGCGG[c/t]	0.3
rs41541917	Antisense	CTGACTGACTGACTGACTTCTGTTCTCCAGCTTCCTGA[g/a]	0.2

To achieve adequate separation of the products during electrophoresis, the primers are required to be separated by 4 to 6 bp. Polymorphisms with complimentary pairs could be the same length (Figure A1.1A), but in hindsight it would have been more beneficial, for analysis purposes, to have them at slightly different lengths (Figure A1.1B).

Figure A1.1: Spacing of the SNaPshot primers.

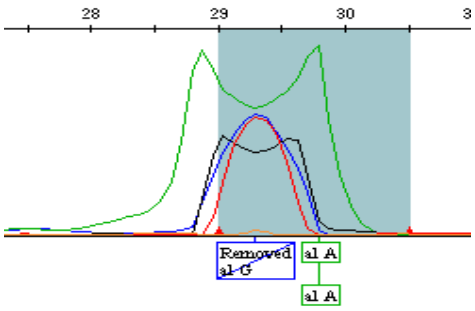
A: As advised and used in this thesis for *SP110* gene polymorphisms.

B: Separation that in hindsight would have eased analysis.



This suggested spacing would have avoided the incorrect automated annotation when there was spectral pull up (Figure A1.2), a phenomenon which occurs when there is overloading and the [F]ddNTP extended primer calling 'pulls' up all other fluorescent colours as well.

Figure A1.2: An example of spectral pull up. This occurs when an A allele (green) is called, but because of spectral pull up the G allele (blue) was also called, which was manually removed. If the G bin had not fallen within the A allele bin, this would not have occurred.



When all SNPs have the same alleles present, a complimentary pair can be created by designing the primer on the antisense strand, as the reverse complementary base will then be called, i.e. the dyes will be different. Is important to note when a primer has been designed on the antisense strand as then the results will be the reverse complementary of the expected polymorphism. The advice of Applied Biosystems is that primers over 30 oligonucleotides in length should be high performance liquid chromatography (HPLC) purified, however all primers were HPLC purified in this study.

Despite primer design being done correctly, there was no guarantee that they would work successfully in a SNaPshot multiplex, so optimisation was carried out. This required testing individual primers in a singleplex and then in a multiplex. Samples with a known genotype run in a singleplex, allowed us to determine the electrophoretic mobility for each primer, before combining them in a multiplex. Heterozygotes were preferably run as this allowed for the identification of the run length for each allele and therefore we were able to determine the bins.

The optimisation of the SNaPshot multiplex was done on samples that had been sequenced. The fragments that had been amplified for sequencing in singleplexes (seven separate PCR reactions) were pooled for the optimisation of the SNaPshot multiplex. Initially the concentration of each primer was 0.5µM, but this was adjusted according to the intensity of the fluorescence detected during the optimisation runs. Ideally all fluorescent peaks should be of a similar height and decreasing or increasing primer concentrations in the multiplex this can be corrected (Table A1.2).

A1.5.1.2 Preparation of the PCR template for SNaPshot

Before the SNaPshot reaction can be run, an amplification PCR is required to amplify the fragments of the *SP110* gene containing the polymorphisms of interest. The seven polymorphisms that had been optimised in the SNaPshot reaction were found in seven different fragments. To reduce the number of amplification reactions from seven, which was time consuming and required 7µl of DNA, a multiplex amplification PCR was optimised. When

possible the sequencing primers were used, however these PCR products had similar lengths and varying annealing temperatures (Table A1.1). This made it difficult to optimise one multiplex reaction as the bands would overlap when run on an agarose gel. Initially there were seven single amplifications, but with redesign of primer SP110 Exon 4 and 5 this was reduced to four, where one multiplex PCR contained four primer sets (Table A1.3).

Table A1.3: Primers used in amplification of samples for SNaPshot of the *SP110* gene.

Primer Name	Orientation	Sequence 5' → 3'	Fragment Length
*SP110*Exon4,5B*	Sense	GGT ACA AAC CCA AAC TCA ACT TTT AT	1440
	Antisense	ACT AGT GTG AGT GTT ACG CAG GTT AC	
*SP110*Exon7*	Sense	GCA GAG CTT TAT ATG TCT TTG CTG	821
	Antisense	GTC ACA TAG TGG TGC TCT TGC CA	
*SP110*Exon8	Sense	CCT TTC AAA CCT CAA GCC CT	781
	Antisense	TGG CTT CCC ATT GCA TTT A	
*SP110*Exon11*	Sense	TGA GCA AGA CAG ACA CAA ATC C	765
	Antisense	CAA TCC TGC AAA TGT GTC CA	
*SP110*Exon14	Sense	TGG AAA GGT AGA AGG CCA CA	797
	Antisense	CCA TTC CTT TCT CCT TCC ATT T	
*SP110*Exon15*	Sense	CCA CAG GGT CAG CAT TCG TG	397
	Antisense	CTG TCC AGG GAA TAG CAT ACT AG	
*SP110*Exon17,18	Sense	CTT GGA CAC TCT GGG AGG TG	789
	Antisense	ACC AGG TAG TCC CTC TCC AGA	

* in the multiplex PCR

Amplification of *SP110* Exon 8, 14 and 17 were done in singleplex reactions using 0.13µl of Taq with 2.5µl of 10x PCR buffer, 2µl of dNTPs, 13µl of distilled water and 0.35µl of each primer (forward and reverse), giving a total volume of 19µl per sample before 1µl of DNA template was added. The final 20µl reaction mix was amplified according to the following program: 1 cycle at 94°C for 15 minutes followed by 35 cycles at 94°C for 30 seconds, annealing temperature in Table A1.1 (°C) for 30 seconds and 72°C for 45 seconds, then 1 cycle at 72 °C for 10 minutes and finally a 15°C hold step.

The multiplex PCR consisted of the four primer sets; *SP110*Exon11, *SP110*Exon7, *SP110*Exon4,5B and *SP110*Exon15 (Table A1.3). During optimization we adjusted the primer concentrations to get similar quantities of product for each fragment. The final volumes per primer were 0.35µl for *SP110*Exon11 and *SP110*Exon7, 0.70µl for *SP110*Exon4,5B and 0.18µl for *SP110*Exon15 (final concentrations were therefore 0.175µM, 0.35µM and 0.09µM respectively). The amplification was done as above but with an annealing temperature of 62°C.

Prior to SNaPshot analysis, quality control was done by running samples on a 1% agarose gel at 150V for 40 minutes to ensure amplification of all fragments. The expected fragment sizes are in Table A1.2. Once amplified, the four separate PCR products of the same sample were mixed in equal volumes into one sample to use in the SNaPshot reaction. A plate format was used and as quality control, each plate contained a negative control and a sample that had previously been sequenced.

Purification was required to remove primers, dNTPs, enzymes and buffer. Purification for SNaPshot was done with a Shrimp Alkaline Phosphatase (SAP) (Promega Cooperation, USA) and *ExoI* (USB, USA) treatment (not ExoSAP-IT). The use of this method reduces the background level of 'noise'. In the SP110 project it was found that 1.5 times the usual SAP and *ExoI* ratio worked more effectively. 1µl of PCR product was used and 0.5µl (stock 1U/ µl) of SAP and 0.02µl (stock 5U/µl) of *ExoI* was added. The SAP-*ExoI* mixture was thoroughly mixed with the PCR template and incubated at 37°C for 1 hour. To inactivate the enzymes an additional incubation step at 75°C for 30 minutes was done. Subsequently the samples were kept on ice.

A1.5.1.3 SNaPshot reaction

The SNaPshot reaction incorporates the [F]ddNTPs which are present in the SNaPshot Multiplex Ready Reaction Mix (Applied Biosystems, Part Number: 4323154).

The following master mix was prepared (per sample): 5µl of SNaPshot Multiplex Ready reaction mix, 1µl of the pooled SNaPshot primers (Table A1.2) and 1µl of deionized water. A volume of 7µl of the master mix was aliquoted into plates and 3µl of purified PCR products was added, giving a total volume of 10µl.

A1.5.1.4 Thermal cycling and SNaPshot Post-extension treatment

The plates were placed in a GeneAmp 9700 thermal cycler (Applied Biosystems) and the following cycles were repeated 27 times: 96°C for 10 seconds, 58°C for 5 seconds, 60°C for 30 seconds, followed by a hold at 4°C. Instead of 58°C a temperature of 50°C is usually used, but with this primer combination a higher temperature worked better.

The post-extension treatment was done by adding 1 Unit of SAP (Promega, Catalogue number: M8201) to the SNaPshot post-extension product. This removed [F]ddNTPs that would otherwise co-migrate with the fragments of interest and could cause interference. Once mixed thoroughly with the SAP it was incubated at 37°C for 1 hour. Deactivation of the enzyme required a further incubation at 75°C for 30 minutes. Samples were then stored at 4°C.

A1.5.1.5 Preparation of Sample for Electrophoresis on the ABI PRISM 3130 xl DNA analyzer

The ABI Prism 3130xl Genetic Analyzer was set up with a 36cm capillary array and POP-4 polymer (Applied Biosystems, Part number: 4352755) with dye set E5 for analysis by SNaPshot.

Sample preparation requires genetic analysis grade Hi-Di formamide (Applied Biosystems part number: 4311320), SNaPshot PCR product and GeneScan-120LIZ size standard (Applied Biosystems, Part number 4324287). 10µl of Hi-Di formamide is added to each well of a new 96-well optical reaction plate. 1µl of SNaPshot post-extension product is added to each well, along with 0.4µl of GeneScan-120LIZ size standard. The septa sealed plate was then vortexed and spun

down. Denaturation of the samples was done by placing them at 95°C for 5 minutes in a 9700 thermal cycler (Applied Biosystems). As soon as the 5 minutes was completed the samples were placed on ice, to hinder re-annealing of the product. The samples were then ready to load on the ABI Prism 3130xl Genetic Analyzer. The protocol used was the preset run module "SNP_E5_36_POP4_run". In this preset run SNaPshot products were injected electronically for 22 seconds at 2kV and electrophoresed at 15kV and 5μA at 60°C.

A1.5.1.6 Analysis of SNaPshot results in GeneMapper

Analysis of SNaPshot results was done on the GeneMapper software version 3.7 (Applied Biosystems). A 'bin' was created for each allele and is defined as the fragment size or bp range and dye colour that define an allele. A bin set (one set per polymorphism) was set up for each primer and included a separate bin for each allele (Figure A1.3). During optimisation, bin sets were determined for all polymorphisms to allow for automated annotation of the fluorescent peaks (Figure A1.4). This is why SNaPshot primers are required to be 4 or 6 bps apart, so that bins which have the same dye colour (ie incorporation of the same base) do not overlap. All primers were designed to be 6bp apart to lessen the chance of this occurring.

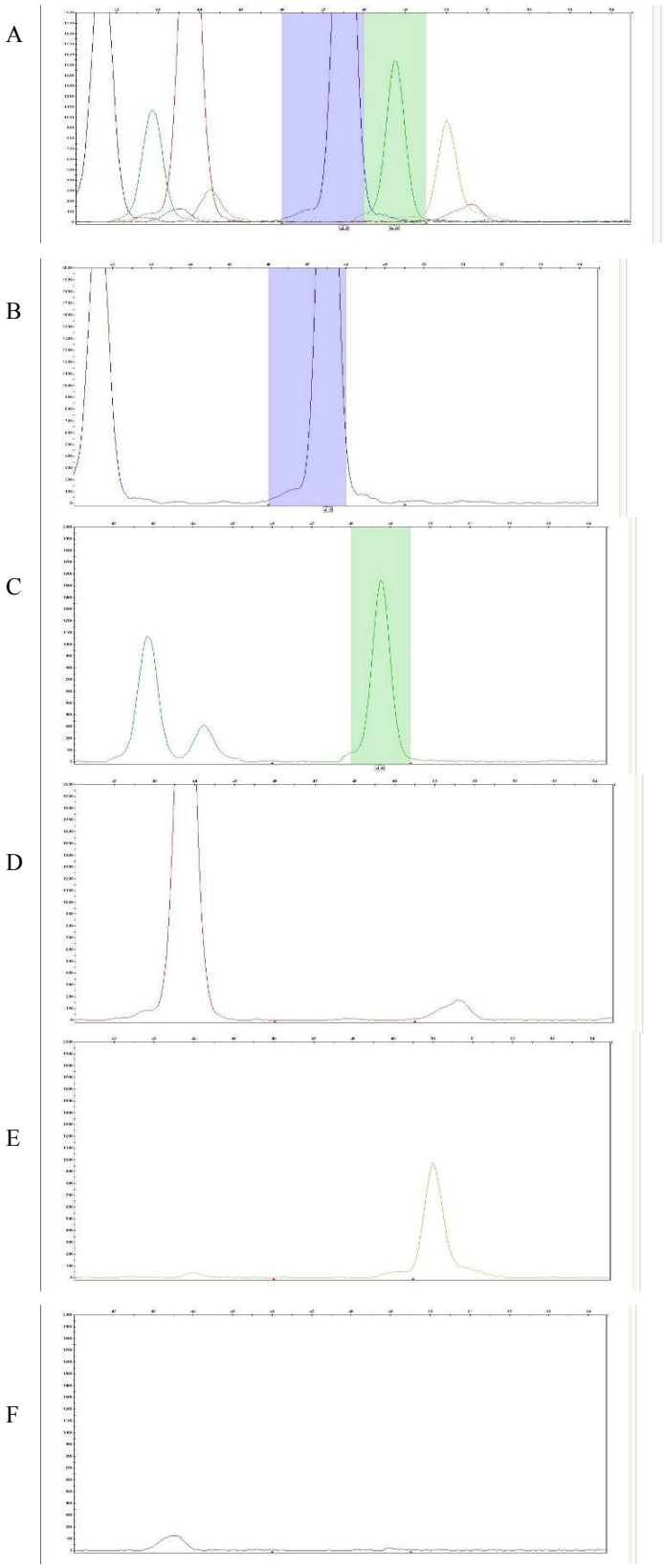


Figure A1.3:

Example of an electropherogram sample run for SNaPshot analysis of SNP rs2114592 (G/A) in the *SP110* gene.

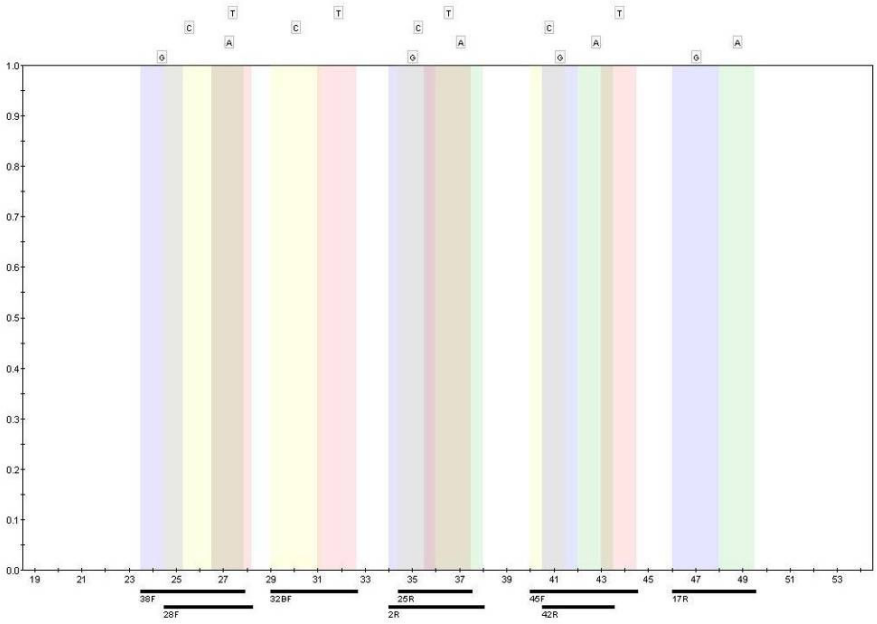
A) All dye colours with bins for alleles in rs2114592

Dyes colours separated:

B) dye blue (G) with bin
C) dye green (A) with bin
D) dye red (T). The peak seen is for another SNP, which will have a different bin for it to be called.
E) dye yellow (C) (appears black on software for easier reading). The peak seen is for another SNP.
F) dye orange- the size standard GeneScan-120 Liz

Fluorescent peaks that fall out of the bin for this SNP, such as in E), are for different primers (i.e. polymorphisms) and are only shown here for demonstration purposes.

Figure A1.4: Bin sets for the analysis of 8 polymorphisms in the *SP110* gene by SNaPshot in the GeneMapper Software version 3.7 (Applied Biosystems).



The size of the primer detected will not always be as expected (Table A1.4). This is because of differences in the electrophoretic mobility of the primers, mainly determined by length, sequence and the dye incorporated [265]. As long as the same colours (ie base calling) do not coincide, this does not inhibit analysis.

Table A1.4: *SP110* SNaPshot primer expected and observed run lengths as determined by GeneMapper.

SP110 SNP	Expected run length	Observed run length, allele 1	Observed run length, allele 2
rs3948463	20	23.5-27.9	23.5-25.3
rs3948464	20	24.5-26.5	26.5-28.2
rs1135791	26	29-32.3	29-31.1
rs1365776	32	34.4-36	35.5-37.5
*rs2114592	44	46-49.5	46-48
rs35495464	38	40-44.5	40-41.5
rs41541917	38	40.5-43.5	40.5-42

A negative control was run at all times, both in the optimization with singleplex and later and all plates had at least one negative control sample with no template. Apart from being able to detect contamination, the negative control would reveal whether primers are annealing to each other. The automated genotype calling was checked by the technician and subsequently by the researcher and discrepancies were further investigated.

As mentioned previously, there is the possible problem of spectral pull up (Figure A1.2). If a bin, for a different fluorescent colour, was overlapping another fluorescent bin and spectral pull up occurred this could cause incorrect annotation. However the software flagged samples with spectral pull up and manual annotation of the bin was done.

A1.5.2 Restriction enzyme digestion

Restriction enzyme digestion, also known as allele specific restriction enzyme array (ASREA), makes use of an enzyme that cuts at a specific location of a PCR amplified fragment. The presence/absence of a SNP will often determine if the recognition sequence of the enzyme exists. It is advisable to have an internal control cut site (where the enzyme always cuts) to ensure that enough enzyme was added and the conditions were correct for a reaction to take place. Appendix 1 lists the enzymes and their cut sites that were used for the analysis of the polymorphisms *CCR5*-59029, *CCR2*-64I, *SDFI* 3'A, *VDR FokI*, *VDR ApaI* and *VDR TaqI*.

A1.5.2.1 Amplification of the CCR5*,* CCR2*, and* SDFI *fragments for restriction enzyme digestion*

The method of analysis for *CCR5*-59029, *CCR2*-64I and *SDFI* 3'A was as reported by Kristiansen *et al* [160], with some modifications. *CCR2*-64I required antisense primer (5' GAG CCC ACA ATG GGA GAG TA 3'). Amplification of *CCR2*-64I was done using 0.13µl of Taq with 2.5µl of 10x PCR buffer, 2µl of dNTPs, 18.7µl of distilled water and 0.35µl of each primer (forward and reverse) (Table A1.5), giving a total volume of 24µl per sample before 1µl of DNA template was added. The final 25µl reaction mix required an annealing step of 60°C. 10µl of PCR product was aliquoted out for digestion with 1µl of the appropriate enzyme.

Table A1.5: Primers used in the PCR amplification of *CCR5*, *SDF1*, *CCR2* and *VDR* fragments for restriction enzyme digestion.

Polymorphism	Orientation	Sequence 5' → 3'	Annealing temperature (°C)
CCR5-59029	Sense	TGG GGT GGG ATA GGG GAT AC	Multiplex 66
	Antisense	TGT ATT GAA GGC GAA AAG AAT CAG	
SDF1 3'	Sense	CAG TCA ACC TGG GCA AAG CC	Multiplex 66
	Antisense	AGC TTT GGT CCT GAG AGT CC	
*CCR5*Δ32*	Sense	CTT CAT CAT CCT CCT GAC AAT CG	60.3
	Antisense	GAC CAG CCC CAA GTT GAC TAT C	
CCR2-64I	Sense	TTG TGG GCA ACA TGA TGG	Multiplex 66
	Antisense	GAG CCC ACA ATG GGA GAG TA	
VDR-FokI	Sense	AGC TGG CCC TGG CAC TGA CTC TGC TCT	67.5
	Antisense	ATG GAA ACA CCT TGC TTC TTC TCC CTC	
VDR-ApaI	Sense	same as *TaqI*	67
	Antisense	same as *TaqI*	
VDR-TaqI	Sense	GGG ACG ATG AGG GAT GGA CAG AGC	67
	Antisense	GGA AAG GGG TTA GGT TGG ACA GGA	

*(used in indel genotyping, Section A1.5.4)

The multiplex PCR for CCR5-59029, CCR5Δ32 and SDF1 3'A, was done using 0.13µl of Taq with 2.5µl of 10x PCR buffer, 2µl of dNTPs, 16.7µl of distilled water and 0.35µl of each forward and each reverse primer (Table A1.5), giving a total volume of 14µl per sample, before 1µl of DNA template was added. The final 24µl reaction mix was amplified with an annealing temperature of 66°C. 10µl aliquots were taken for digestion with 1µl of the appropriate enzyme, except CCR5Δ32 (Section A1.5.4) insertion/deletion detection.

A1.5.2.2 Amplification of the VDR *fragments for restriction enzyme digestion*

Primers used are listed in Table A1.6 [391]. Samples were amplified using an annealing temperature of 67°C. The *FokI* fragment was amplified by using 0.13µl Taq with 2.5µl of 10x PCR buffer, 2µl of dNTPs, 13.5µl of distilled water and 0.35µl of each primer, giving a total volume of 19µl per sample, before 1µl of DNA template was added. Conversely the *TaqI* fragment (including *ApaI* fragment), per sample, used 0.16µl of Taq with 3.2µl of 10x PCR buffer, 2.5µl of dNTPs, 17.3µl of distilled water and 0.44µl of each primer, giving a total volume of 24µl before 1µl of DNA template was added.

Table A1.6: Restriction enzyme digestion and the expected fragment lengths that would be visible on a gel.

			Fragment lengths in base pairs		
Polymorphisms investigated	**Restriction enzyme**	**Digestion temp. (°C)**	**wild type**	**heterozygote**	**homozygote**
CCR5-59029	*Bsp1286I*	37	453	453,408	408
SDF1'3A	*MspI*	37	201, 101	201,101,302	302
CCR2-64I	*BsaI*	60	129	(21),108,121	(21),108
VDR-FokI	*FokI*	55	270	270, 208, 62	208,62
VDR-ApaI	*ApaI*	37	716	716, 485, 231	485, 231
VDR-TaqI	*TaqI*	67	512, 204	512, 311, 204, 201	311, 204, 201

A1.5.2.3 Enzyme digestion

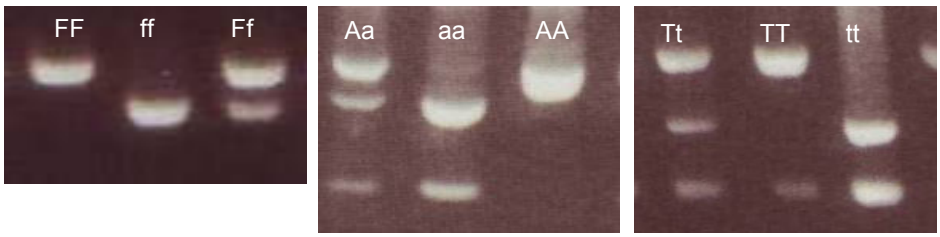
Enzymatic digestion was described previously [37, 160, 285, 391, 399] with the following conditions; The *Bsp1286I* (Amersham Biosciences, USA) (*CCR5*-59029) and the *MspI* (Promega, USA) (*SDF1* 3'A) digests were incubated at 37°C for at least 1 hour. *CCR2*-64I PCR products were digested by *BsaI* (New England Biolabs, USA) at 60ºC for at least 3 hours. Analysis of *CCR5*Δ32 does not require restriction enzyme digestion and the fragment does not contain a cut site for any of the restriction enzymes used, as a mismatch was added into the reverse primer to prevent the amplification of a non-mutation specific *BsaI* endonuclease cut site [160]. Therefore the *CCR5*Δ32 results could be read in conjunction with another digested sample (Section A1.5.4). All *VDR* fragments were digested overnight at 65°C, 37°C and 55°C with *TaqI*, *ApaI* and *BseGI* / *FokI* restriction enzymes (Fermentas, Canada), respectively. The restriction enzymes, their recognition cut sites and incubation temperatures are listed in Table A1.7.

Table A1.7: Alphabetical list of the restriction enzymes, their recognition cut sites and incubation temperatures used for the analysis of SNPs in the genes *VDR*, *CCR2*, *CCR5* and *SDF1*.

Restriction enzymes	**Polymorphism Analysed**	**Recognition site**	**Incubation temperature**
ApaI	*VDR ApaI*	5' – G GGCC ↓ C - 3' 3' – C ↑ CCGG G – 5'	37°C
*Bsa*I	*CCR2*-64I	5' – GATNN ↓ NNATC – 3' 3' - CTANN ↑ NNTAG – 5'	60 ºC
Bsp1286I	*CCR5*-59029 A/G	5' – G (A/G/T)GC(A/C/T) ↓ C – 3' 3' – C ↑ (T/A/C)CG(T/G/A) G – 5'	30ºC
FokI	*VDR FokI*	5' – GGATG NN ↓ – 3' 3' – CCTAC ↑ NN – 5'	55°C
MspI	*SDF1*-3'A	5' – C ↓ CG G - -3' 3' – G GC ↑ C – 5'	37 ºC
TaqI	*VDR TaqI*	5' – T ↓ CG A – 3' 3' – A GC ↑ T – 5'	67°C

The enzymatically digested PCR aliquots were separated on a 2 or 3% Agarose gel (Hispanagar, Spain), using SB buffer [42] containing 5µg ethidium bromide / 100ml. Gels were run at 180V for an hour and viewed under UV light (Figure A1.5). Resulting fragment sizes are listed in Table A1.6. For the *VDR* polymorphisms a lower case ('f', 'a' or 't') was used to indicate the presence of an endonuclease site.

Figure A1.5: Gel electrophorectic analysis after restriction enzyme digestion of the *VDR* gene: genotyping of the SNPs *FokI*, *ApaI* and *TaqI* (left to right).



A1.5.3 Amplification Refractory Mutation System (ARMS)

ARMS [235] is an oligonucleotide ligation technique that has a primer designed for both of the alleles present at a polymorphic site. The technique generally requires two reactions (one for each allele specific primer) but has been modified to be done in a single reaction [89].

Genotyping of the *RANTES* polymorphisms -403, -109, -28 and +1092 was done by means of a two-tube ARMS and was done by Hanno Nel and Erika Truter of our Division Molecular Biology and Human Genetics. Amplification of the regions surrounding the polymorphism of interest was done by external primers and an allele specific primer (Table A1.8). A mismatch was introduced in the allele specific primer to destabilise non-specific allele binding where the non-specific primer partially primes to the DNA. The SNPs are named according to the numerical system used by Liu *et al* [176] that considered +1 as the first nucleotide of the RANTES mRNA.

Table A1.8: Primers used for amplification refractory mutation system (ARMS) analysis of the *RANTES* polymorphisms.

Polymorphism	Orientation	Sequence 5' → 3'	Annealing temp. (°C)
RANTES-403	Sense	GGT CGC CTT AGC AAG TAA ATG G	60
	Antisense	GTC CAC GTG CTG TCT TGA TC	
	Allele Specific	CAT GGA TGA GGG AAG G<u>C</u>G [G/A]	
RANTES-109	Sense	GGT CGC TTA GCA AGT AAA TGG	58
	Antisense	GTC CAC GTG CAG TCT TGA TC	
	Allele Specific	CGG AGG CTA TTT CAG TTT T<u>G</u>T [C/T]	
RANTES-28	Sense	ACT CTA GAT GAG AGA GCA GT	50.4
	Antisense	GAC AGT ATT CAT GCT ACA GTT G	
	Allele Specific	CCT AGG GAT GC C<u>A</u>T [G/C]	
RANTES+1092	Sense	AGA GCT TCT GAG GCG CTG	58
	Antisense	CGG GAG GAA ATC AAG AGT C	
	Allele Specific	GGA AGC TTA AGA GTG C<u>T</u>G [C/T]	

[ ] indicate the two possible oligonucleotides that were added to the ends of the allele specific primers.

Bases are underlined where a mismatch was introduced to destabilise non-specific allele primer binding.

PCR amplification was done with the following protocol using an Eppendorf Mastercycler Gradient PCR Machine (Eppendorf, Hamburg, Germany) and a Hybaid Touchdown Thermocycler system. The final reaction contained 1x Buffer A, 300µM dNTPs each, 3-4mM $MgCl_2$, 25-5pmol primers and 1-2U of Taq. Specific concentrations of $MgCl_2$ and primers for each ARMS-PCR analysis are shown in Table A1.9.

Table A1.9: Magnesium Chloride and primer concentrations for the set up of ARMS analysis of the *RANTES* gene.

RANTES **Polymorphism**	**$MgCl_2$ (mM)**	**Allele specific primer (pmol)**	**Shared primers (pmol)**	**Control primers (pmol)**
-403A/G	4	7-8	4.5	1.6
-109T/C	2.5	10-12	10	4
-28C/G	4	2.8-3	5.5	0.7
+1092A/G	3	9-11	6.3	2.3

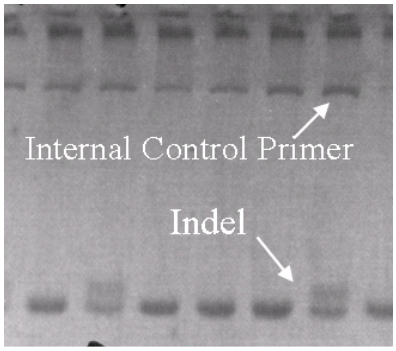
The following program was used: 1 cycle at 95°C for 5 minutes followed by 35 cycles at 94°C for 30 seconds, annealing temperature in Table A1.7 °C for 30 seconds and 72°C for 60 seconds, then 1 cycle at 72 °C for 10 minutes and finally a 4°C hold step. Per sample this was 0.3µl of Taq with 3µl of $MgCl_2$, 0.6µl of dNTPs, 38µl of distilled water, 0.5µl of the allele specific and reverse primer and 0.2µl of the forward primer.

A1.5.4 Insertion/deletion detection

If there is an insertion/deletion (indel) present in an amplified PCR product of a heterozygote there will be two fragments of different lengths. Depending on the size of the indel and the amplified fragment, the gels used for analysis should be at a sufficient concentration to allow for adequate separation.

Through sequencing, a four base pair indel was identified in intron 10 of the *SP110* gene. Amplification was done using primers *SP110*indelF 5'-TGA GCA AGA CAG ACA CAA ATC C-3' and *SP110*R 5'- CTT CTC ATG TTC CCA GCC C-3'. The PCR reaction, per sample, was 0.07µl of Taq with 1.25µl of 10x PCR buffer, 1µl of dNTPs, 6.7µl of distilled water and 0.18µl of each primer, giving a total volume of 9.5µl before 0.5µl of DNA template was added and amplified with an annealing temperature of 58°C. 5µl of PCR product was run on a 3% gel at 120V for 120 minutes. The fragment lengths that could be seen were 216 and/or 212bp in length (Figure A1.6).

Figure A1.6: Insertion deletion in the *SP110* gene, intron 10. Fragment run on a 3% agarose gel.



Amplification of the indel *CCR5*Δ32 was done using the primers in Table A1.5, where a multiplex PCR was performed to analyse other polymorphisms by restriction enzyme digestion (Section A1.5.2). Analysis of *CCR5*Δ32 did not require restriction enzyme digestion [160]. The presence of the 32bp deletion, and the length of the fragment meant that the *CCR5*Δ32 could be detected on a gel in conjunction with another digested sample (Section A1.5.2).

A1.5.5 Template-directed Dye-terminator Incorporation with Fluorescence Polarization detection (TDI-FP) using Victor (Perkin Elmer) (primer extension method)

The AcycloPrime™-FP SNP Detection Kit (Perkin Elmer, USA) utilizes Template-directed Dye-terminator Incorporation with Fluorescence Polarization (TDI-FP) [97, 164] technology with the use of the Victor instrument (Perkin Elmer, USA). This procedure involves primer extension, similarly to the SNaPshot technique, and primer design should follow the same principles (Section A1.5.1). However, unlike SNaPshot, this procedure was used to analyze one SNP per reaction. The kits contain [F]ddNTPs which are labeled with either the fluorescent dyes TAMRA

or R110. The detection kit required was determined by the polymorphism present and the orientation of the extension primer. The various combinations of polymorphisms and the dye labelling the base are listed in Table A1.10.

Table A1.10: AcycloPrime ™ -FP SNP detection kit combinations of polymorphisms that can be detected with fluorescently labeled ddNTPs.

R110/TAMRA
G/A
G/C
G/T
C/A
A/T
C/T

For analysis using TDI-FP, it was necessary to do an amplification PCR and then a second extension PCR using the extension/Victor primer that incorporates the [F]ddNTP at the site of interest. Polymorphisms were selected after the sequencing (Section 2.4) of 28 randomly chosen samples and determination of tagging SNPs (Section 2.6.3).

The amplification PCR used 0.1µl of Taq, 0.15µl of $MgCl_2$ at a concentration of 25mM, 2.15µl of distilled water, 0.6µl of dNTPs at a concentration of 1.25mM and finally 0.25µl of each amplification primer (Table A1.11) at concentrations of 3µM. This gave a total volume of 4µl per sample to which 1µl of DNA template was added. Template DNA concentration needed to be 10ng/µl as this was vital to ensure similar PCR yields for the extension step, and was important for the fluorescent polarization intensity. We ensured correct concentrations by diluting all samples after conducting spectrometry readings. One possible way to correct for irregular concentrations of DNA is to first do whole genome amplification on the samples prior to testing. This ensures that there is sufficient sample for all the tests and that the DNA concentration is constant in all samples.

The PCR program used for amplification was: 94°C for 10 minutes followed by 40 cycles of 94°C for 15 sec, annealing temperature from Table A1.11 °C for 30 seconds, 72°C for one minute, followed by a step at 72°C for 10 minutes and a hold at 15°C.

Random samples were chosen and run on a 2% agarose gel to check for amplification. Then a PCR Clean-Up was done using enzymes supplied in the AcycloPrime™-FP SNP detection kit, where 0.2µl of PCR Clean-up enzyme and 1.8µl of buffer was added to 2µl of PCR product. Subsequently the primer extension with the [F]ddNTPS could be performed, as in Table A1.12.

Table A1.11: Primers used in the amplification and primer extension for analysis by Victor to genotype *DC-SIGN* polymorphisms.

					Victor PCR			
***DC-SIGN* SNP**	**Primer F 5'→3'**	**Primer R 5'→3'**	**Allele specific Primer 5'→3'**	**Annealing temperature for amplification (°C)**	**Number of initial cycles for primer extension**	**Annealing temperature for allele specific primer (°C)**	**Orientation**	**Kit combination R110/ TAMRA**
-939	ACA CAT GTG GTA AAT GCA CA	ACC AGC AGG TGA ATG ATA AA	ACA CAC TGT AAG ATT TGA TTT T	50	25	55	Antisense	G/A
-336	GAG GAC AGC AGC AGC TCA AA	TGG CTG AGC AGT GGG ATG CTT	TCC ACT AGG GCA AGG GT	60	35	55	Antisense	G/A
2392	TAC ATG GAT GTG TGC ATG TCA G	CTG CTC CTC AGC ACT TTT GA	GAG GCT GGT CAG GGC TGG	55	40	55	Antisense	G/A
3220	TGC ATG GAT GTG TGC ATG TCA G	CAT TTT CCA GGC TCT GTC TC	GCC TAC AAG CAG AAC TTC CT	58	35	55	Sense	G/A
3838	CAG TAA GGC CAC CAG CTC AG	CCC AAT CCT AAG CCT GTT CA	GAC CTT TCC TTT CCA CAG AA	63	40	55	Antisense	G/T
4235	CAC ACA CTC AAT AGG TGG AG	CTA AAT TCC GCG CAG TCT TC	GAA TGT GAG CCT CTG TCC CC	55	25	65	Antisense	G/C

Table A1.12: Victor PCR mixture that incorporates the [F]ddNTP.

Ingredient	µl (per sample)
AcyloPol	0.05
10x Buffer	2.0
AcycloTerminator mix (R110/TAMRA)	1.0
SNP Primer V (10µM)	0.5
dH_2O	9.45
Total	13
PCR product per aliquot	7

Since fluorescent polarization was used for detection, specific plates are required and black microplates (MJ research) were used. Each plate had eight control samples to ensure that the dyes were read correctly. The controls included 2 wildtype homozygotes, 2 heterozygotes, 2 homozygotes for the polymorphism and 2 blanks. Nine 96-well micro-plates of DNA stocks were set up for the case-control analysis of DC-SIGN and L-SIGN genes (Chapter 4). The first column in the plate was left empty to allow for the control samples.

A1.5.5.1 Analysis of results using Victor and fluorescent polarization

The SNP 'macro Victor 96-well', an Excel-based workbook for data analysis and allele calling (available for download at http://lifesciences.perkinelmer.com/products/snp.asp) automatically analyzed the polarization of labelled samples and, with the use of the known control samples and the blanks, the allele-calling could be done. The TAMRA polarization and R110 polarization for each sample was plotted on an Excel graph. The dye which was incorporated at the site of the polymorphism determined where a sample fell in the plot, e.g. the bottom left cluster contains the blanks and the top right cluster contains the heterozygotes (Figure A1.7 A).

The primer extension PCR done with the commercially available detection kit is a very stable reaction and samples could be subjected to additional thermocycles if the clustering of samples was not appropriate for scoring (Figure A1.7 A and B). In addition the centrifugation of the plate also affected the reading of a sample (Figure A1.7 C and D). This was useful for repeated analysis of samples that fell outside an allocated quadrant.

Figure A1.7: SNP detection by TDI-FP using VICTOR.

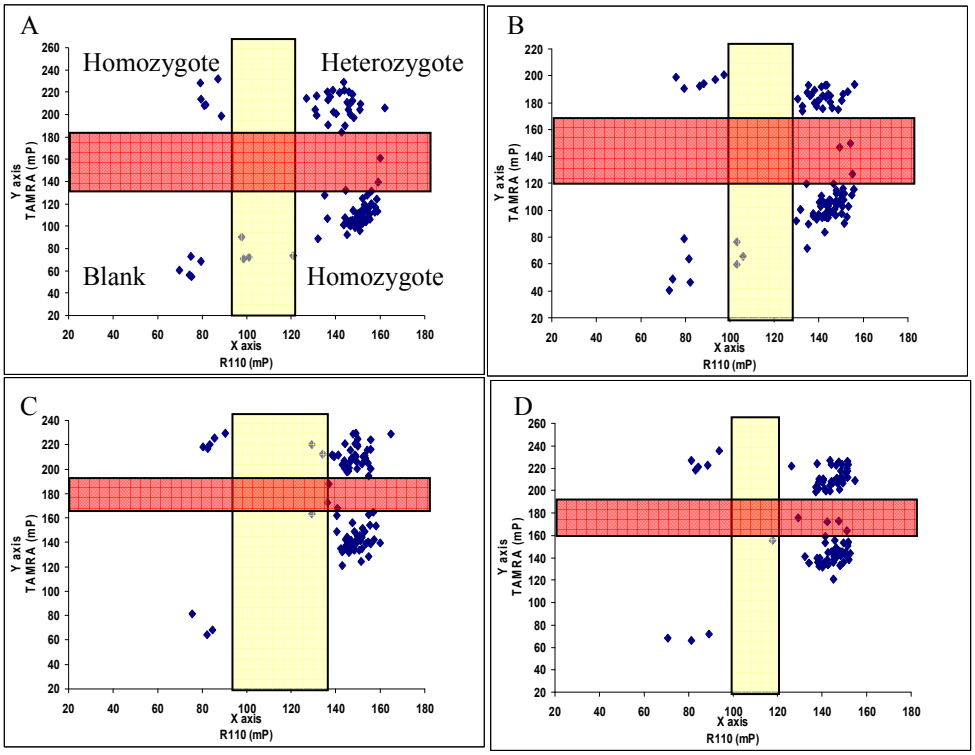
A) 25 thermocycles

B) an additional 5 cycles to plate in A), giving a total of 30 cycles

C) mild centrifugation

D) additional centrifugation on the same plate as C)

The red bar is the division of positive (above) and negative (below) for TAMRA incorporation. If a sample falls within the box, it is considered as a 'no call' result. The yellow bar applies to R110 incorporation.



A1.6 Analysis

Databases

Apart from the "EileenTB" database used for management of the case-control data, the "Surrogate Marker (SM) Database version 3" was used in this thesis. These had the relevant information on subjects and different collection procedures and ethic approvals. The SM Database was used for the VDR study (Chapter 6). All databases were managed by the software package Microsoft-ACCESS. As dual entry had not been used, considerable time was spent checking and cleaning the data as well as managing and maintaining the databases. Publicly available online databases of nucleotide sequences and polymorphism used include NCBI (http://www.ncbi.nlm.nih.gov/) and Ensembl (http://www.ensembl.org/index.html). Microsoft-Excel worksheets were used during active analysis.

A1.6.1 Hardy-Weinberg Equilibrium

The Hardy Weinberg Equilibrium (HWE) states that the segregation of two alleles in a large, randomly mating population will be such that there will be constant proportions of the three possible genotypes. The Hardy Weinberg Theorem predicts genotype frequencies AA, Aa, and aa are p^2, 2pq, and q^2, respectively, and uses a binomial equation: $p^2 + 2pq + p^2 = 1$, where p and q are the proportions of the alleles present. It is primarily used as a data quality check. When an allele falls out of HWE it could be for a number of reasons such as poor genotyping (primarily occurring when heterozygotes are genotyped as homozygotes) inbreeding, population stratification, or selection, however it could be a sign of disease association particularly if only the case population is out of HWE. A locus that falls out of HWE should not be immediately discarded, but carefully investigated to determine the cause [396].

The binomial equation will calculate the expected frequencies of the genotypes from the known frequencies of the alleles. The χ^2 statistic calculated using the equation $\chi^2=\Sigma(\text{observed-expected})^2/\text{expected}$, should be less then 3.84 (one degree of freedom), not to be significantly out of HWE. This is also known as the Pearson's χ^2 analysis which tests for deviations from HW between the observed and expected frequencies of the polymorphism. HWE was checked in this population for all genotypes tested.

A1.6.2 Testing for an Association

To test for an association with tuberculosis, genotype frequencies were compared between cases and controls. The frequency of the genotypes, alleles, minor allele containing genotypes versus the remaining, haplotypes and diplotypes were also compared between cases and controls.

GraphPad Prism version 4.03 and version 5 were used for statistical analysis in the case-control studies. Associations were tested by construction of a 2xn contingency table. For genotypes this was a 2x3 table containing the number of individuals with each genotype in the groups of cases and controls. Chi-square (χ^2) tests were subsequently carried out with a confidence interval of 95%. In addition, a dominant/recessive model was constructed where the genotypes with the most frequent allele were compared to the homozygote of the rare allele, and then the genotypes with the rare allele were compared to the homozygote of the most frequent allele. Associations for the dominant/recessive model were analysed using a 2x2 contingency table and a Fishers exact test where the odds ratios and relative risks with 95% confidence intervals were calculated as well as the p value. This is not always recommended, as it requires an assumption of HWE in cases and controls combined [18]. For the analysis of the alleles, 2x2 contingency tables with Fishers exact test were done to test for an association with tuberculosis.

When analyzing the *VDR* gene in relation to sputum conversion time, a number of additional statistical analyses were done, including Kaplan-Meier survival curves using PRISM and Cox

regression analysis using the program R: A language and environment for statistical computing, Base R [266] as well as packages survival [345], genetics [381], and haplo.stats [312]. This was discussed in Chapter 6.

Statistical genetics has the difficult issue of multiple testing. It is important to consider in association studies but there is no 'best fit' approach to handling this thorny issue. Within our analyses, we made corrections for multiple testing as much as possible. There is always the risk of over-correction, in particular when multiple polymorphisms are investigated in one gene [238]. The Bonferroni correction is widely used. This is a conservative correction where the number of independent tests done (usually individual SNPs analysed) is used to 'correct' the p value. However, polymorphisms in a single gene are seldom independent, and associations might suffer a multiple testing penalty [18].

To correct for population stratification is another issue. As discussed in Chapter 2 the population used in this study is admixed and the concern of population stratification is not unwarranted but, considering the history of the population, is unlikely. With the analysis of 25 unlinked SNPs the level of stratification (μ) could be represented by the mean χ^2 statistic among the SNPs investigated and could be used to correct for stratification [272].

A1.6.3 Linkage Disequilibrium Analysis and Haplotypes

Linkage disequilibrium (LD) refers to the situation where two apparently independent loci occur more or less frequently than would be expected. Loci that are in complete LD are considered to be linked and are co-inherited. The programs HaploView version 3.32 [23], PHASE version 2.1 [329] and GOLD version 1.0 [3] were used to analyse LD and infer haplotypes for the various genes investigated (weblinks for the various programs are listed in Table A1.13). The most frequently used LD coefficients are *D'* and r^2. *D'* is biased upward inversely with sample size [386] and is known to fluctuate when assessing small numbers or rare alleles. Loci are considered to be in 'complete' LD when *D'*=1 and are completely reshuffled when *D'*=0. *D'* is useful in assessing the probability of historical recombination in a given population whereas the square of the correlation coefficient between the loci, r^2, is more useful in the context of association studies [225, 259, 386].

Table A1.13: Weblinks for downloading of haplotype analysis programs.

PHASE	http://www.stat.washington.edu/stephens/
Haploview	http://www.broad.mit.edu/mpg/haploview
Unphased	http://www.hgmp.mrc.ac.uk/~fdudbrid/software/unphased/
GOLD	http://www.well.ox.ac.uk/asthma/GOLD

Through using PHASE and HaploView the determination of haplotypes was possible. As cases and controls were used, with no familial data, haplotypes could only be inferred, i.e. the data was unphased. HaploView uses an expectation-maximization algorithm for the assignment of alleles to a

haplotype with a high probability. The only limitation with this algorithm is that it assumes HWE in cases and controls. The reconstruction of haplotypes was therefore also done using PHASE which implements a Bayesian statistical method [328, 329]. HaploView [23] was used to determine haplotype tagging SNPs. Tagging SNPs partially or totally report the state of other polymorphisms in the same haplotype which are considered tagged SNPs [121]. An association analysis between the haplotypes in the cases versus controls was carried out. In addition 'Cocaphase' within the program Unphased version 2.404 [81] was used to test for association of multilocus haplotypes for unphased genotype data.

Power Calculations were done using EpiInfo 2000 version 1.1 (http://www.cdc.gov/epiinfo) subprogram Statcalc. Sample size and power calculations for unmatched cases and controls, at a confidence interval of 95% and a power of 80% were performed and these, depending on the frequency of the polymorphism, indicated the odds ratio of the association found.

A1.7 Submission of SNPs into NCBI database

All SNPs, novel and known, were submitted to the NCBI SNP database (dbSNP) [305]. The handle that is used by our group is EILEEN. To check that the SNP was not already in the database, or to find the NCBI dbSNP rs number, the online program BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) was used. It requires a short sequence surrounding the polymorphism to be entered and will identify, if present, the dbSNP rs number.

A1.8 Materials

Agarose Gels

1% w/v to 2% w/v were made up with 1x TBE \ SB running buffer

0.5M ethylenediaminetetraacetic acid/ ethylene-diamino-tetra-acetate (EDTA)

93.06g EDTA.2H_2O
Approximately 10g NaOH pellets
Add 400ml distilled water, adjust the pH to 8.0 using the 10M NaOH pellets.
Make up to 500ml
Autoclave and store at room temperature

Ethidium Bromide stock solution

Ethidium bromide was dissolved in TE buffer to a final concentration of 10mg/ml

Orange G Gel loading dye

0.1% Orange G
20% Ficoll
10mM EDTA pH 7
Filter through a 0.22µm filter

20 x Sodium Boric (SB) buffer

19.1 g of disodium tetraborate decahydrate in 500 ml of distilled water.

1x TBE

200ml of 5x TBE made up to 1000ml with distilled water

TE buffer

1.21g Tris-HCl
0.372g EDTA
Add approximately 800ml distilled water and mix. Once dissolved adjust the pH to 8.0 with concentrated HCl and make up to a final volume of 1000ml.
Autoclave

5x Tris-Borate-EDTA buffer (TBE)

54g Tris-HCl (1M pH 8.0 or 7.6)
27.5g Boric acid
20ml 0.5M EDTA (pH8.0)
Make up to 1000ml with distilled water

1 M Tris-HCl

121.1g in 800ml dH_2O pH to 8.0 or 7.6 make up to 1000ml using 10M Hydrogen Chloride (HCl)

Appendix 2

Supplementary Documentation for the Surrogate Marker database version 4 provided by K. Lawrence, Desmond Tutu TB centre, Tygerberg Campus, Stellenbosch University.

Initial file descriptions for the SM database:

SUMMARY: DEFINITIONS OF OUTCOMES FOR SURROGATE MARKER

1. ENROLLED

313 Patients were enrolled for the Surrogate Marker study on an initial SAIMR smear result from the clinic. Of these 14 were not included as they did not meet the Inclusion Criteria, which are listed below.

2. INCLUSION CRITERIA

1. Two sputum samples: Both smear positive, OR
2. Two sputum samples, one smear positive, one culture positive OR
3. One sputum sample, one smear positive, chest X-ray typical of pulmonary TB.

3. ACCEPTABLE SAMPLES

3.1 Inclusion

- Samples sent by the clinic to State lab (SAIMR / NHLS or Tygerburg Hospital AND / OR
- Samples sent to the Research lab at Diagnosis, Day 1 or Day 3.

3.2 Outcome

- Only Research Laboratory samples were accepted for outcomes.

3.3 Analysis

- Only Research Laboratory samples were accepted for analysis.

The above samples were accepted for both Inclusion (first Episode) and for Retreatments (second Episode).

Final Inclusion: N=299 Patients.

OUTCOME DEFINITIONS

All outcomes are based on the International definitions as stated in the source documents below.

New Patient

Def: On treatment is from start of treatment until all prescribed tablets have been taken or on treatment not longer than 15 months from the start.

1. **Cure**: a patient who is sputum smear negative in the last month of treatment **and** on at least one previous occasion.
2. **Treatment Completed**: A patient who has completed treatment, but who does not meet the criteria to be classified as either a cure or a failure, due to a smear not being done, or the smear being contaminated.
3. **Failure**: Any new patient that has remains or becomes smear positive at 5 months or later during treatment.
4. **Defaulter**: Default treatment for at least 2 consecutive months OR prescribed treatment not completed within 15 months.
5. **Transferred Out**: Transferred while on treatment to another reporting or recording unit, where treatment is continued, and for whom outcome is not known.
6. **Death while on TB treatment**: Any death, not necessarily death from TB.
7. **Not Treated**: < 10 dosages taken.

Retreatment
Same rules as above, but treatment is 8 months instead of 6.
The last negative sputum should be taken at 7 months or later.

SOURCES

WHO, IUATLD, KNCV. Revised international definitions in tuberculosis control. Int J Tuberc Lung Dis. 2001: 5(3): 213-215.
Tuberculosis Guide for Low Income Countries. Fifth Edition 2000 IUATLD

Updates to the above file:

SURROGATE MARKER EXCLUSION CRITERIA AND DEFINITIONS

Only patients who qualified for inclusion into the study will be included in this database.

INCLUSION CRITERIA

•Two sputum samples: Both smear positive, OR
•Two sputum samples: One smear positive, one culture positive OR
•One sputum sample: Both smear positive and culture positive•One sputum sample: One smear positive, chest X-ray typical of pulmonary TB
Results were accepted from either the Stellenbosch laboratory or the SAIMR/NHLS laboratory.

INITIAL EXCLUSIONS

Any patient who does not have an initial exclusion, must have an Outcome.

Initial exclusion criteria according to the study protocol (pg3) will include all patients who:

- are HIV positive
- have had previous treatment for TB
- have drug resistant TB at any time during treatment(MDR)
- have any disease or medication known to affect the immune system e.g. diabetes mellitus, malignancies, oral or inhaled steroids.
- MOTT
- NTB
- Other
- Initial protocol violators

The initial exclusions include any patients that presented with any one of the above during their 6 months of treatment. In other words, they qualified for the study according to the inclusion criteria, but then become an initial exclusion due to one above the above reasons, **at any time during their first treatment period.**

Any person who "qualifies" for an initial exclusion, must have InitExcl in the Outcome column. All other patients **must** have an outcome. If an outcome is not known, "U" (Unknown) must be entered. (After this exercise, there should be **NO** unknowns for any first episodes of the SM study).

SURROGATE MARKER STUDY

Patient Table

Microsoft Access - [Patient : Table]

File Edit View Insert Tools Window Help

Field Name	Data Type	Description
PatientID	Number	
BirthDate	Date/Time	
Sex	Text	
ExcludedDate_Epi1	Date/Time	Excluded date will be the same as the enrollment date of the first episode
ExcludedReason_Epi1	Text	A patient may have been enrolled but not included due to HIV, NotTB, pregnancy etc
ExcludedDate_Epi2	Date/Time	Excluded date will be the same as the enrollment date of the second episode
ExcludedReason_Epi2	Text	A patient may have been enrolled as a retreatment but not included due to HIV, NotTB, pregnancy etc
DeathDate	Date/Time	
DeathReason	Text	
RetreatmentPatient	Number	All patients to be aware of as a 2nd epi (Yes = 1, No = 2) whether or not included as a retreatment case
Clinic	Text	(U)itsig; (R)avensmead; (A)driaanse; (E)sies; (L)eondale
M30_Outcome	Text	The final outcome for each patient at completion of the study or if on retreatment

Unique Key:

PatientID: Use this to link to the Episode Table

Episode Table

Microsoft Access - [Episode : Table]

File Edit View Insert Tools Window Help

Field Name	Data Type	Description
EpisodeID	AutoNumber	Unique Key for the Episode table - this ID links to most other tables
DiagnosisDate	Date/Time	Date of Diagnosis
PatientID	Number	This links the Episode to the Patient
EpisodeNo	Text	Episode Number - check if 1 for 1st episode; 2 for 2nd episode
PatientCategory	Text	New or Retreatment as a result of Failure, Cure, TC etc
DiseaseCode	Text	All should be PTB, unless an initial exclusion (then NTB, PleuralEff)
InfectionID	Number	DO NOT USE THIS FIELD
Outcome_Int	Text	Outcome based on International (WHO: IUALTD) definitions - only smear results
Outcome_Res	Text	Outcome based on research laboratory results (including Culture results)
DebbieNo	Text	Special No for Research Lab
ENo	Text	Special No for Research Lab

Unique Key:

EpisodeID: Use this to link to most of the other tables. This key is unique for each episode.

Clinical Table

Microsoft Access - [Clinical : Table]

File Edit View Insert Tools Window Help

Field Name	Data Type	Descrip
ClinicalID	AutoNumber	Unique Key for each clinical record
EpisodeID	Number	Links to the specific episode of the patient
WhenID	Number	Links to the specific visit of the patient
Date	Date/Time	Date of clinic visit
BCG	Text	BCG Scar
BodyMass	Number	Weight
Length	Number	Height of patient
HIV	Text	HIV Result: Be careful of this field: multiple records per patient per episode
BMI	Number	To be calculated if both Mass and Length are present
Mantoux	Number	Mantoux result
Temperature	Number	
OtherMedProblems	Text	Not dual entered

Unique Key:

ClinicalID: This key is unique for each clinical record.

Other Keys:

EpisodeID: Links the clinic record to the specific episode
WhenID: Links the clinic record to the visit time-point.

Blood Table

Microsoft Access - [Blood : Table]

File Edit View Insert Tools Window Help

Field Name	Data Type	
BloodID	AutoNumber	Unique key for each blood sample
EpisodeID	Number	Linked to a specific patient episode
WhenID	Number	Linked to a specific visit per episode
BarcodeID	Number	DO NOT USE (db management)
Date	Date/Time	Date blood was taken
Hb	Number	g/l
WBC	Number	
NeutrPerc	Number	% of total WBC
LymphPerc	Number	% of total WBC
MonoPerc	Number	% of total WBC
EosPerc	Number	% of total WBC
Platelets	Number	

Unique Key:

BloodID: This key is unique for each FBC record.

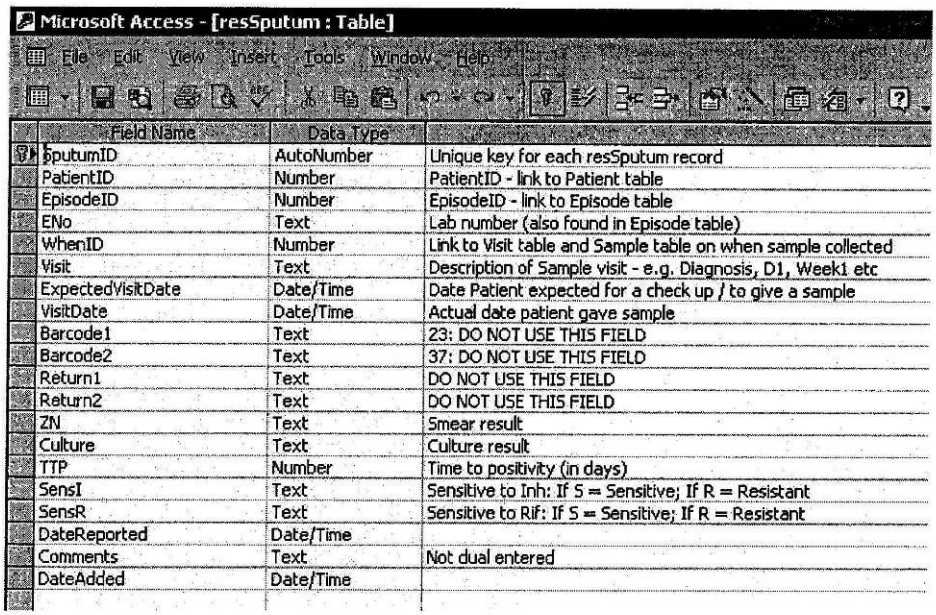
Other Keys:

EpisodeID: Links the clinic record to the specific episode

WhenID: Links the clinic record to the visit time-point.

---

resSputum Table


Microsoft Access - [resSputum : Table]

File Edit View Insert Tools Window Help

Field Name	Data Type	
SputumID	AutoNumber	Unique key for each resSputum record
PatientID	Number	PatientID - link to Patient table
EpisodeID	Number	EpisodeID - link to Episode table
ENo	Text	Lab number (also found in Episode table)
WhenID	Number	Link to Visit table and Sample table on when sample collected
Visit	Text	Description of Sample visit - e.g. Diagnosis, D1, Week1 etc
ExpectedVisitDate	Date/Time	Date Patient expected for a check up / to give a sample
VisitDate	Date/Time	Actual date patient gave sample
Barcode1	Text	23: DO NOT USE THIS FIELD
Barcode2	Text	37: DO NOT USE THIS FIELD
Return1	Text	DO NOT USE THIS FIELD
Return2	Text	DO NOT USE THIS FIELD
ZN	Text	Smear result
Culture	Text	Culture result
TTP	Number	Time to positivity (in days)
SensI	Text	Sensitive to Inh: If S = Sensitive; If R = Resistant
SensR	Text	Sensitive to Rif: If S = Sensitive; If R = Resistant
DateReported	Date/Time	
Comments	Text	Not dual entered
DateAdded	Date/Time	

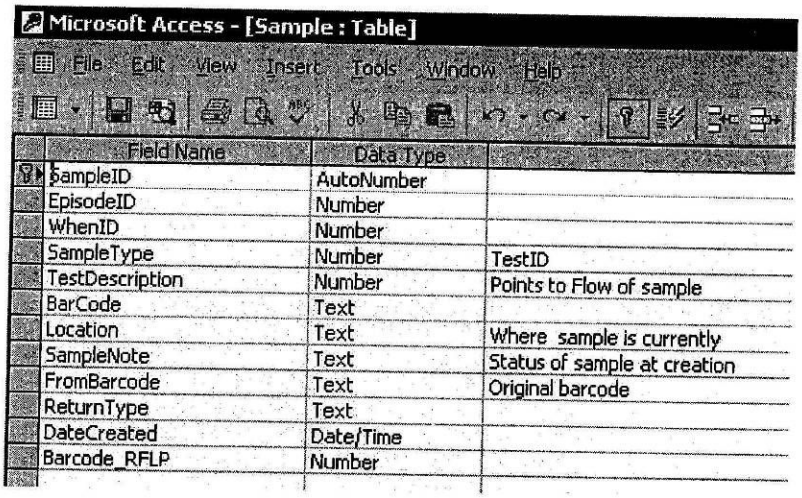
Unique Key:

SputumID

Other Keys:

PatientID: Links to the patient table. Be careful – it does not discriminate between episodes
EpisodeID: Links to the specific episode.
WhenID: Links to the specific time-point within an episode

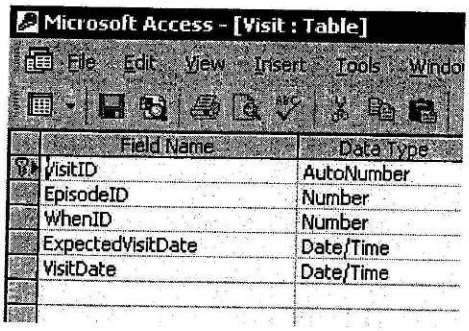
Sample Table



Field Name	Data Type	
SampleID	AutoNumber	
EpisodeID	Number	
WhenID	Number	
SampleType	Number	TestID
TestDescription	Number	Points to Flow of sample
BarCode	Text	
Location	Text	Where sample is currently
SampleNote	Text	Status of sample at creation
FromBarcode	Text	Original barcode
ReturnType	Text	
DateCreated	Date/Time	
Barcode_RFLP	Number	

This table is / was largely used for management and storage of samples.

Visit Table



Field Name	Data Type
VisitID	AutoNumber
EpisodeID	Number
WhenID	Number
ExpectedVisitDate	Date/Time
VisitDate	Date/Time

This table records each Expected Visit Date and Actual Visit Date for each person per episode.

Test Table

This table lists all the types of tests and samples at each time point. It is more of a complicated lookup table than of any practical use for anaylsis.

TestCategory Table

This table categorises the various tests. It is a lookup table, used mainly for weekly reports.

TestDescription Table

This table describes what needed to happen to each test/sample. It is a lookup table, used mainly for management.

SampleUse Table

This table lists which samples have been removed from storage, and will be deleted from the final analysis database. It is used for management of Samples.

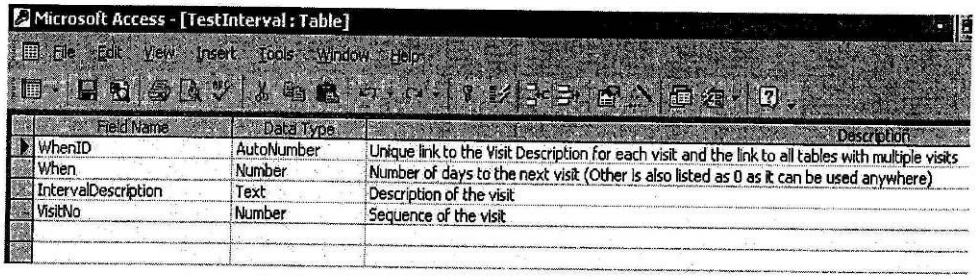
TestInterval Table

Although a lookup table, this table is very important as it lists the WhenID and Description of the visits.

TestInterval

WhenID	When	IntervalDescription	VisitNo
1	0	Diagnosis	1
2	7	W1	4
3	14	W2	5
4	21	W3	6
5	28	W4	7
6	35	W5	8
7	42	W6	9
8	49	W7	10
9	56	W8(M2)	11
11	91	M3(W13)	13
12	126	M4(W18)	14
13	154	M5	15
14	182	M6(W26)	16
15	273	M9	18
16	364	M12	19
17	546	M18	20
18	728	M24	21
19	910	M30	22
20	1	D1	2
21	3	D3	3
22	0	Other	0
23	59	W8+3	12
24	245	M8	17

Always use WhenID when looking up specific visits

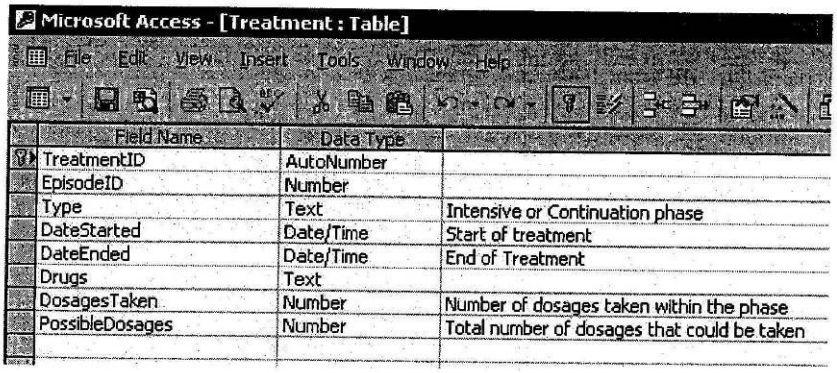


Field Name	Data Type	Description
WhenID	AutoNumber	Unique link to the Visit Description for each visit and the link to all tables with multiple visits
When	Number	Number of days to the next visit (Other is also listed as 0 as it can be used anywhere)
IntervalDescription	Text	Description of the visit
VisitNo	Number	Sequence of the visit

Unique Key

WhenID

Treatment Table



Field Name	Data Type	
TreatmentID	AutoNumber	
EpisodeID	Number	
Type	Text	Intensive or Continuation phase
DateStarted	Date/Time	Start of treatment
DateEnded	Date/Time	End of Treatment
Drugs	Text	
DosagesTaken	Number	Number of dosages taken within the phase
PossibleDosages	Number	Total number of dosages that could be taken

Unique Key:

TreatmentID

CXR Table

Microsoft Access - [CXR : Table]

File Edit View Insert Tools Window Help

Field Name	Data Type	De
CXRID	AutoNumber	
EpisodeID	Number	
WhenID	Number	
Date	Date/Time	
Normal	Text	
Cavities	Text	
Consolidation	Text	
Collapse	Text	
CavitiesLocation	Text	
Amount	Text	Required only if there are cavities (no no value in Cavities Location)
SizeCavity	Text	New;Required only if there are cavities (no no value in Cavities Location)
WallSize	Text	Required only if there are cavities (no no value in Cavities Location)
Ghon	Text	
Milliary	Text	
SizeEffusion	Text	
SideEffusion	Text	
Calcific	Text	
HilarNodes	Text	
Fibro	Text	
Bronchiect	Text	
AlveolarDisease	Text	
ExtentAlvDis	Text	Required if AlvDis is not none
CXRQuality	Text	
TypeTB	Text	
Notes	Text	
Improved	Text	

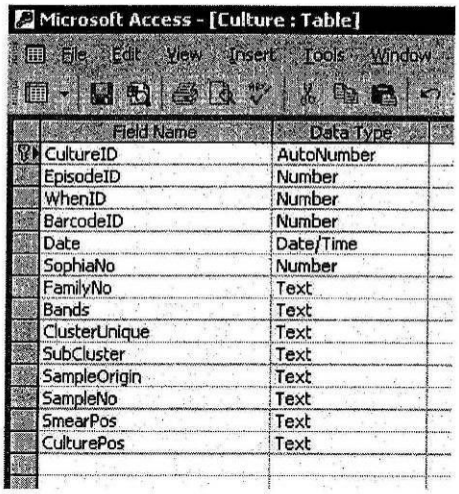
resUrine Table

Microsoft Access - [resUrine : Table]

File Edit View Insert Tools Windo

Field Name	Data Type
UrineID	AutoNumber
SampleID	Number
DateOfAnalysis	Date/Time
INH	Text
Hydrazine	Text
TypeOfTest	Number
DrugsTaken	Number
DrugsPrescribed	Number
LongWeekend	Yes/No

Culture Table



M30_Outcome_LU 29/07/2005

M30_OutcomeID	M30_Outcome_Desc
1	SEEN AT M30: WELL
2	NOT SEEN AT M30
3	NOT SEEN AT M30: LOST
4	NOT SEEN AT M30: MOVED AFTER M6
5	NOT SEEN AT M30: DIED AFTER M6
6	UNKNOWN
7	RETREATMENT: CURED AT M6
8	RETREATMENT: TREATMENT COMPLETED AT M
9	UNFAVOURABLE: FAILED AT M6
10	UNFAVOURABLE: DEFAULTER AT M6
11	UNFAVOURABLE: PROTOCOL VIOLATOR AT M6
12	UNFAVOURABLE: TRANSFERRED OUT AT M6
13	INITIAL EXCLUSION

Keys for the Surrogate Marker Database

KEY education	
Nil	0
<Std3	1
Std 4-7	2
Std8-10	3
Diploma	4

1 = yes
0 = no

Gender

m =male
f= female

Sensitivity

1=R
0=S

Key Income	
Nil	0
<R250 pm	1
R251-500 pm	2
R501-1000 pm	3
R 1001-1500 pm	4
R1501-2000	5
>2001 pm	6

Key extent Alv Disease	
1	<RUL
2	= RUL
3	>RUL
4	>1 LUNG

Outcome episode table		
1	C	cured
2	TC	treatment complete
3	F	failed
4	Defaulter	2 consecutive months no medication taken
no	ProtViol	left study
no	Trans	transferred out of study area

Censored Date	
1	Cured, all information available
0	censored

Last positive	
-1	never had a positive result
0	day 0

Genotypic Groups	Family (RFLP)
1	31,29,27,25,20
2	1,2,3,4,6,7,9,10,11,13,14,15,19,24,26,28,32,110,120,130,140,150
3	5,8,16,17,18,21,22,23
4	Unclassified
0	Pseudofamily 0

Appendix 3

Website Information

Websites used and additional sites of interest

Program	Link and description
AutoDimer	http://www.cstl.nist.gov/div831/strbase//AutoDimerHomepage/AutoDimerProgramHomepage.htm Rapidly screens previously selected PCR primers for primer-dimer and hairpin interactions in short DNA oligomers (< 30 nucleotides).
Biocarta	http://www.biocarta.com/genes/index.asp Illustrates how genes interact in dynamic graphical models.
Bioedit	http://www.mbio.ncsu.edu/BioEdit/bioedit.html Allows for sequence alignment.
BLAST	http://www.ncbi.nlm.nih.gov/BLAST/ Finds regions of local similarity between sequences.
ClustalW	http://align.genome.jp/ Allows for the aligment of multiple sequences.
DNAMan	http://www.lynnon.com/ For molecular biology applications for high efficiency sequence analysis.
EMBL	http://www.ebi.ac.uk/embl/ Europe's primary nucleotide sequence resource.
Ensembl	http://www.ensembl.org/index.html A joint project between EMBL - European Bioinformatics Institute and the Wellcome Trust Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes.
GAIA	http://bbu.cf.ac.uk/html/research/biostats.htm Tests for statistical interactions between loci.
GOLD	http://www.sph.umich.edu/csg/abecasis/GOLD/ A graphical summary of linkage disequilibrium in human genetic data.
Haploview	http://www.broad.mit.edu/mpg/haploview Simplifies and expedites the process of haplotype analysis.
HapMap	http://www.hapmap.org/ The International HapMap Project is a multi-country effort to identify and catalog genetic similarities and differences in human beings.
Harvester	http://harvester.fzk.de/harvester/ Harvester is a bioinformatic tool that crosslinks a large number of databases including : BLAST, ensEMBL, Entrez, gopubmed, HomoloGene, MapView, Mitocheck, OMIM, STRING, Unigene, UniprotKB.

Program	Link and description
Human Blat Search	http://genome.ucsc.edu/cgi-bin/hgBlat With the entry of about 30 base pairs of sequence results surrounding an identified SNP, databases are screened to identify if the SNP is already known in any online databases.
NCBI	http://www.ncbi.nlm.nih.gov/ National Centre for Biotechnology Information. Public database porthole.
Oligocalculator	http://trishul.sci.gu.edu.au/tools/OligoCalculator.html Checks the melting temperature and the GC content of primers.
primer3	http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi Identifies primers in a DNA sequence.
Pubmed	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed Journal and citation search engine.
Scitools Oligoanalyzer	http://scitools.idtdna.com/Analyzer/oligocalc.asp Tests for the presence of hairpins and dimers in primers.
Sequencher	http://www.genecodes.com/ Used for alignment of multiple sequences (very useful)
Software for Genetic analysis	http://dmoz.org/Science/Biology/Genetics/Software/ http://linkage.rockefeller.edu/soft/list.html Lists software available for genetic analysis
String	http://string.embl.de/ A database of known and predicted protein-protein interactions.
Virtual PCR	http://grup.cribi.unipd.it/cgi-bin/mateo/vpcr2.cgi Conducts a virtual PCR using information from the NCBI database

Appendix 4

Published articles

From Chapter 2

Lombard Z, Brune AE, Hoal EG, **Babb C**, van Helden PD, Epplen JT, Bornman L. HLA class II disease associations in southern Africa. Tissue Antigens 2006;67:97-110.

From Chapter 4 and partly Chapter 2

Barreiro LB, Neyrolles O, **Babb CL**, Tailleux L, Quach H, McElreavey K, Helden PD, Hoal EG, Gicquel B, Quintana-Murci L. Promoter Variation in the DC-SIGN Encoding Gene CD209 Is Associated with Tuberculosis. PLoS Medicine 2006; 3: e20.

From Chapter 4

Barreiro LB, Neyrolles O, **Babb CL**, van Helden PD, Gicquel B, Hoal EG, Quintana-Murci L. Length Variation of DC-SIGN and L-SIGN Neck-Region has no Impact on Tuberculosis Susceptibility. Human Immunology 2007; 68: 106-12.

From Chapter 5

Babb C, Keet EH, van Helden PD, Hoal EG. SP110 polymorphisms are not associated with pulmonary tuberculosis in a South African population. Human Genetics 2007; 121: 521-22.

From Chapter 6

Babb C, van der Merwe L., Beyers N, Pheiffer C, Walzl G, Duncan K, van Helden P, Hoal EG. Vitamin D receptor gene polymorphisms and sputum conversion time in pulmonary tuberculosis patients. Tuberculosis 2007; 87: 295-302.

van Helden PD, Moller M, **Babb C**, Warren R, Walzl G, Uys P, Hoal E. TB epidemiology and human genetics. Novartis Found Symposium 2006; 279: 17-31.

Tissue Antigens ISSN 0001-2815

REVIEW ARTICLE

HLA class II disease associations in southern Africa

Z. Lombard[1], A. E. Brune[1], E. G. Hoal[2], C. Babb[2], P. D. Van Helden[2], J. T. Epplen[3] & L. Bornman[1]

1 Department of Biochemistry, University of Johannesburg, PO Box 524, Auckland Park 2006, South Africa
2 MRC Center for Molecular and Cellular Biology, Department of Medical Biochemistry, Stellenbosch University, Faculty of Health Sciences, PO Box 19063, Tygerberg 7505, South Africa
3 Department of Human Genetics, Ruhr-University, D-44780 Bochum, Germany

Key words
disease association; HLA class II; resistance; southern Africa; susceptibility

Correspondence
Eileen G. Hoal
Department of Medical Biochemistry
Stellenbosch University
Faculty of Health Sciences
PO Box 19063
Tygerberg 7505
South Africa
Tel: 27 21 938 9412
Fax: 27 21 938 9476
e-mail: egvh@sun.ac.za

Received 2 September 2005; revised 4 October 2005; accepted 10 October 2005

doi: 10.1111/j.1399-0039.2005.00530.x

Abstract

Southern Africa harbors several population groups representing a diversity of gene pool origins. This provides a unique opportunity to study genetic disease predisposition in these populations against a common environmental background. Human leukocyte antigen (HLA) association studies of these populations could improve knowledge on inter-population variation and HLA-related disease susceptibility. The aim of this paper is to review HLA class II disease associations reported for southern African population groups, compare them with findings in other populations and identify those unique to southern Africa. A number of HLA class II disease associations appear to be unique to southern African populations. These include DRB1*14011 association with insulin-dependent diabetes mellitus susceptibility in the Xhosa and DRB1*10 and DQB1*0302 with rheumatoid arthritis susceptibility in the South African (SA) Indian and SA Coloreds, respectively. A noteworthy similarity in class II disease association was observed among southern African Caucasoid and their European parental populations. Unique HLA class II disease associations observed in southern Africa are consistent with the notion that unique environmental and natural selective factors have resulted in certain ethnic-specific HLA class II disease associations, while common HLA class II disease associations found across different populations support the notion that common diseases are caused by common, ancient alleles present in indigenous African populations.

Introduction

The African continent is home to more than 2000 distinct ethnic groups and languages, responsible for a considerable genetic diversity among native African populations (1). Southern Africa has seen waves of immigration from the African continent to the north in pre-recorded times and Caucasoid and Asian during the past three and a half centuries. This population diversity within a uniform environment creates a unique opportunity to study population genetics and complex diseases of genetic etiology.

The southern African population includes native African groups such as the Zulu, Xhosa, Sotho, Tswana, Ndebele, Swazi, Pedi, Lemba, Shona, Khoi, and San; the white population descended from Dutch, British, German, and French immigrants; admixture population groups such as the South African (SA) Coloreds and groups of Asian descent (2, 3).

Archaeological evidence indicates that southern Africa was first peopled by the ancestors of the San people, who exhibit the most ancient of African mitochondrial DNA lineages, the L1 lineage (1). The Khoi migrated into southern Africa about 2000 years ago from the north and coexisted with the San. Bantu-speaking pastoralists moving south entered the eastern part of southern Africa from approximately AD 250 (3). Pressure from these migration waves and subsequent European colonization drove the remaining San into the semi-desert areas of southern Africa. It is likely that there was substantial admixture between southern African blacks and the Khoi and San, as evidenced by the presence of San-specific Gm markers

in the majority of black SA groups, particularly the Xhosa population (2, 4). South Africa's largest population group, the Zulu, belongs to the Nguni linguistic division of the Niger–Congo people, as do the Xhosa (5). The Gokomere people, a Bantu-speaking group of migrant farmers, inhabited the Great Zimbabwe site from about AD 500, displacing earlier Khoisan people. These were the ancestors of the Shona, who consititute approximately 80% of the current Zimbabwean population (2).

The Venda population in the north of South Africa is regarded as one of the last population groups to migrate to the country and have experienced little admixture with other SA groups. The SA Coloreds are an ethnic group of mixed ancestry. This distinct group emerged about 300 years ago and received genetic input from Malaysia and Indonesia, as well as from the European Caucasoid and indigenous Khoi, San, and other black Africans (2, 6). The Indian population in South Africa originated from different regions of the Indian subcontinent, including the Hindu group originating from South India, Muslims from the Gujerat and Maharashtra, and the Tamil community, and is considered to be a hybrid population with some genetic contribution from Caucasoid, Black, and Chinese (2, 7). The genetic diversity in southern Africa is illustrated by the varying frequencies of both the human leukocyte antigen (HLA) class I (8) (not discussed in this review) and HLA class II alleles in the different groups (Table 1). The traditional geographic locations of various groups are indicated in Figure 1.

HLA genes and disease susceptibility

The HLA complex plays a fundamental role in disease predisposition. The HLA class I antigens were discovered first, and many diseases were initially coupled to this genetic locus. The subsequent characterization of the HLA class II antigens and their related genes revealed that this association was most likely due to linkage disequilibrium with the HLA class II region (9). A vast body of evidence relates various HLA class II alleles with susceptibility to multiple diseases, including those of infectious etiology.

Several hypotheses explaining the role of HLA genes in disease susceptibility involve the ability of HLA molecules to govern the immune response to antigens. Most of the molecular mechanisms underlying HLA disease associations have not yet been fully established, as the contribution of HLA is usually studied after disease onset. At this stage of disease, the task of the HLA molecules in initiating the immune response is surpassed by the cascades and cellular processes associated with human immunity, obscuring the exact molecular role of the HLA molecules (9, 10).

The southern Africa context

Southern Africa lends itself to study genetic predisposition to disease in different populations and offers a unique opportunity to compare candidate genes between (i) European populations residing in the contrasting environments of Europe or South Africa, (ii) various native ethnic groups, and (iii) admixture populations and their parental population groups.

Such studies may provide support for the common disease/common variant hypothesis (11). According to this hypothesis, susceptibility alleles are assumed to be ancient and therefore most likely present in native African populations. African-based genetic susceptibility studies are also valuable in establishing cases where less-frequent population-specific alleles predispose to disease (1).

The majority of research centers reside in economically developed areas (1, 12), explaining the lack of sampling of certain southern African population groups (Table 1).

Here, we summarize the literature available on HLA class II-associated disease susceptibility in southern African populations, some unique to the area, and compare it with data from other populations. Knowledge of unique patterns of disease susceptibility defined by HLA class II molecules could improve the understanding of the molecular mechanisms contributing to disease susceptibility in this geographic region and may in future contribute to the development of novel therapeutic approaches.

HLA class II diversity and disease susceptibility of southern African population groups: genetic and molecular implications

Insulin-dependent diabetes mellitus

Insulin-dependent diabetes mellitus (IDDM) results from hypo-secretion of insulin caused by Th1 cell-mediated autoimmune destruction of the pancreatic islet β cells (13, 14).

IDDM was one of the first autoimmune diseases to be associated with the HLA complex (15). Worldwide, IDDM susceptibility is associated with the HLA DRB1*03 and DRB1*04 alleles, with DRB1*03/*04 heterozygous individuals exhibiting the greatest risk of developing IDDM (16, 17). An increased frequency of DRB1*03-DQA1*0501-DQB1*02 and DRB1*0401-DQA1*03-DQB1*0302 haplotypes was reported in Brazilian type I diabetes patients (18) and DRB1*0402-DQA1*03-DQB1*0302 and DRB1*0301-DQA1*0501-DQB1*02 in IDDM patients from Turkey (19).

Similar associations of IDDM susceptibility with DRB1*03 and DRB1*04 have been observed in southern African populations (Table 2) – DRB1*03 in the Zulu (5), Shona (20) and SA Indian population (21); DRB1*04 in the Zulu (22) and SA Indians of North Indian descent (21)

Table 1 HLA class II allele frequencies in southern African populations

Serological Specificity	Allele	Black (87, 88)	Black (Cape Town) (88)	Zulu (22, 27, 40) Zulu + Sotho (57)	Xhosa (23, 29, 89, 90)	Shona (20, 31, 32, 89)	Khoi (88)	San (88)	SA Colored (23, 26, 29, 58, 60, 90, 91)	Indians (92, 93)	Tamils (30)	Hindu (30)	Muslim (30)	White (26, 88, 90)
DR1		0.03	0.047	0.047, 0.027	0.0428, 0.111	0.075, 0.14	–	0	0.147, 0.11, 0.057, 0.12	0.055, 0.026	0.05	0.05	0.18	0.1056
	DRB1*0101	0.084	–	0.07	0.02	0	0.017	–	0.03	–	–	–	–	–
	DRB1*0102	–	–	–	0.01	0.11	0.052	–	0.02	–	–	–	–	–
	DRB1*0103	–	–	–	–	0	0.058	–	0.01	–	–	–	–	–
DR2		–	–	0.23	0.1153, 0.21	0.216, 0.38	–	–	0.326, 0.1902, 0.344	0.223, 0.394	0.35	0.43	0.49	0.1495
	DRB1*1501	0.165	–	–	0.0769	–	0.3	0.015	0.0833	–	–	–	–	–
	DRB1*1502	–	–	–	–	–	0.012	–	0.0282	–	–	–	–	–
	DRB1*1503	–	–	–	–	0.24	–	–	0.0244	–	–	–	–	–
	DRB1*1505	–	–	–	0.0576	–	–	–	–	–	–	–	–	–
	DRB1*02	–	–	0.13	–	0.26	–	–	–	–	–	–	–	–
DR3		0.29	0.104	0.34	0.2379, 0.413	0.14, 0.26	–	0.019	0.216, 0.1099	0.0672	0.12	0.11	0.2	0.0979
									0.2077	0.121				
	DRB1*03	–	–	0.36	–	–	–	–	–	–	–	–	–	–
	DRB1*0301	0.195	–	–	0.1026	0.16	–	–	0.083	–	–	–	–	–
	DRB1*0302	–	–	–	0.1538	0.11	0.012	0.007	0.028	–	–	–	–	–
DR4		0.028, 0.13	0.104	0.123, 0.096	0.069, 0.141	0.039, 0.08	–	0.435	0.21, 0.11, 0.25	0.0189, 0.205	0.27	0.08	0.06	0.1516, 0.28
	DRB1*0401	–	–	–	0.04	0	0.135	0.409	0.07	–	–	–	–	–
	DRB1*0403	–	–	–	–	–	–	–	0.01	–	–	–	–	–
	DRB1*0404	–	–	–	0.01	0.04	0.036	0.044	0.01	–	–	–	–	–
	DRB1*0405	–	–	–	0.01	–	0.006	–	0.02	–	–	–	–	–
	DRB1*0408	–	–	–	–	–	–	–	0.01	–	–	–	–	–
	DRB1*0401/05	0.061	–	–	–	–	–	–	–	–	–	–	–	–
	DRB1*04	–	–	0.06	–	0.04	–	0.063	–	–	–	–	–	–
DR5		–	–	0.33	0.172, 0.323	0.134, 0.25	–	–	0.16, 0.30, 0.29	0.084, 0.166	0.18	0.2	0.16	0.132
	DRB1**1101/1201	0.23	–	–	–	–	–	–	–	–		–	–	–
DRw5		–	–	0.16	0.181, 0.305	0.186, 0.34	–	–	0.11, 0.15, 0.21	0.0737	–	–	–	0.082
DR6		–	–	0.16	–	–	–	–	0.21	0.146	0.15	0.14	0.1	–
	DRB1*1302	0.2	–	–	–	–	0.115	0.116	–	–	–	–	–	–
	0.115	0.116	–	–	–	–	–	–	–	–	–	–	–	–
DR7		0.07	0.145	0.123, 0.154	0.0742, 0.137	0.085, 0.16	–	0.073	0.24, 0.14, 0.25	0.153, 0.297	0.28	0.3	0.24	0.132
	DRB1*07	–	–	0.13	–	0.15	–	–	–	–	–	–	–	–
	DRB1*0701	0.11	–	–	0.0577	–	0.041	0.063	0.1387	–	–	–	–	–
DRw8		–	–	0.087, 0.039	0.006, 0.021	0.019, 0.04	–	–	0.027, 0.054	0.012	–	–	–	0.032

Table 1 Continued

Serological Specificity	Allele	Black (87, 88)	Black (Cape Town) (88)	Zulu (22, 27, 40) Zulu + Sotho (57)	Xhosa (23, 29, 89, 90)	Shona (20, 31, 32, 89)	Khoi (88)	San (88)	SA Colored (23, 26, 29, 58, 60, 90, 91)	Indians (92, 93)	Tamils (30)	Hindu (30)	Muslim (30)	White (26, 88, 90)
DR8		0.01	0.019	0.01	–	–	–	0.056	0.05	0.079	0.1	0.09	0.04	–
	DRB1*08	–	–	0.10	–	0.12	–	–	–	–	–	–	–	–
	DRB1*0802	0.109	–	–	0.006	–	–	–	–	–	–	–	–	–
	DRB1*0803	–	–	–	0.006	–	–	–	–	–	–	–	–	–
	DRB1*0804	–	–	–	0.006	–	0.029	0.036	–	–	–	–	–	–
DRw9		–	–	0.02	0.0078, 0.012	–	–	–	0.0091, 0.023, 0.018	0.003	–	–	–	0.017
DR9		0.01	0.01	0.005, 0.008	0.01	0.015, 0.03	–	0.019	0.02	0.008	0.01	0	0	–
	DRB1*09	–	–	0.02	–	0	–	–	–	–	–	–	–	–
	DRB1*0901	0.047	–	–	0.01	–	–	0.015	0.0167	–	–	–	–	–
DRw10		–	–	0.02	0.028, 0.051	0.029, 0.06	–	–	0.021, 0.0082, 0.017	0.045	–	–	–	0.014
DR10		0.02	0.026	0.03	–	–	–	0	0.016	0.104	0.11	0.07	0.1	–
	DRB1*10	–	–	0.01	–	0.02	–	–	–	–	–	–	–	–
	DRB1*1001	–	–	–	0.03	0.02	0.017	–	0.033	–	–	–	–	–
	None	0.014	–	–	–	–	–	–	–	–	–	–	–	–
DR11		0.247	0.182	–	–	–	–	0.063	–	–	–	–	–	–
	DRB1*11	–	–	0.36	–	0.4	0.006	–	–	–	–	–	–	–
	DRB1*1101	–	–	–	0.10	–	–	0.015	0.130	–	–	–	–	–
	DRB1*1102	–	–	–	0.03	–	0.017	0.022	0.017	–	–	–	–	–
	DRB1*1104	–	–	–	0.04	–	0.006	–	0.028	–	–	–	–	–
	DRB1*12	–	–	0.08	–	0.06, 0.07	–	–	–	–	–	–	–	–
	DRB1*1201	–	–	–	0.03	–	–	–	0.022	–	–	–	–	–
	DRB1*1202	–	–	–	–	–	0.012	–	0.050	–	–	–	–	–
	DRB1*13	–	–	0.33	–	0.19	–	–	–	–	–	–	–	–
	DRB1*1301	–	–	–	0.06	–	0.070	0.123	0.022	–	–	–	–	–
	DRB1*1302	–	–	–	0.07	–	–	–	0.017	–	–	–	–	–
	DRB1*1303	–	–	–	0.01	–	–	–	0.006	–	–	–	–	–
	DRB1*1305	–	–	–	0.01	–	–	–	–	–	–	–	–	–
	DRB1*1308	–	–	–	–	–	–	–	0.006	–	–	–	–	–
	DRB1*14	–	–	0.01	–	–	–	–	–	–	–	–	–	–
	DRB1*1401	–	–	–	0.03	–	0.012	0.007	0.022	–	–	–	–	–
	DRB1*1404	–	–	–	0.01	–	–	0	–	–	–	–	–	–
	DRB1*1601	–	–	–	–	–	0.006	–	–	–	–	–	–	–
	DRB1*1602	–	–	–	–	–	–	0.019	0.006	–	–	–	–	–
DR12		0.05	0.083	–	–	–	–	0.019	–	–	–	–	–	–
DR13		0.14	0.078	–	–	–	–	0.245	–	–	–	–	–	–
DR14		0.03	0.021	–	–	–	–	0	–	–	–	–	–	–
DR15		0.05	0.099	–	–	–	–	0.019	–	–	–	–	–	–
DR16		0	0.021	–	–	–	–	0	–	–	–	–	–	–
DR51	DRB5*0101	0.165	–	–	–	–	–	–	–	–	–	–	–	–
DR52	DRB3*0101	0.552	–	–	–	–	–	–	–	–	–	–	–	–
DR53	DRB4*0101	0.211	–	–	–	–	–	–	–	–	–	–	–	–

DQw1		–	–	0.66	0.44	–	–	–	0.420	–	0.69	0.77	0.88	0.39
DQ1		0.43	0.408	–	–	–	–	0.313	0.660	0.72	–	–	–	–
DQw2		–	–	0.23	0.11	–	–	–	0.150	–	0.22	0.27	0.12	–
DQ2		0.17	0.193	–	–	–	–	0.102	0.270	0.23	–	–	–	–
DQw3		–	–	0.31	0.25	–	–	–	0.44, 0.25	–	0.42	0.32	0.28	0.27
	DQB1*02	–	–	0.311	0.006	–	–	–	–	–	–	–	–	–
	DQB1*0201	–	–	–	0.167	0.28	0.111	0.084	0.193	–	–	–	–	–
	DQB1*0301	–	–	–	0.103	0.21	–	–	0.217	–	–	–	–	–
	DQB1*0302	–	–	–	0.039	0.04	0.144	0.497	0.089	–	–	–	–	–
	DQB1*0303	–	–	–	0.013	–	–	–	0.028	–	–	–	–	–
	DQB1*04	–	–	0.189	–	–	–	–	–	–	–	–	–	–
	DQB1*0402	–	–	–	0.160	0.11	0.055	0.058	0.050	–	–	–	–	–
	DQB1*05	–	–	0.23	–	–	–	–	–	–	–	–	–	–
	DQB1*0501	–	–	–	0.121	0.39	0.105	–	0.136	–	–	–	–	–
	DQB1*0502	–	–	–	–	–	0.017	–	0.033	–	–	–	–	–
	DQB1*06	–	–	0.554	–	–	–	–	0.006	–	–	–	–	–
	DQB1*0601	–	–	–	–	–	0.006	–	0.033	–	–	–	–	–
	DQB1*0602	–	–	–	0.221	0.49	0.273	0.130	0.132	–	–	–	–	–
	DQB1*0603	–	–	–	0.103	0.11	0.041	0.029	0.050	–	–	–	–	–
	DQB1*0604	–	–	–	0.032	0.04	0.017	0.108	–	–	–	–	–	–
	DQB1*0605	–	–	–	0.026	0.02	0.116	0.029	0.006	–	–	–	–	–
	DQB1*0609	–	–	–		–	–	–	0.006	–	–	–	–	–
	DQ*0602	–	–	0.27	–	–	–	–	–	–	–	–	–	–
	DQ*0603-0608	–	–	0.162	–	–	–	–	–	–	–	–	–	–
	DQB1*07	–	–	0.338	–	–	–	–	–	–	–	–	–	–
	DQB1*08	–	–	0.04	–	–	–	–	–	–	–	–	–	–
DQ3		0.02	0.069	–	–	–	–	0.283	0.437	0.38	–	–	–	–
DQ4		0.23	0.047	–	–	–	–	0.058	–	0.076	–	–	–	–
DQ7		0.16	0.255	–	–	–	–	0.15	–	–	–	–	–	–
DRw52		–	–	–	0.633	–	–	–	0.475	–	–	–	–	0.4462
DRw53		–	–	–	0.150	–	–	–	0.260	–	–	–	–	0.34
	DQA1*01	–	–	–	0.064	–	–	–	0.048	–	–	–	–	–
	DQA1*0101	–	–	–	0.071	0.30	0.122	0.015	0.083	–	–	–	–	–
	DQA1*0102	–	–	–	0.251	0.56	0.426	0.166	0.133	–	–	–	–	–
	DQA1*0103	–	–	–	0.100	0.15	–	–	0.050	–	–	–	–	–
	DQA1*0201	–	–	–	0.032	0.10	0.081	0.02	0.116	–	–	–	–	–
	DQA1*03	–	–	–	–	–	0.017	–	–	–	–	–	–	–
	DQA1*0301	–	–	–	0.090	0.07	0.214	0.605	0.121	–	–	–	–	–
	DQA1*0401	–	–	–	0.1538	0.11	0.023	0.029	0.033	–	–	–	–	–
	DQA1*0501	–	–	–	0.1529	0.38	0.126	0.036	0.229	–	–	–	–	–
	DQA1*0601	–	–	–	–	0.01	–	–	0.033	–	–	–	–	–
	DPB1*0101	–	–	–	–	–	0.175	0.235	0.128	–	–	–	–	–
	DPB1*0102	–	–	–	0.1026	–	–	–	–	–	–	–	–	–
	DPB1*0201	–	–	–	0.0897	–	0.164	0.029	0.100	–	–	–	–	–
	DPB1*0202	–	–	–	0.0128	–	–	–	–	–	–	–	–	–
	DPB1*0210	–	–	–	–	–	–	–	0.006	–	–	–	–	–
	DPB1*0301	–	–	–	–	–	0.057	0.04	0.033	–	–	–	–	–
	DPB1*0401	–	–	–	0.06	–	0.114	0.29	0.172	–	–	–	–	–
	DPB1*0402	–	–	–	0.04	–	0.260	0.12	0.115	–	–	–	–	–
	DPB1*0403	–	–	–	–	–	–	–	0.006	–	–	–	–	–
	DPB1*0501	–	–	–	–	–	0.006	–	0.039	–	–	–	–	–
	DPB1*0601	–	–	–	0.01	–	–	–	–	–	–	–	–	–
	DPB1*0801	–	–	–	–	–	0.006	0.022	–	–	–	–	–	–

Table 1 Continued

Serological Specificity	Allele	Black (87, 88)	Black (Cape Town) (88)	Zulu (22, 27, 40) Zulu + Sotho (57)	Xhosa (23, 29, 89, 90)	Shona (20, 31, 32, 89)	Khoi (88)	San (88)	SA Colored (23, 26, 29, 58, 60, 90, 91)	Indians (92, 93)	Tamils (30)	Hindu (30)	Muslim (30)	White (26, 88, 90)
	DPB1*0901	–	–	–	–	–	–	–	0.022	–	–	–	–	–
	DPB1*1101	–	–	–	0.02	–	0.058	0.007	0.006	–	–	–	–	–
	DPB1*1201	–	–	–	–	–	–	–	0.006	–	–	–	–	–
	DPB1*1301	–	–	–	0.04	–	0.006	–	0.072	–	–	–	–	–
	DPB1*1401	–	–	–	–	–	0.023	0.007	0.028	–	–	–	–	–
	DPB1*1501	–	–	–	0.01	–	0.017	0.015	0.006	–	–	–	–	–
	DPB1*1601	–	–	–	–	–	–	–	0.011	–	–	–	–	–
	DPB1*1701	–	–	–	0.01	–	–	–	0.006	–	–	–	–	–
	DPB1*1801	–	–	–	0.06	–	0.006	–	0.017	–	–	–	–	–
	DPB1*1901	–	–	–	–	–	0.006	–	0.011	–	–	–	–	–
	DPB1*2001	–	–	–	0.02	–	–	–	0.006	–	–	–	–	–
	DPB1*2101	–	–	–	0.01	–	–	–	0.006	–	–	–	–	–
	DPB1*2201	–	–	–	–	–	0.006	–	–	–	–	–	–	–
	DPB1*2301	–	–	–	0.006	–	0.006	0.015	0.006	–	–	–	–	–
	DPB1*2501	–	–	–	0.006	–	–	–	–	–	–	–	–	–
	DPB1*2601	–	–	–	0.074	–	–	–	0.006	–	–	–	–	–
	DPB1*2701	–	–	–	0.019	–	–	0.007	0.017	–	–	–	–	–
	DPB1*2801	–	–	–	0.013	–	–	–	0.006	–	–	–	–	–
	DPB1*3001	–	–	–	–	–	0.012	0.060	–	–	–	–	–	–
	DPB1*3101	–	–	–	0.013	–	–	–	0.006	–	–	–	–	–
	DPB1*3201	–	–	–	0.039	–	–	–	–	–	–	–	–	–
	DPB1*3301	–	–	–	0.032	–	–	–	–	–	–	–	–	–
	DPB1*3401	–	–	–	0.013	–	0.017	0.022	–	–	–	–	–	–
	DPB1*3901	–	–	–	0.006	–	–	–	0.006	–	–	–	–	–
	DPB1*4001	–	–	–	0.006	–	–	–	–	–	–	–	–	–
	DPB1*4101	–	–	–	0.026	–	–	–	–	–	–	–	–	–
	DPB1*4401	–	–	–	0.006	–	–	–	–	–	–	–	–	–
	DPB1*4601	–	–	–	0.006	–	–	–	–	–	–	–	–	–
	DPB1*4701	–	–	–	0.019	–	–	–	–	–	–	–	–	–
	DPB1*4901	–	–	–	0.019	–	–	–	–	–	–	–	–	–
	DPB1*5001	–	–	–	0.045	–	–	–	0.006	–	–	–	–	–
	DPB1*5101	–	–	–	0.013	–	–	–	0.006	–	–	–	–	–
	DPB1*5501	–	–	–	–	–	0.035	0.108	0.006	–	–	–	–	–

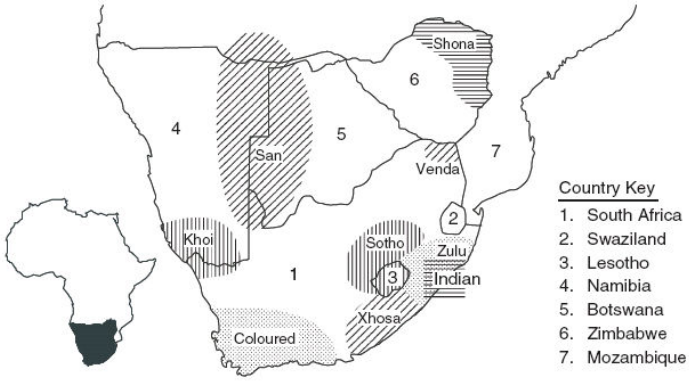


Figure 1 Populations studied in Southern Africa. Shaded areas indicate the approximate traditional distributions of the various groups. Subsequent urbanization has resulted in large-scale migration.

and DRB1*03/*04 heterozygosity in the Zulu and SA Indian population (21, 22). In contrast, the SA Coloreds showed no significant association between HLA DRB1*03 and IDDM (23). Xhosa patients with IDDM showed a unique DRB1*14011 susceptibility association (22), whereas a study on the Zulu population showed various susceptibility associations, of which HLA-DRB1*09 was most significant (5). A study conducted on the Shona population found a significant association between IDDM and HLA DRB1*0301, DQA*0301, DQB*0201, and DQB*0302 (20). HLA class II alleles associated with IDDM also differed significantly between southern African populations and the rest of the world. Whereas DRB1*15, DQB1*0602, DQB1*0301, and DQA1*0102 conferred protection from IDDM in most other populations (24), it was found that DRB1*0302 and DQA1*04

Table 2 HLA class II alleles associated with IDDM in southern African populations

Population group	HLA class II allele	Association (*P* value)	Cases	Controls	*P* value corrected	Validation[a]	Reference
Susceptibility							
Indian	DRB1*03 (DRB1*03/*04 heterozygote)	<0.035	35	235	Yes	No	(21)
Colored	DRB1*04	<0.02	57	Not stated	Yes	No	(23)
Zulu	DRB1*03/*04 (heterozygote)	Not stated	56	105	Yes	No	(22)
	DRB1*04	0.002	47	3	Yes	No	(5)
	DRB1*03	0.003					
	DRB1*09	<0.001					
	DQB*0302	<0.001					
	DQA1*03	<0.035					
	DQB1*02	<0.004					
Xhosa	**DRB1*14011**	<0.05	50	Not stated	Yes	No	(23)
Shona	DRB1*03	<0.005	40	82	No	Yes	(20)
	DRB1*04	<0.01					
	DQA1*0301	<0.001					
	DQB1*0201	<0.001					
	DQB1*0302	<0.01					
Resistance							
Zulu	**DRB1*0302**	Not stated	47	3	Yes	No	(5)
	DQA1*0102	Not stated					
	DQA1*04	Not stated					
Shona	**DRB1*11**	Not stated	40	82	No	Yes	(20)
	DQB1*0602	Not stated					
Colored	DR2	<0.01	57	Not stated	Yes	No	(23)

[a]Validation is interpreted to be the replication of the study findings by the same authors on a different selection from the same population group, in the same study, or replication of the findings by another research group on the same population group.
Allele associations shown in bold are unique to southern Africa.

were negatively associated with IDDM in Zulus (5). Similarly, DRB1*11 and DR2 were observed to be associated with IDDM resistance in the Shona (20) and the SA Coloreds, respectively (23)

Although it is thought that HLA molecules contribute to a discrepancy in binding affinity of autoantigenic peptides to distinct polymorphic alleles, the precise molecular mechanism for HLA-associated susceptibility or resistance in IDDM is not clear. A proposed mechanism is that favored binding of certain β-cell peptides to the above-mentioned susceptibility-associated HLA molecules occurs, leading to the activation of autoreactive T cells (10).

Rheumatoid arthritis

The connective tissue in the joints, in particular the synovial membrane, is the target of autoimmune attack in rheumatoid arthritis (RA), leading to inflammation and thickening of the membrane and joint deformation (13, 14).

Several studies suggest that the main contributors to RA susceptibility are HLA-DRB1 alleles coding for a five amino acid motif in the third hypervariable region of the DRβ1 molecule. These alleles code for a positively charged amino acid (Arg or Lys) at residue DRβ71, making these molecules prone to binding potential arthritogenic peptides containing a negatively charged amino acid at residue four (10, 25). HLA-DRB1*04 encodes this motif and is most often associated with RA (25–27). Similarly, HLA-DRB1*04 is associated with RA susceptibility in most southern African populations, including the Zulu (27) Sotho (28), Xhosa (29), SA Muslim Indians (30), the Shona (31, 32), SA Coloreds (29), and Afrikaner Whites (26) (Table 3). An association with DR10 was observed in Indian Hindi and Tamil communities in South Africa (30) and in the Shona (32). The Shona and SA Colored populations exhibited a further association of RA susceptibility with HLA-DQB1*0302, which is unique to these groups (26, 32). The DRB1*0301 and *0302 alleles both showed a unique, significant protective effect for RA in both the Xhosa and SA Colored (29).

Other diseases

Juvenile chronic arthritis

Juvenile chronic arthritis (JCA) is a combination of many clinically heterogeneous diseases, with the common clinical feature being the persistence of one or several inflamed joints caused by autoimmunity (33). Few disparate population groups have been studied, but it is evident that, similar to RA, DRB1*04 is associated with JCA susceptibility (34) in addition to DR8 and DR11 – each depending on the JCA subgroups represented (35). A distinct southern African association between JCA and DR10 has been found in SA Coloreds (36) (Table 4).

Chronic rheumatic heart disease

The structural homology between the Streptococcal M protein and human myocardial and valve proteins leads to a late-onset autoimmune reaction directed against the cardiac tissues of the host (13, 37). DRB4 super-specificities (DRB1*04 and DRB1*07) have been associated with chronic rheumatic heart disease (CRHD) susceptibility in Egyptian (38) and Brazilian patients (39), and an increased DRB1*0403 frequency was observed in Mexican patients (38). A unique association between CRHD and DRB1*01 and DR6 was observed in black SAs (40) (Table 4). At a molecular level, it has been proposed that HLA-DRB4 and its super-specificities preferentially present peptides from the M protein to T cells, instigating the autoimmune process (39).

Idiopathic dilated cardiomyopathy

In idiopathic dilated cardiomyopathy (IDC), T-cell autoreactivity and increased expression of HLA class I and II molecules on myocytes of cardiomyopathy sufferers may contribute to sustained myocardial damage (41). IDC was found to be associated with HLA-DRB1*10 in black SAs (42) (Table 4).

Graves' disease

Graves' disease is due to the lymphocytic infiltration of the thyroid and presence of autoantibodies directed against thyroid proteins, often manifesting as goiter and hyperthyroidism (13, 14). Susceptibility to Graves' disease is associated with DRB1*03 and DRB1*01 in the Zulu population (43). Similar associations are observed in French, German, and Indian patients (44).

Human T-cell leukemia virus-1-associated myelopathy

Human T-cell lymphotrophic virus-1 is a retrovirus that infects human $CD4^+$ T cells (45). A small percentage of infected individuals develop either adult T-cell leukemia (ATL) or a disabling progressive neurological disorder named human T-cell leukemia virus-1-associated myelopathy (HAM), where immunogenetic factors influence outcome (46). Globally, the DRB1*0101-DQB1*0501 haplotype has been implicated with HAM susceptibility, whereas this haplotype is rare in ATL sufferers (47, 48). A weak association between HAM and DR2 was observed in Zulus (49).

Systemic lupus erythematosus

Systemic lupus erythematosus (SLE) is a chronic inflammatory, multisystemic disorder of the connective tissue

Table 3 HLA class II alleles associated with RA in southern African populations

Population group	HLA class II allele	Association (*P* value)	Cases	Controls	*P* value corrected	Validation[a]	Reference
Susceptibility							
Zulu	DRB1*04	<0.001	100	513	Yes	No	(27)
Sotho	DRB1*04	<0.001	25	43	No	No	(28)
Xhosa	DRB1*04	<0.001	25	94	Yes	No	(29)
Indians			121	446	Yes	No	(30)
Muslim	DRB1*04	<0.01	30	49			
Hindi	DRB1*10	<0.001	38	135			
Tamil	DRB1*10	<0.01	53	262			
Colored	DRB1*04	<0.001	65	114	Yes	Yes	(29)
			183	3716	Yes	Yes	(26)
	DQB1*0302	<0.001	65	114	Yes	Yes	(29)
		0.03	183	3716	Yes	Yes	(26)
Whites	DRB1*04	<0.001	66	1059	Yes	No	(26)
Shona	DRB1*04	Not stated	72	82	No	Yes	(32)
Shona	DRB1*04	Not stated	26	119	No	Yes	(31)
			69	82	No	Yes	(32)
	DRB1*10	Not stated	69	82	No	No	(32)
	DQB1*0302	Not stated	26	119	No	No	(31)
Resistance							
Xhosa	**DRB1*03**	0.028	25	94	Yes	No	(29)
Colored	**DRB1*03**	<0.001	65	114	Yes	No	(29)

[a]Validation is interpreted to be the replication of the study findings by the same authors on a different selection from the same population group, in the same study, or replication of the findings by another research group on the same population group.
Allele associations shown in bold are unique to southern Africa.

and is thought to be the result of the malfunction of the regulatory mechanisms of the autoimmune system (13, 14). A multitude of genes have been shown to influence SLE susceptibility, among them CTLA4 (50, 51), the apoptosis genes FAS and FASL (52), FCGR2B (53), and various candidates within the HLA class II region. Three distinct haplotypes within the class II region have been identified that exhibit transmission distortion, in a large collection of

Table 4 HLA class II alleles associated with susceptibility to other diseases in southern African populations

Disease	Population Group	HLA class II allele	*P* value	Cases	Controls	*P* value corrected	Reference
Juvenile chronic arthritis	Colored	**DRB1*10**	<0.001	91	549	Yes	(36)
Chronic rheumatic heart disease	Black	**DRB1*01**	<0.045	103	220	Yes	(40)
		DR6	<0.045				
Idiopathic dilated cardiomyopathy	Black	**DRB1*10**	<0.02	57	220	Yes	(42)
Grave-s disease	Zulu	DRB1*01	<0.014	63	330	Yes	(43)
		DRB1*03	<0.023				
Human T-cell leukemia virus-1-associated myelopathy	Zulu	DR2	Borderline	40	556	Yes	(49)
Systemic lupus erythematosus	Zulu	DRB1*02	<0.005	49	87	No	(57)
		DQB1*0602	<0.025				
	Sotho	DRB1*02	<0.005	49	87	No	(57)
		DQB1*0602	<0.025				
	Colored	DR2	0.0005	75	549	No	(58)
		DRB1*07	0.02				
Multiple sclerosis	Colored	DR2	0.25–0.1	24	1042	No	(60)

Allele associations shown in bold are unique to southern Africa. None of these studies were validated. Validation is interpreted to be the replication of the study findings by the same authors on a different selection from the same population group, in the same study, or replication of the findings by another research group on the same population group.

families with SLE (54). DR2 is significantly associated with SLE in American Caucasians (55) and Malaysians (56) and has also been observed in the Sotho, Zulu, and SA Colored patients (57) (Table 4). Unique SLE associations with DRB1*07 were observed in the SA Colored population (58) and in the Sotho and Zulu populations with DQB1*0602 (57).

Multiple sclerosis

Multiple sclerosis (MS) is characterized by small, scattered areas of degeneration of the myelin sheath of nerve fibers (13). MS progression begins when viral antigens are presented to T cells by specific HLA molecules. Activated T cells cross the blood–brain barrier into the central nervous system, interact with cells bearing HLA class II molecules presenting the myelin antigens to which the T cells were originally primed, which leads to inflammation and myelin damage (59). DR2 has been implicated in MS susceptibility in SA Coloreds (60), similar to reports on Northern Europeans (59), North Americans, Californian Caucasians, and Italians (61).

Infectious diseases

In comparison with autoimmune diseases, research on HLA and susceptibility to infectious diseases within the southern African context is limited.

Mycobacterial disease (tuberculosis and leprosy)

Mycobacterium tuberculosis and *Mycobacterium leprae* are the causative agents of tuberculosis (TB) and leprosy, respectively. In general, leprosy is characterized by bacteria-induced skin and nerve lesions, whereas with TB, the most common clinical manifestation is pulmonary TB, resulting in similar lesions within the lungs (62). HLA-DR2 is most consistently associated with TB in a diversity of populations, including Indian (63, 64), Polish (65), Thai (66), Indonesian (67), and Russian (68). Similarly, DR2 has been associated with susceptibility to leprosy in Brazilian (69) and Indian populations (70). Nonetheless, inter-population variations in HLA/TB associations have been reported. HLA-DQB1*0503 was found to influence TB progression in the Cambodian population (71); DQB1*0601 was associated with TB susceptibility in the Thai and South Indian population (64, 66), and the HLA haplotype DRB1*08032-DQB1*0601 was associated with genetic susceptibility to multidrug-resistant TB in Korean patients (72). A study of the Venda population (92 TB patients and 117 controls) showed an association of DRB1*1302 with TB susceptibility (Lombard et al., unpublished data), whereas Boshoff et al. (unpublished data) have shown a marginal association of DRB1*03 with TB in the SA Colored population (106 TB patients and 107 controls).

Human immunodeficiency virus susceptibility

The hallmark of human immunodeficiency virus susceptibility (HIV) infection is a decline in $CD4^+$ T cells, resulting in compromised immunity (73). Several reports have suggested an association of HLA class II with HIV infection progression – HLA-DRB1*03 and DQB1*02 are associated with developing symptomatic HIV infection in Italians, whereas DR2 has been associated with disease progression (74). DQB1*0603 and DQB1*0201 have been found to be associated with HIV-1 susceptibility in American Caucasians and African Americans, respectively, whereas DQB1*03032 was associated with resistance against HIV infection in American Caucasians but not in African Americans (75, 76). HLA-DRB1*01 was reported to confer resistance to HIV infection in East African populations (77), and HLA-DRB1*1302 was found least frequently in Ugandan HIV-positive patients (78). In southern Africa, HLA-DRB1 homozygosity was significantly associated with high viral load in Zimbabwean patients (73). Recently, a significant association between certain HLA-DRB1 and DQB1 alleles and transmission of HIV type I was observed in Zambian couples (79). Seronegative partners with either DRB1*0301-DQB1*0201 or DRB1*1503-DQB1*0602 demonstrated accelerated seroconversion, whereas presence of DRB1*1301 in initially seropositive partners led to delayed transmission of HIV.

Other infectious diseases

Resistance to malaria is conferred by the HLA-DRB1*1352 and DRB1*1302 alleles in the Gambian population (80). Hepatitis B resistance in Gambians is also conferred by DRB1*1302 (81), whereas the presence of DRB1*07 is associated with hepatitis B susceptibility (82). The presence of HLA-DQB1*0603 in black SA patients with hepatitis B infection led to a more rapid development of membranous nephropathy (83).

Conclusion

The study of the genetic basis of a complex disease is critical for an improved understanding of its development, progression, and treatment. The genetic diversity and unique demographic history of southern African population groups suggests that they may be invaluable in the fine mapping of such diseases. Similar to other studies conducted worldwide, association between the HLA class II genetic region and disease has been found in southern African populations, with several population-specific allelic associations revealing the distinct genetic character of

the population groups residing in this area (Tables 2–4). Unique associations were specifically observed among native southern African population groups and groups which have received genetic input from them. The only exception to this observation is the association of HLA-DRB1*10 with RA in the Indian Hindi and Tamil communities of South Africa (30), illustrating the ethnic diversity within the Indian subcontinent. In most cases, one or more of the alleles associated with a specific disease in the rest of the world also occur in the southern African population. These findings support the hypothesis that common diseases are caused by common, ancient alleles present in indigenous African populations (1). All HLA disease associations observed in the southern African Caucasoid population corresponded to those observed in European Caucasian, as would be expected if the genetic influence was greater than the environmental.

While a limited number of studies have been conducted in southern Africa to analyze the role of HLA class II polymorphisms in susceptibility to infectious disease, evidence for unique genetic associations among Africans is apparent. Efforts to develop vaccines and other pharmacological therapies for HIV infection, TB, and malaria are delayed by a poor understanding of the genetic mechanisms involved in these infectious diseases (84). With one of the world's most rapidly increasing TB- and HIV-incidence rates (12, 85), southern Africa should be a focal point for research approaches to resolve these global epidemics.

The practical implications of understanding gene–disease associations include the identification of novel molecular targets for intervention, the design of vaccines and immunotherapeutics, and the prediction of therapeutic efficiency according to an individuals' genotype (86). The distinctive southern African milieu, in which diverse populations are harbored against common environmental backgrounds, provides a genetic resource that has the potential to be utilized for the health benefit of all these groups.

Acknowledgment

The Medical Research Council and the National Research Foundation of South Africa (NRF Grant number 2053199 to LB) supported this review.

References

1. Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and complex disease. *Nat Rev Genet* 2002: **3**: 611–21.
2. Nurse GT, Weiner JS, Jenkins T. *The People of Southern Africa and their Affinities. Oxford, UK: Clarendon Press, 1985.*
3. Mountain A. *The First People of the Cape. Cape Town, South Africa: David Philip Publishers, 2003.*
4. Hammond MG, Du Toit ED, Sanchez-Mazas A et al. HLA in sub-Saharan Africa. In: Charron D, ed. *HLA: Genetic Diversity of HLA; Functional and Medical Implication. Proceedings of the 12th International Histocompatibility Workshop & Conference. Paris, France: EDK*, 1997, 345–53.
5. Pirie FJ, Hammond MG, Motala AA, Omar MA. HLA class II antigens in South African Blacks with type I diabetes. *Tissue Antigens* 2001: **57**: 348–52.
6. Van der Ross R. *100 Questions about Coloured South Africans. Cape Town, South Africa: University of Western Cape Printing Department, 1993.*
7. Rani R, Fernandez-Vina MA, Stastny P. Associations between HLA class II alleles in a North Indian population. *Tissue Antigens* 1998: **52**: 37–43.
8. Cao K, Moormann AM, Lyke KE et al. Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue Antigens* 2004: **63**: 293–325.
9. Marsh SGE, Parham P, Barber LD. *The HLA Facts Book. London, UK: Academic Press, 2000.*
10. Thorsby E. Invited anniversary review: HLA associated diseases. *Hum Immunol* 1997: **53**: 1–11.
11. Lander ES. The new genomics: global views of biology. *Science* 1996: **274**: 536–9.
12. Sibanda EN. Research and clinical aspects of immunology in Africa. *Curr Opin Immunol* 2001: **13**: 528–32.
13. Youngson RM. *Collins Dictionary of Medicine. London, UK: Harper Collins, 1992.*
14. Marieb EN. *Human Anatomy and Physiology. California, USA: Benjamin Cummings, 1995.*
15. Singal DP, Blajchman MA. Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. *Diabetes* 1973: **22**: 429–32.
16. Rodey GE, White N, Frazer TE, Duquesnoy RJ, Santiago JV. HLA-DR specificities among black Americans with juvenile-onset diabetes. *N Engl J Med* 1979: **301**: 810–2.
17. Cudworth AG, Wolf E. The genetic susceptibility to type I (insulin-dependent) diabetes mellitus. *Clin Endocrinol Metab* 1982: **11**: 389–408.
18. Volpini WM, Testa GV, Marques SB et al. Family-based association of HLA class II alleles and haplotypes with type I diabetes in Brazilians reveals some characteristics of a highly diversified population. *Hum Immunol* 2001: **62**: 1226–33.
19. Saruhan-Direskeneli G, Uyar FA, Bas F et al. HLA-DR and -DQ associations with insulin-dependent diabetes mellitus in a population of Turkey. *Hum Immunol* 2000: **61**: 296–302.
20. Garcia-Pacheco JM, Herbut B, Cutbush S et al. Distribution of HLA-DQA1-DQB1 and DRB1 alleles in black IDDM patients and controls from Zimbabwe. *Tissue Antigens* 1992: **40**: 145–9.
21. Omar MA, Hammond MG, Rajput MC, Asmal AC. HLA A, B, C and DR antigens in young South African Indians with insulin-dependent diabetes mellitus. *S Afr Med J* 1984: **66**: 765–7.
22. Omar MA, Hammond MG, Asmal AC. HLA-A, B, C and DR antigens in young South African blacks with type 1 (insulin-dependent) diabetes mellitus. *Diabetologia* 1984: **26**: 20–3.

23. Orren A, Taljaard D, de Toit E. HLA-A, B, C and DR antigen associations in insulin dependent diabetes mellitus (IDDM) in South African Negro (black) and Cape coloured people. *Tissue Antigens* 1985: **26**: 332–9.
24. Caillat-Zucman S, Djilali-Saiah I, Timsit J et al. Insulin dependent diabetes mellitus (IDDM). In: Charron D, ed. *HLA: Genetic Diversity of HLA; Functional and Medical Implication, Proceedings of the 12th International Histocompatibility Workshop & Conference. Paris, France*: *EDK*, 1997, 389–98.
25. Toussirot E, Sauvageot C, Chabod J, Ferrand C, Tiberghien P, Wendling D. The association of HLA-DM genes with rheumatoid arthritis in Eastern France. *Hum Immunol* 2000: **61**: 303–8.
26. Martell RW, du Toit ED, Kalla AA, Meyers OL. Association of rheumatoid arthritis with HLA in three South African populations – whites, blacks and a population of mixed ancestry. *S Afr Med J* 1989: **76**: 189–90.
27. Mody GM, Hammond MG, Naidoo PD. HLA associations with rheumatoid arthritis in African blacks. *J Rheumatol* 1989: **16**: 1326–8.
28. Pile KD, Tikly M, Bell JI, Wordsworth BP. HLA-DR antigens and rheumatoid arthritis in black South Africans: a study of ethnic groups. *Tissue Antigens* 1992: **39**: 138–40.
29. Rousseau J, Pokorny L, Glaser J, Creemers PC. Predisposing and protective HLA-DR and DQ alleles for rheumatoid arthritis in South African mixed-ancestry and Xhosa populations. *S Afr J Sci* 2003: **99**: 89–91.
30. Mody GM, Hammond MG. Differences in HLA-DR association with rheumatoid arthritis among migrant Indian communities in South Africa. *Br J Rheumatol* 1994: **33**: 425–7.
31. Martell RW, Stein M, Davis P, West G, Emmanuel J, du Toit ED. The association between HLA and rheumatoid arthritis in Zimbabwean blacks. *Tissue Antigens* 1990: **36**: 125–6.
32. Cutbush S, Chikanza IC, Biro PA et al. Sequence-specific oligonucleotide typing in Shona patients with rheumatoid arthritis and healthy controls from Zimbabwe. *Tissue Antigens* 1993: **41**: 169–72.
33. Miller ML. Juvenile rheumatoid arthritis. *Curr Probl Pediatr* 1994: **24**: 190–8.
34. Forre O, Dobloug JH, Hoyeraal HM, Thorsby E. HLA antigens in juvenile arthritis. Genetic basis for the different subtypes. *Arthritis Rheum* 1983: **26**: 35–8.
35. Goedde R, Miterski B, Drynda S, Böschow G, Opperman J, Epplen JT. Numerous genetic predisposition factors for rheumatic arthritis in adults and juveniles. *Recent Res Dev Human Genet* 2004: **2**: 183–203.
36. Bhettay E, Martell R, Creemers PC. Association of HLA-DR10 with juvenile chronic arthritis in South Africans of mixed ancestry. *Tissue Antigens* 1994: **43**: 47–9.
37. Guilherme L, Cunha-Neto E, Coelho V et al. Human heart-infiltrating T-cell clones from rheumatic heart disease patients recognize both streptococcal and cardiac proteins. *Circulation* 1995: **92**: 415–20.
38. Goldberg AC, Kalil J, Gorodezky C et al. HLA and rheumatic fever. In: Charron D, ed. *HLA: Genetic Diversity of HLA; Functional and Medical Implication, Proceedings of the 12th International Histocompatibility Workshop & Conference. Paris, France*: *EDK*, 1997, 413–8.
39. Guilherme L, Oshiro SE, Fae KC et al. T-cell reactivity against streptococcal antigens in the periphery mirrors reactivity of heart-infiltrating T lymphocytes in rheumatic heart disease patients. *Infect Immun* 2001: **69**: 5345–51.
40. Maharaj B, Hammond MG, Appadoo B, Leary WP, Pudifin DJ. HLA-A, B, DR, and DQ antigens in black patients with severe chronic rheumatic heart disease. *Circulation* 1987: **76**: 259–61.
41. Malkiel S, Kuan AP, Diamond B. Autoimmunity in heart disease: mechanisms and genetic susceptibility. *Mol Med Today* 1996: **2**: 336–42.
42. Maharaj B, Hammond MG. HLA-A, B, DR, and DQ antigens in black patients with idiopathic dilated cardiomyopathy. *Am J Cardiol* 1990: **65**: 1402–3.
43. Omar MA, Hammond MG, Desai RK, Motala AA, Aboo N, Seedat MA. HLA class I and II antigens in South African blacks with Graves' disease. *Clin Immunol Immunopathol* 1990: **54**: 98–102.
44. Semana G, Mehra NK, Lepetit JC et al. HLA and Graves' disease. In: Charron D, ed. *HLA: Genetic Diversity of HLA; Functional and Medical Implication, Proceedings of the 12th International Histocompatibility Workshop & Conference. Paris, France*: *EDK*, 1997, 407–12.
45. Osame M, Usuku K, Izumo S et al. HTLV-I associated myelopathy, a new clinical entity. *Lancet* 1986: **1**: 1031–2.
46. Nishimura M, Maeda M, Matsuoka M et al. Tumor necrosis factor, tumor necrosis factor receptors type 1 and 2, lymphotoxin-alpha, and HLA-DRB1 gene polymorphisms in human T-cell lymphotropic virus type I associated myelopathy. *Hum Immunol* 2000: **61**: 1262–9.
47. Sonoda S, Fujiyoshi T, Yashiki S. Immunogenetics of HTLV-I/II and associated diseases. *J Acquir Immune Defic Syndr Hum Retrovirol* 1996: **13**: S119–23.
48. Kitze B, Usuku K, Yamano Y et al. Human $CD4^+$ T lymphocytes recognize a highly conserved epitope of human T lymphotropic virus type 1 (HTLV-1) env gp21 restricted by HLA DRB1*0101. *Clin Exp Immunol* 1998: **111**: 278–85.
49. Bhigjee AI, Bill PL, Hammond MG, Windsor IM. HLA profile and HTLV-I associated myelopathy (HAM/TSP) in Natal, South Africa. *J Neurol Neurosurg Psychiatry* 1992: **55**: 329–30.
50. Barreto M, Santos E, Ferreira R et al. Evidence for CTLA4 as a susceptibility gene for systemic lupus erythematosus. *Eur J Hum Genet* 2004: **12**: 620–6.
51. Lee YH, Harley JB, Nath SK. CTLA-4 polymorphisms and systemic lupus erythematosus (SLE): a meta-analysis. *Hum Genet* 2005: **116**: 361–7.
52. Wu J, Wilson J, He J, Xiang L, Schur PH, Mountz JD. Fas ligand mutation in a patient with systemic lupus erythematosus and lymphoproliferative disease. *J Clin Invest* 1996: **98**: 1107–13.
53. Blank MC, Stefanescu RN, Masuda E et al. Decreased transcription of the human FCGR2B gene mediated by the -343 G/C promoter polymorphism and association with systemic lupus erythematosus. *Hum Genet* 2005: **117**: 220–7.
54. Graham RR, Ortmann WA, Langefeld CD et al. Visualizing human leukocyte antigen class II risk haplotypes in human

systemic lupus erythematosus. *Am J Hum Genet* 2002: **71**: 543–53.
55. Howard PF, Hochberg MC, Bias WB, Arnett FC, McLean RH. Relationship between C4 null genes, HLA-D region antigens, and genetic susceptibility to systemic lupus erythematosus in Caucasian and black Americans. *Am J Med 1986: **81**: 187–93.*
56. Azizah MR, Ainol SS, Kong NC, Normaznah Y, Rahim MN. HLA antigens in Malay patients with systemic lupus erythematosus: association with clinical and autoantibody expression. *Korean J Intern Med* 2001: **16**: 123–31.
57. Rudwaleit M, Tikly M, Gibson K, Pile K, Wordsworth P. HLA class II antigens associated with systemic lupus erythematosus in black South Africans. *Ann Rheum Dis* 1995: **54**: 678–80.
58. Klemp P, du Toit ED, Issacs S, Mody GM, Oudshoorn M. HLA-A,B,C, and DR antigens, GLO I and Bf marker profiles in 75 Cape coloured patients with systemic lupus erythematosus (SLE). *Tissue Antigens* 1988: **32**: 12–6.
59. 't Hart BA, Brok HP, Amor S, Bontrop RE. The major histocompatibility complex influences the ethiopathogenesis of MS-like disease in primates at multiple levels. *Hum Immunol* 2001: **62**: 1371–81.
60. Fewster ME, Kies B. HLA antigens in multiple sclerosis in coloured South Africans. *J Neurol Sci* 1984: **66**: 175–81.
61. Barcellos LF, Lin P, Schafer J, Thomson G, Klitz WT. The contribution of two unlinked regions to genetic susceptibility in multiple sclerosis. In: Charron D, ed. *HLA: Genetic Diversity of HLA; Functional and Medical Implication, Proceedings of the 12th International Histocompatibility Workshop & Conference. Paris, France: EDK*, 1997, 725–7.
62. Russell DG. Mycobacterium tuberculosis: here today, and here tomorrow. *Nat Rev Mol Cell Biol* 2001: **2**: 569–77.
63. Meyer CG, May J, Stark K. Human leukocyte antigens in tuberculosis and leprosy. *Trends Microbiol* 1998: **6**: 148–54.
64. Ravikumar M, Dheenadhayalan V, Rajaram K et al. Associations of HLA-DRB1, DQB1 and DPB1 alleles with pulmonary tuberculosis in south India. *Tuber Lung Dis* 1999: **79**: 309–17.
65. Dubaniewicz A, Lewko B, Moszkowska G, Zamorska B, Stepinski J. Molecular subtypes of the HLA-DR antigens in pulmonary tuberculosis. *Int J Infect Dis* 2000: **4**: 129–33.
66. Vejbaesya S, Chierakul N, Luangtrakool K, Srinak D, Stephens HA. Associations of HLA class II alleles with pulmonary tuberculosis in Thais. *Eur J Immunogenet* 2002: **29**: 431–4.
67. Selvaraj P, Uma H, Reetha AM, Xavier T, Prabhakar R, Narayanan PR. Influence of HLA-DR2 phenotype on humoral immunity & lymphocyte response to Mycobacterium tuberculosis culture filtrate antigens in pulmonary tuberculosis. *Indian J Med Res* 1998: **107**: 208–17.
68. Khomenko AG, Litvinov VI, Chukanova VP, Pospelov LE. Tuberculosis in patients with various HLA phenotypes. *Tubercle* 1990: **71**: 187–92.
69. Visentainer JE, Tsuneto LT, Serra MF, Peixoto PR, Petzl-Erler ML. Association of leprosy with HLA-DR2 in a Southern Brazilian population. *Braz J Med Biol Res* 1997: **30**: 51–9.
70. van Eden W, de Vries RR, Mehra NK, Vaidya MC, D'Amaro J, van Rood JJ. HLA segregation of tuberculoid leprosy: confirmation of the DR2 marker. *J Infect Dis* 1980: **141**: 693–701.
71. Goldfeld AE, Delgado JC, Thim S et al. Association of an HLA-DQ allele with clinical tuberculosis. *JAMA* 1998: **279**: 226–8.
72. Park M, Song E, Park H, Kwon S, Han S, Shim Y. HLA-DRB1 and DQB1 gene polymorphism is associated with multidrug-resistant tuberculosis in Korean patients. *Hum Immunol* 2002: **63** (Suppl): S33.
73. Zijenah LS, Hartogensis WE, Katzenstein DA et al. Association of high HIV-1 RNA levels and homozygosity at HLA class II DRB1 in adults coinfected with Mycobacterium tuberculosis in Harare, Zimbabwe. *Hum Immunol* 2002: **63**: 1026–32.
74. Al Jabri AA. HLA and in vitro susceptibility to HIV infection. *Mol Immunol* 2002: **38**: 959–67.
75. Winchester R, Charron D, Louie L et al. The role of HLA in influencing the time of development of a particular outcome of HIV-1 infection. In: Charron D, ed. *Genetic Diversity of HLA Functional and Medical Implication. Paris, France: EDK*, 1997, 423–8.
76. Roe DL, Lewis RE, Cruse JM. Association of HLA-DQ and -DR alleles with protection from or infection with HIV-1. *Exp Mol Pathol* 2000: **68**: 21–8.
77. MacDonald KS, Fowke KR, Kimani J et al. Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. *J Infect Dis* 2000: **181**: 1581–9.
78. Ali SF, Lyagoba B, Biryawaho A et al. MHC polymorphisms associated with HIV infection in Uganda. *Hum Immunol* 1996: **47**: 120.
79. Tang J, Penman-Aguilar A, Lobashevsky E, Allen S, Kaslow RA. HLA-DRB1 and -DQB1 alleles and haplotypes in Zambian couples and their associations with heterosexual transmission of HIV type 1. *J Infect Dis* 2004: **189**: 1696–704.
80. Hill AV, Allsopp CE, Kwiatkowski D et al. Common west African HLA antigens are associated with protection from severe malaria. *Nature* 1991: **352**: 595–600.
81. Thursz MR, Kwiatkowski D, Allsopp CE, Greenwood BM, Thomas HC, Hill AV. Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia. *N Engl J Med* 1995: **332**: 1065–9.
82. Almarri A, Batchelor JR. HLA and hepatitis B infection. *Lancet* 1994: **344**: 1194–5.
83. Bhimma R, Hammond MG, Coovadia HM, Adhikari M, Connolly CA. HLA class I and II in black children with hepatitis B virus-associated membranous nephropathy. *Kidney Int* 2002: **61**: 1510–5.
84. Kwiatkowski D. Science, medicine, and the future: susceptibility to infection. *BMJ* 2000: **321**: 1061–5.
85. Bates JH, Stead WW. The history of tuberculosis as a global epidemic. *Med Clin North Am* 1993: **77**: 1205–17.
86. Hill AV. Immunogenetics and genomics. *Lancet* 2001: **357**: 2037–41.
87. Doolan DL, Southwood S, Chesnut R et al. HLA-DR-promiscuous T cell epitopes from Plasmodium falciparum

pre-erythrocytic-stage antigens restricted by multiple HLA class II alleles. *J Immunol* 2000: **165**: 1123–37.
88. Imanishi T, Akaza T, Kimura A, Tokunaga T, Gojobori T. Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. In: Tsuji K, Aizawa M, Sasazuki T, eds. *HLA 1991. Proceedings of the Eleventh International Histocompatibility Workshop and Conference. Oxford: Oxford University Press*, 1992, 1065–74.
89. du Toit ED, Emmanuel JC, West G, Taljaard DG, Oudshoorn M. HLA frequencies in black Zimbabweans. *Tissue Antigens* 1990: **36**: 122–4.
90. du Toit ED, MacGregor KJ, Taljaard DG, Oudshoorn M. HLA-A, B, C, DR and DQ polymorphisms in three South African population groups: South African negroes, Cape coloureds South African Caucasoids. *Tissue Antigens* 1988: **31**: 109–25.
91. Martell RW, van Eeden PJ, Madden MV, du Toit ED, Dent DM. Lack of an association between carcinoma of the stomach and the major histocompatibility complex (HLA) in Cape Coloureds. *Tissue Antigens* 1988: **32**: 150–4.
92. van Wyk CW, Grobler-Rabie AF, Martell RW, Hammond MG. HLA-antigens in oral submucous fibrosis. *J Oral Pathol Med* 1994: **23**: 23–7.
93. Omar MAK, Hammond M, Motala AA. HLA class II antigens in South African Indians with NIDDM. *Diabetes* 1988: **37**: 796–9.

OPEN ACCESS Freely available online PLoS MEDICINE

Promoter Variation in the DC-SIGN–Encoding Gene *CD209* Is Associated with Tuberculosis

Luis B. Barreiro[1,2], Olivier Neyrolles[2], Chantal L. Babb[3], Ludovic Tailleux[2], Hélène Quach[1], Ken McElreavey[4], Paul D. van Helden[3], Eileen G. Hoal[3], Brigitte Gicquel[2], Lluís Quintana-Murci[1]*

1 CNRS FRE2849, Unit of Molecular Prevention and Therapy of Human Diseases, Institut Pasteur, Paris, France, **2** Unité de Génétique Mycobactérienne, Institut Pasteur, Paris, France, **3** Faculty of Health Sciences, Stellenbosch University, Tygerberg, South Africa, **4** Reproduction, Fertilité et Populations, Institut Pasteur, Paris, France

Competing Interests: The authors have declared that no competing interests exist.

Author Contributions: LBB, ON, BG, and LQ designed the study. LBB, ON, LT, KM, BG, and LQ analyzed the data. LBB, CLB, and HQ performed the sequencing and genotyping analyses. PDH and EGH coordinated the collection, clinical information, and DNA extraction of the cohort samples. LBB, ON, CLB, LT, HQ, KM, PDH, EGH, BG, and LQ contributed to writing the paper.

Citation: Barreiro LB, Neyrolles O, Babb CL, Tailleux L, Quach H, et al. (2006) Promoter variation in the DC-SIGN–encoding gene *CD209* is associated with tuberculosis. PLoS Med 3(2): e20.

Received: July 8, 2005
Accepted: October 19, 2005
Published: January 3, 2006

DOI: 10.1371/journal.pmed.0030020

Abbreviations: DC, dendritic cell; OR, odds ratio; SNP, single nucleotide polymorphism; TB, tuberculosis

* To whom correspondence should be addressed. E-mail: quintana@pasteur.fr

ABSTRACT

Background

Tuberculosis, which is caused by *Mycobacterium tuberculosis*, remains one of the leading causes of mortality worldwide. The C-type lectin DC-SIGN is known to be the major *M. tuberculosis* receptor on human dendritic cells. We reasoned that if DC-SIGN interacts with *M. tuberculosis*, as well as with other pathogens, variation in this gene might have a broad range of influence in the pathogenesis of a number of infectious diseases, including tuberculosis.

Methods and Findings

We tested whether polymorphisms in *CD209*, the gene encoding DC-SIGN, are associated with susceptibility to tuberculosis through sequencing and genotyping analyses in a South African cohort. After exclusion of significant population stratification in our cohort, we observed an association between two *CD209* promoter variants (–871G and –336A) and decreased risk of developing tuberculosis. By looking at the geographical distribution of these variants, we observed that their allelic combination is mainly confined to Eurasian populations.

Conclusions

Our observations suggest that the two –871G and –336A variants confer protection against tuberculosis. In addition, the geographic distribution of these two alleles, together with their phylogenetic status, suggest that they may have increased in frequency in non-African populations as a result of host genetic adaptation to a longer history of exposure to tuberculosis. Further characterization of the biological consequences of DC-SIGN variation in tuberculosis will be crucial to better appreciate the role of this lectin in interactions between the host immune system and the tubercle bacillus as well as other pathogens.

Introduction

One-third of the world's population is estimated to be infected with *Mycobacterium tuberculosis,* the etiological agent of tuberculosis (TB). This disease tops the World Health Organization list of deaths due to a single infectious agent, with the death toll between 2 and 3 million people per year [1]. A perplexing, and yet unsolved, feature of TB is that less than 10% of infected individuals develop the disease. Substantial epidemiological evidence supports that host-related factors, such as sex, age, HIV infection, malnutrition, and BCG (bacille Calmette-Guérin) vaccination, influence the balance between the tubercle bacilli and host immune defences [2,3]. In addition, there is increasing evidence that host genetic factors determine differences in host susceptibility to mycobacterial infection and might contribute therefore to the pattern of clinical disease [4–7]. From a host perspective, the innate immunity system acts as the first line of host defense against microbial pathogens [8]. Initial recognition of pathogens by the innate immunity system is mediated by phagocytic cells, such as dendritic cells (DCs) or macrophages, through germline-encoded receptors, known as pattern recognition receptors [9]. DCs bear a range of pattern recognition receptors, such as C-type lectins and Toll-like receptors, involved both in recognition of conserved products of microbial metabolism and in the induction of adaptive immunity [8,10–12]. In particular, C-type lectins detect pathogens by their characteristic carbohydrate structures and internalise them for further antigen processing and presentation [13]. We have recently shown that a prototypic C-type lectin, DC-SIGN (dendritic cell–specific ICAM-3 grabbing nonintegrin), is the major *Mycobacterium tuberculosis* receptor on human DCs [14]. DC-SIGN is specifically, though not exclusively, expressed on DCs and functions both as a cell adhesion and as a pathogen recognition receptor [15]. As an adhesion receptor, it plays an important role in many DC functions, such as DC-T cell interaction and DC migration [16,17]. Besides its cellular recognition role, DC-SIGN serves as pathogen uptake receptor and mediates interactions with a plethora of pathogens other than *M. tuberculosis* [18]. Indeed, it has been shown that DC-SIGN allows DCs to capture other bacteria such as *Helicobacter pylori* and certain *Klebsiella pneumonia* strains, but also viruses such as HIV-1, Ebola, cytomegalovirus, hepatitis-C, dengue, and SARS-coV, and parasites like *Leishmania pifanoi* and *Schistosoma mansoni* [19–27]. In addition, recent data suggest that DC-SIGN may mediate intracellular signalling events leading to cytokine secretion and, on this basis, it has been proposed that the lectin could be used by pathogens, including *M. tuberculosis,* as a part of an immune evasion strategy to their own advantage [28,29].

In light of the ability of DC-SIGN to interact with *M. tuberculosis* and other pathogens, it is plausible that variation in its gene may influence the pathogenesis of a number of infectious diseases, including TB. We have therefore explored the relationship between *CD209* polymorphisms and susceptibility to TB by determining *CD209* sequence variation in a cohort of South African Coloured origin.

Methods

Patients and Methods

The study was conducted in a cohort of 711 individuals, including 351 TB patients and 360 healthy controls, living in the Cape Town area. Certain suburbs of metropolitan Cape Town have some of the highest reported incidence rates of TB in the world, despite extensive BCG vaccination. Indeed, our study population comes from two suburbs that have been extensively studied because of their uniform ethnicity (known as South African Coloured) and socio-economic status as well as high incidence of TB and low prevalence of HIV [30]. In addition, our study group represents a present-day homogenous population [31] that previously received genetic input from Khoisan, Malaysian, Bantu, and European descent populations [32]. Thus, it represents a community originating from populations with different susceptibilities to TB and offers a unique opportunity to dissect the contributing genetic variants and their probable geographic/ethnic origins. TB patients were bacteriologically-confirmed (smear-positive and/or culture-positive) to present pulmonary tuberculosis. Their mean age (± standard deviation) was 36.7 ± 10.9 y, and 51.8% were male. Controls were unrelated healthy individuals from the same community, with the same socio-economic status, access to health facilities, and chance of diagnosis, and with neither signs nor previous history of TB (mean age 34.6 ± 12.5 y, 22% male). The annual risk of infection in this suburb was estimated at 2.5% in 1987 and at 2.8%–3.5% in 1999, and it is therefore highly likely that, in such an environment, the vast majority of controls have been exposed to *M. tuberculosis* [33,34]. All subjects were HIV-negative and older than 18 y. Informed consent was obtained from all participants, and the study was approved by the ethics committee of the Faculty of Health Sciences, Stellenbosch University (South Africa).

Laboratory Procedures and Statistical Analysis

To identify informative *CD209* single nucleotide polymorphisms (SNPs) and to avoid ascertainment bias in the choice of markers to be tested, we first sequenced the whole *CD209* genomic region (seven coding exons, flanking intronic regions, and 1,000 base pairs situated 5′ of the start codon) in 28 randomly chosen individuals (56 chromosomes). Using polymorphisms with a minimum allele frequency of 0.05, unphased genotypic data were converted into haplotypes using the accelerated EM (Expectation Maximization) algorithm implemented in Haploview v3.1 [35]. To evaluate the accuracy of the EM algorithm, haplotype reconstruction was performed in parallel using the Bayesian statistical method [36] implemented in Phase v.2.1.1. Equivalent results were obtained using both methods, with all haplotypes presenting high levels of statistical support. In order to define a minimal number of SNPs explaining most haplotypic diversity, we used the BEST v1.0 software [37]. Eight haplotype-tagging SNPs were then selected to genotype the entire panel of 711 individuals. Further, potential population stratification between cases and controls was tested by genotyping 25 unlinked SNP markers in the entire study cohort. DNA samples were genotyped by either fluorescence polarization (VICTOR-2TM technology; PerkinElmer, Wellesley, California, United States) or TaqMan (ABI Prism-7000 Sequence Detection System; Applied Biosystems, Foster City, California, United States) assays. Statistical testing for genotypic and haplotypic associations were performed using STATA 8.2 and Haploview v3.1, respectively. The haplotype frequencies were obtained by summing the fractional likelihood of each haplotype for each individual (i.e., if a particular individual has been determined to have a 40% likelihood of haplotype A

and 60% likelihood of haplotype B, 0.4 and 0.6 would be added to the counts for A and B, respectively) [35].

Results and Discussion

Two variants located in the *CD209* promoter region (−871 A/G and −336 A/G) exhibited a frequency distribution significantly distorted between TB patients and controls, as indicated by a Chi-square test (Table 1). For the −871 variant, genotypes GG and GA were less frequently observed in cases (16.8%) compared to the control group (27.2%) ($p = 8.2 \times 10^{-4}$). For the −336 variant, genotypes GG and GA were more frequent in cases (70.6%) than in controls (61.9%) ($p = 0.01$). These observations suggest that the alleles −871A (odds ratio [OR]: 1.85; 95% CI: 1.29–2.66) and −336G (OR: 1.48; 95% CI: 1.08–2.02) increase the risk of developing TB in our South African cohort. At the haplotype level (Table 2), a Chi-square test first revealed that the global distribution of haplotype frequencies was significantly different between cases and controls ($p = 1.2 \times 10^{-3}$). One haplotype (H3) turned out to be the main haplotype responsible for such a distorted frequency distribution (Table 2). This haplotype, which contains both −871G and −336A, was found to be strongly associated with the control group ($p = 1.6 \times 10^{-3}$; OR: 1.7; 95% CI: 1.22–2.38). The associations with this haplotype, and with −871, remained highly significant ($p = 1.3 \times 10^{-2}$ and 6.6×10^{-3} respectively), even after the conservative Bonferroni correction for multiple testing.

Although our cohort is considered a present-day homogeneous community that has received genetic contribution from different populations multiple generations ago [31,32], population stratification between cases and controls can be a confounding factor leading to a spurious positive association. Indeed, the use of admixed populations in association-mapping studies can be very useful to identify disease-causing genetic variants that differ in frequency across parental populations. However, when the admixture event is too recent, allelic frequencies can differ coincidentally among cases and controls, reflecting a nonuniform genetic contribution from the parental populations to each subpopulation (i.e., cases and controls), rather than a genuine association between a given genetic variant and the phenotype under study. In this case, the study-cohort is said to present population stratification. To formally test and quantify the levels of background genetic differences [38], if any, between cases and controls, we genotyped the entire cohort for a panel of 25 independent SNPs markers which are (1) not in linkage disequilibrium with the candidate *CD209* locus and with any other known gene, (2) randomly distributed along the genome, and (3) polymorphic among the major ethnic groups (Table 3). The mean χ^2 statistic among the 25 SNPs for the comparison of allele frequencies between cases and controls, which represents the levels of stratification (μ) between the two groups [39], was 1.25 ($p = 0.26$), implying that the two groups were not significantly stratified. As an additional correction for stratification, we divided the χ^2 values obtained for our candidate gene *CD209* by the level of stratification detected (1.25) [39]. Even after such a conservative correction, the associations observed with −336 and −871 as well as with H3 remained significant (−336 $p = 2.8 \times 10^{-2}$; −871 $p = 2.7 \times 10^{-3}$; H3 $p = 4.8 \times 10^{-3}$). These observations support therefore the idea that the −871G and −336A variants are indeed genuinely associated with a protective role against TB.

In order to gain insights into the frequency distribution of these two SNPs, we genotyped them in 254 human chromosomes from sub-Saharan Africa, Europe, and East Asia as well as in eight chimpanzee chromosomes. We observed that the −871G and −336A forms, which we propose as offering protection against TB, corresponded to the derived allele in humans; we also observed that these forms are present at higher frequencies in Eurasians as compared to Africans (Table 4). Indeed, the −871G is absent in African populations whereas it reaches high frequencies (20%–40%) in European and Asian populations. Given the absence of the haplotypic combination of −871G and −336A among sub-Saharan Africans, its presence among South African Coloureds suggests that it was introduced through the historically well-known admixture with Europeans and Asians [31]. This observation highlights the power of using admixed populations to better understand historical issues associated with the geographic/ethnic origin of disease-affecting alleles, provided that their prevalence varies in the ancestors of the admixed population (i.e., different frequency of H3 in Africans versus non-Africans; Table 4).

In the context of TB, it has been suggested that present-day

Table 1. DC-SIGN Genotype Distributions in Patients with Tuberculosis and in Healthy Controls

				Genotype Frequencies (%)							
				Cases ($n = 351$)			Controls ($n = 360$)			Cases vs. Controls[b]	
htSNP[a]	1		2	11	12	22	11	12	22	*p*	OR
−939	G	>	A	52.1	39	8.8	45.3	46.9	7.8	0.07	0.76 [0.57–1.02]
−871	A	>	G	83.2	15.7	1.1	72.8	26.1	1.1	8.2×10^{-4}	1.85 [1.29–2.66]
−336	A	>	G	29.3	50.4	20.2	38.1	43.3	18.6	0.01	1.48 [1.08–2.02]
−139	A	>	G	8.3	37.6	54.1	7.2	43.1	49.7	0.24	0.84 [0.62–1.13]
2392	G	>	A	96.6	3.4	0	95	4.7	0.3	0.29	0.67 [0.32–1.42]
3220	T	>	C	74.6	23.4	2	75.8	22.8	1.4	0.71	1.07 [0.76–1.50]
3838	A	>	C	80.1	17.9	2	84.7	14.4	0.8	0.1	1.38 [0.94–2.04]
4235	G	>	C	51.6	41	7.4	57.2	37.8	5	0.13	1.26 [0.94–1.69]

[a]All htSNPs (haplotype tagging SNPs) were in Hardy-Weinberg *equilibrium* in both the global sample and in cases and controls, separately.
[b]The homozygotes for the most frequent allele were compared with the sum of the homozygotes and heterozygotes for the rare allele.
DOI: 10.1371/journal.pmed.0030020.t001

Table 2. Haplotype Distributions in Patients with Tuberculosis and in Healthy Controls

Haplotype[a]	Haplotype Frequencies (%)		Cases vs. Controls[b]	
	Cases	Controls	*p*	OR
H1-AAAGGTCG	10	6.2	0.01	1.70 [1.15–2.52]
H2-AAAGGTAG	7.6	9.3	0.25	0.80 [0.55–1.16]
H3-AGAGGTAG	8.9	14.2	1.6×10^{-3}	1.70 [1.22–2.38]
H4-GAAAGTAG	26	27.9	0.41	0.91 [0.72–1.15]
H5-GAGGGTAG	17.3	15	0.25	1.17 [0.88–1.55]
H6-GAGGGTAC	12.2	8.5	0.02	1.49 [1.05–2.11]
H7-GAGGATAC	1.7	2.2	0.44	0.77 [0.36–1.63]
H8-GAGGGCAC	12.8	11.2	0.35	1.16 [0.84–1.60]

[a]Haplotypes with frequency greater than 1%. The alleles are ordered from SNP-939 until SNP4235.
[b]The frequency of each haplotype in cases and controls was compared with the sum of all the others.
DOI: 10.1371/journal.pmed.0030020.t002

susceptibility to TB is determined by previous history of exposure [40]. There is fairly convincing evidence that TB has been endemic in Europe for several hundred years, whereas in Africa it has probably been rare before contact was initiated with Europeans [41–43]. It is expected therefore that *M. tuberculosis* has exerted stronger selective pressures on European than African populations [42]. Our results lend support to this hypothesis and suggest that the protective alleles –871G and –336A increased in frequency in non-African populations as a result of genetic adaptation to a longer period of TB exposure. The potential impact of tuberculosis on the frequency of resistant alleles in European populations has been recently addressed using epidemiological data and statistical modeling [44]. The authors have sought to evaluate the expected changes in resistant allele frequencies, during the 300-y period corresponding to the peak epidemics of TB in Europe. They concluded that if a given resistant allele was at a low frequency in the beginning of an epidemic, selection by *M. tuberculosis* alone would increase the frequency of this allele, but not enough to bring it to epidemiologically significant levels. In this context, since DC-SIGN is known to interact with a vast range of pathogens, it is indeed likely that

Table 3. Frequency Distribution in the Study Cohort of the 25 SNPs Used to Test for Population Stratification

Name	rs Number	Location	ObsHET[a]	PredHET[b]	HWP[c]	MAF	MAF	χ^2	*p*-Value
						Cases (%)	Controls (%)		
SNP1	rs2048022	Chr4	0.509	0.497	0.588	43.4	48.3	3.469	0.0625
SNP2	rs1380229	Chr8	0.429	0.468	0.034	36.1	38.4	0.815	0.3667
SNP3	rs650389	Chr10	0.276	0.287	0.369	15.2	19.5	4.518	0.0335
SNP4	rs870384	Chr12	0.511	0.499	0.591	48.3	46.8	0.311	0.5771
SNP5	rs695982	Chr12	0.416	0.419	0.888	32.1	27.7	3.376	0.0662
SNP6	rs708682	Chr15	0.216	0.212	0.826	11.8	12.3	0.086	0.7693
SNP7	rs715774	Chr15	0.244	0.257	0.209	14.3	16.1	0.883	0.3473
SNP8	rs1433456	Chr15	0.313	0.317	0.784	19.9	19.7	0.011	0.9169
SNP9	rs807131	Chr17	0.478	0.453	0.183	35.5	33.9	0.398	0.5279
SNP10	rs11672183	Chr19	0.214	0.209	0.649	12	11.7	0.018	0.8947
SNP11	rs2024628	Chr20	0.513	0.494	0.333	42.2	46.5	2.627	0.1051
SNP12	rs1028184	Chr2	0.461	0.464	0.9	34.2	39	3.53	0.0603
SNP13	rs2056773	Chr3	0.492	0.473	0.32	39.5	37.1	0.808	0.3687
SNP14	rs1479067	Chr5	0.398	0.395	0.972	25.9	28.4	1.099	0.2945
SNP15	rs327747	Chr7	0.374	0.399	0.111	25.8	29.2	2.084	0.1488
SNP16	rs12665321	Chr6	0.239	0.234	0.693	14.2	12.9	0.462	0.4969
SNP17	rs1566838	Chr9	0.475	0.497	0.265	46.5	45.8	0.081	0.7762
SNP18	rs12785524	Chr11	0.452	0.483	0.106	39	42.4	1.708	0.1912
SNP19	rs975423	Chr13	0.429	0.464	0.056	35.1	37.8	1.134	0.2868
SNP20	rs914904	Chr13	0.397	0.409	0.462	29.2	28.2	0.179	0.6723
SNP21	rs876287	Chr14	0.459	0.484	0.182	41.3	40.9	0.024	0.8765
SNP22	rs1582598	Chr16	0.378	0.394	0.321	27.5	26.5	0.164	0.6852
SNP23	rs1364198	Chr16	0.361	0.364	0.908	25.2	22.7	1.242	0.2652
SNP24	rs739259	Chr22	0.469	0.469	1	36.1	39	1.217	0.27
SNP25	rs169479	Chr21	0.217	0.217	1	13.3	11.5	1.064	0.3024

[a]Observed heterozygosity.
[b]Predicted heterozygosity.
[c]Hardy–Weinberg equilibrium probability.
MAF, minimum allele frequency.
DOI: 10.1371/journal.pmed.0030020.t003

Table 4. Frequency Distribution of the Eight *CD209* SNPs Genotyped in the Multi-Ethnic Panel of 127 Individuals ($n = 254$ Chromosomes) as Well as in the 711 Individuals of the South African Cohort ($n = 1,422$ Chromosomes)

SNP	Variant[a]	Population Frequencies (%)			
		African ($n = 82$)	Asian ($n = 86$)	European ($n = 86$)	SAC ($n = 1,422$)
−939	A	45.1	29.1	54.7	29.8
	G	54.9	70.9	45.3	70.2
−871	A	100	79.1	61.6	88.4
	G	0	20.9	38.4	11.6
−336	G	37.8	5.8	20.9	42.8
	A	62.2	94.2	79.1	57.2
−139	G	87.8	33.7	75.6	72.1
	A	12.2	66.3	24.4	27.9
2,392	G	100	98.8	91.9	97.8
	A	0	1.2	8.1	2.2
3,220	C	14.6	0	0	13.2
	T	85.4	100	100	86.8
3,838	A	70.7	100	90.7	90.5
	C	29.3	0	9.3	9.5
4,235	G	76.8	96.5	86	74.1
	C	23.2	3.5	14	25.9

[a]The allele in bold corresponds to the derived allele when compared with the sequence of the chimpanzee.
SAC, South African coloured population.
DOI: 10.1371/journal.pmed.0030020.t004

the increased frequencies observed today for both −871G and −336A in non-African populations (specially for −871G which is absent in sub-Saharan Africans) may have been driven, not only by the selective pressures imposed by *M. tuberculosis*, but also by other infectious agents. Indeed, two independent studies have recently reported a genetic association between the −336A variant and protection against parenteral HIV infection [45] and severity of dengue pathogenesis [46]. Although HIV infection, for example, is too recent to have left any signature of selection on *CD209*, these observations emphasize the possible action of other pathogens in shaping the patterns of variability of this gene.

From a functional point of view, the −336A allele has been shown to affect an Sp1-like binding site and to modulate transcriptional activity in vitro by increasing the levels of expression [46]. In the context of TB, increased DC-SIGN expression levels by DCs may result in better capture and processing of mycobacterial antigens, leading to a stronger and wider T-cell response. In addition, we have recently shown that DC-SIGN expression is markedly induced in alveolar macrophages in active TB patients and that *M. tuberculosis* is preferentially phagocytosed by DC-SIGN–expressing macrophages in these individuals [47]. Thus, the higher prevalence observed among healthy individuals of the −336A variant, which is associated with increased DC-SIGN expression, may underlie an increased efficiency of host phagocytes, such as DCs and macrophages, to control the infection. In addition to the −336A variant, our genetic data showed a strong association of the −871G allele with healthy controls, suggesting also a functional consequence of this variant that, either alone or in combination with −336A, remains to be defined.

In conclusion, the significant association found for the *CD209* promoter variants together with their phylogenetic status and frequency distribution strongly suggests that the −871G and −336A alleles may reduce the risk of developing TB. More generally, our results, together with those reporting association of *CD209* promoter variants with both HIV susceptibility and dengue pathogenesis [45,46] suggest that variation in this lectin may be of crucial importance in the outcome of a number of infections due to DC-SIGN–interacting pathogens. Detailed in vitro and in vivo studies assessing the functional consequences of *CD209* variants on the quality of the host immune response against pathogens, including *M. tuberculosis*, are now required to eventually develop knowledge-based and effective pathway-targeted treatments.

Acknowledgments

LBB was supported by a "Fundação para a Ciência e a Tecnologia" fellowship (SFRH/BD/18580/2004). This research project has been co-financed by the European Commission, within the 6th Framework Programme (contract no. LSHP-CT-2003–503367). The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information. We thank M. Kennedy for help with sample collection, and the South African Medical Research Council and Department of Science and Technology/National Research Foundation (DST/NRF) Centres of Excellence for financial assistance. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. ■

References

1. Frieden TR, Sterling TR, Munsiff SS, Watt CJ, Dye C (2003) Tuberculosis. Lancet 362: 887–899.
2. Enarson DA, Rouillon A (1994) The epidemiological basis of tuberculosis control. In: Davies PD, editor. Clinical tuberculosis. London: Chapman and Hall. pp 19–32.
3. Lienhardt C (2001) From exposure to disease: The role of environmental factors in susceptibility to and development of tuberculosis. Epidemiol Rev 23: 288–301.
4. Bellamy R (2003) Susceptibility to mycobacterial infections: The importance of host genetics. Genes Immun 4: 4–11.
5. Blackwell JM (2001) Genetics and genomics in infectious disease susceptibility. Trends Mol Med 7: 521–526.
6. Cooke GS, Hill AV (2001) Genetics of susceptibility to human infectious disease. Nat Rev Genet 2: 967–977.
7. Casanova JL, Abel L (2002) Genetic dissection of immunity to mycobacteria: The human model. Annu Rev Immunol 20: 581–620.
8. Janeway CA Jr., Medzhitov R (2002) Innate immune recognition. Annu Rev Immunol 20: 197–216.

9. Kimbrell DA, Beutler B (2001) The evolution and genetics of innate immunity. Nat Rev Genet 2: 256–267.
10. Taylor PR, Martinez-Pomares L, Stacey M, Lin HH, Brown GD, et al. (2005) Macrophage receptors and immune recognition. Annu Rev Immunol 23: 901–944.
11. Medzhitov R (2001) Toll-like receptors and innate immunity. Nat Rev Immunol 1: 135–145.
12. Fujita T (2002) Evolution of the lectin-complement pathway and its role in innate immunity. Nat Rev Immunol 2: 346–353.
13. McGreal EP, Miller JL, Gordon S (2005) Ligand recognition by antigen-presenting cell C-type lectin receptors. Curr Opin Immunol 17: 18–24.
14. Tailleux L, Schwartz O, Herrmann J, Pivert E, Jackson M, et al. (2003) DC-SIGN is the major *Mycobacterium tuberculosis* receptor on human dendritic cells. J Exp Med 197: 121–127.
15. Soilleux EJ, Morris LS, Leslie G, Chehimi J, Luo Q, et al. (2002) Constitutive and induced expression of DC-SIGN on dendritic cell and macrophage subpopulations in situ and in vitro. J Leukoc Biol 71: 445–457.
16. Geijtenbeek TB, Torensma R, van Vliet SJ, van Duijnhoven GC, Adema GJ, et al. (2000) Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. Cell 100: 575–585.
17. Geijtenbeek TB, Krooshoop DJ, Bleijs DA, van Vliet SJ, van Duijnhoven GC, et al. (2000) DC-SIGN-ICAM-2 interaction mediates dendritic cell trafficking. Nat Immunol 1: 353–357.
18. Geijtenbeek TB, van Vliet SJ, Engering A, 't Hart BA, van Kooyk Y (2004) Self- and nonself-recognition by C-type lectins on dendritic cells. Annu Rev Immunol 22: 33–54.
19. Geijtenbeek TB, Kwon DS, Torensma R, van Vliet SJ, van Duijnhoven GC, et al. (2000) DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. Cell 100: 587–597.
20. Alvarez CP, Lasala F, Carrillo J, Muniz O, Corbi AL, et al. (2002) C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans. J Virol 76: 6841–6844.
21. Halary F, Amara A, Lortat-Jacob H, Messerle M, Delaunay T, et al. (2002) Human cytomegalovirus binding to DC-SIGN is required for dendritic cell infection and target cell trans-infection. Immunity 17: 653–664.
22. Lozach PY, Lortat-Jacob H, de Lacroix de Lavalette A, Staropoli I, Foung S, et al. (2003) DC-SIGN and L-SIGN are high-affinity binding receptors for hepatitis C virus glycoprotein E2. J Biol Chem 278: 20358–20366.
23. Tassaneetrithep B, Burgess TH, Granelli-Piperno A, Trumpfheller C, Finke J, et al. (2003) DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. J Exp Med 197: 823–829.
24. Marzi A, Gramberg T, Simmons G, Moller P, Rennekamp AJ, et al. (2004) DC-SIGN and DC-SIGNR interact with the glycoprotein of Marburg virus and the S protein of severe acute respiratory syndrome coronavirus. J Virol 78: 12090–12095.
25. Bergman MP, Engering A, Smits HH, van Vliet SJ, van Bodegraven AA, et al. (2004) *Helicobacter pylori* modulates the T helper cell 1/T helper cell 2 balance through phase-variable interaction between lipopolysaccharide and DC-SIGN. J Exp Med 200: 979–990.
26. Appelmelk BJ, van Die I, van Vliet SJ, Vandenbroucke-Grauls CM, Geijtenbeek TB, et al. (2003) Carbohydrate profiling identifies new pathogens that interact with dendritic cell-specific ICAM-3-grabbing nonintegrin on dendritic cells. J Immunol 170: 1635–1639.
27. Colmenares M, Puig-Kroger A, Pello OM, Corbi AL, Rivas L (2002) Dendritic cell (DC)-specific intercellular adhesion molecule 3 (ICAM-3)-grabbing nonintegrin (DC-SIGN, CD209), a C-type surface lectin in human DCs, is a receptor for *Leishmania amastigotes*. J Biol Chem 277: 36766–36769.
28. Geijtenbeek T, van Vliet S, Koppel E, Sanchez-Hernandez M, Vandenbroucke-Grauls C, et al. (2003) Mycobacteria target DC-SIGN to suppress dendritic cell function. J Exp Med 197: 7–17.
29. Tailleux L, Maeda N, Nigou J, Gicquel B, Neyrolles O (2003) How is the phagocyte lectin keyboard played? Master class lesson by *Mycobacterium tuberculosis*. Trends Microbiol 11: 259–263.
30. Beyers N, Gie RP, Zietsman HL, Kunneke M, Hauman J, et al. (1996) The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high-incidence community. S Afr Med J 86: 40–41, 44.
31. van der Ross RE (1993) 100 questions about Coloured South Africans. Cape Town (South Africa): UWC Printing Department. 36 p.
32. Nurse GT, Weiner JS, Jenkins T (1985) The peoples of Southern Africa and their affinities. Oxford (United Kingdom): Clarendon Press. 410 p.
33. Munch Z, Van Lill SW, Booysen CN, Zietsman HL, Enarson DA, et al. (2003) Tuberculosis transmission patterns in a high-incidence area: A spatial analysis. Int J Tuberc Lung Dis 7: 271–277.
34. Beyers N, Michaelis I, Gie R, Schaaf HS, Richardson M, et al. (2001) Transmission of tuberculosis (TB) to children in a high incidence area. Int J Tuberc Lung Dis 5: S185.
35. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265.
36. Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73: 1162–1169.
37. Sebastiani P, Lazarus R, Weiss ST, Kunkel LM, Kohane IS, et al. (2003) Minimal haplotype tagging. Proc Natl Acad Sci U S A 100: 9900–9905.
38. Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65: 220–228.
39. Reich DE, Goldstein DB (2001) Detecting association in a case-control study while correcting for population stratification. Genet Epidemiol 20: 4–16.
40. O'Brien SJ (1991) Ghetto legacy. Curr Biol 1: 209–211.
41. Stead WW, Senner JW, Reddick WT, Lofgren JP (1990) Racial differences in susceptibility to infection by *Mycobacterium tuberculosis*. N Engl J Med 322: 422–427.
42. Stead WW (1992) Genetics and resistance to tuberculosis. Could resistance be enhanced by genetic engineering? Ann Intern Med 116: 937–941.
43. Daniel TM, Bates JH, Downes KA (1994) History of tuberculosis. In: Bloom BR, editor. Tuberculosis: Pathogenesis, protection, control. Washington, DC: ASM Press. pp 13–24.
44. Lipsitch M, Sousa AO (2002) Historical intensity of natural selection for resistance to tuberculosis. Genetics 161: 1599–1607.
45. Martin MP, Lederman MM, Hutcheson HB, Goedert JJ, Nelson GW, et al. (2004) Association of DC-SIGN promoter polymorphism with increased risk for parenteral, but not mucosal, acquisition of human immunodeficiency virus type 1 infection. J Virol 78: 14053–14056.
46. Sakuntabhai A, Turbpaiboon C, Casademont I, Chuansumrit A, Lowhnoo T, et al. (2005) A variant in the CD209 promoter is associated with severity of dengue disease. Nat Genet 37: 507–513.
47. Tailleux L, Pham-Thi N, Bergeron-Lafaurie N, Herrmann JL, Charles P, et al. (2005) DC-SIGN induction in alveolar macrophages defines a privileged target host cell population for mycobacteria in patients with tuberculosis. PLoS Med (2)12: e381.

Patient Summary

Background. Tuberculosis (TB) is an infectious disease caused by the bacterium *Mycobacterium tuberculosis*. The disease kills between 2 million and 3 million people each year, and almost one-third of the world's population, or 2 billion people, are thought to be infected with the bacterium. However, only about 10% of infected individuals become sick. In the other cases, the body's immune system is capable of controlling the growth of the bacteria. Whether a person's immune system is strong enough to keep the bacterium in check depends on a variety of factors, such as age, how well nourished the person is, other infections, and genetic make-up.

Why Was This Study Done? Understanding the genetic factors that influence whether an infected person is more or less susceptible to getting sick with TB should help doctors to fight the disease. A number of research groups are studying these genetic factors, and some genes have previously been identified that seem to increase a person's risk of TB. Some of the authors of this study work on a molecule that is part of the immune system. The molecule is called DC-SIGN, and it plays a role in infection of cells in the lung by the Mycobacterium. Together with colleagues who are genetics experts, they wanted to see whether variation in the gene (called *CD209*) that codes for DC-SIGN influences the risk of someone infected with *M. tuberculosis* getting sick.

What Did the Researchers Do and Find? They worked with a group of South African participants from an area close to Cape Town where TB is very common. Half of the participants had TB; the other half did not (even though, because TB is so common in that area, it was likely that most if not all of them had been infected with *M. tuberculosis* at some point). The researchers compared make-up of the gene for DC-SIGN in all participants and found that some variants of the gene were more common in the group of healthy individuals. In other words, having this particular type of DC-SIGN-encoding gene seemed to protect individuals from getting TB.

What Does This Mean? DC-SIGN is a central part of the immune system, and others scientists have reported links between variants in the gene for DC-SIGN and the risk of picking up other infectious diseases, including dengue fever and HIV/AIDS. The present study lends more support to the notion that DC-SIGN is a key player in the control of infectious diseases. Understanding more about DC-SIGN could help to develop better treatments for these infections.

Where Can I Find More Information Online? The following Web sites provide information about tuberculosis.
World Health Organization pages on TB:
http://www.who.int/mediacentre/factsheets/fs104/en/index.html
OMNI pages on TB:
http://omni.ac.uk/browse/mesh/D014376.html
Medline PLUS pages on TB:
http://medlineplus.nlm.nih.gov/medlineplus/tuberculosis.html
European TB Vaccine Cluster:
http://www.tb-vac.org

ELSEVIER

Length Variation of DC-SIGN and L-SIGN Neck-Region has no Impact on Tuberculosis Susceptibility

Luis B. Barreiro, Olivier Neyrolles, Chantal L. Babb, Paul D. van Helden, Brigitte Gicquel, Eileen G. Hoal, and Lluís Quintana-Murci

ABSTRACT: The C-type lectins DC-SIGN and L-SIGN are important pathogen-recognition receptors of the human innate immune system. Both lectins have been shown to interact with a vast range of infectious agents, including *Mycobacterium tuberculosis*, the etiologic agent of tuberculosis in humans. In addition, DC-SIGN and L-SIGN possess a neck region, made up of a variable number of 23 amino acid tandem repeats, which plays a crucial role in the tetramerization of these proteins and support of the carbohydrate recognition domain. The length of the neck region, which shows variable levels of polymorphism, can critically influence the pathogen binding properties of these two receptors. We therefore investigated the impact of the DC-SIGN and L-SIGN neck-region length variation on the outcome of tuberculosis by screening this polymorphism in a large cohort of Coloured South African origin. The analyses of 711 individuals, including 351 tuberculosis patients and 360 healthy controls, revealed that none of the DC-SIGN and L-SIGN neck-region variants or genotypes seems to influence the individual susceptibility to develop tuberculosis. *Human Immunology 68, 106–112 (2007).* © American Society for Histocompatibility and Immunogenetics, 2007. Published by Elsevier Inc.

KEYWORDS: Tuberculosis; susceptibility; DC-SIGN; L-SIGN; neck region; genetics

ABBREVIATIONS

DC-SIGN	dendritic cell-specific ICAM-3 grabbing nonintegrin
L-SIGN	dendritic cell-specific ICAM-3 grabbing nonintegrin related
PRRs	pattern recognition receptors
TB	Tuberculosis

INTRODUCTION

The innate immune system is the first line of host defense against pathogens [1]. Early recognition and uptake of microbes by host professional phagocytes, such as macrophages and dendritic cells, are crucial for downstream immune responses and pathogen clearance. Phagocytic cells express a range of cellular receptors, known as pattern recognition receptors (PRRs), involved in the sensing of microorganisms [1]. These proteins bind to conserved microbial ligands, promoting phagocytosis and antigen presentation, and trigger intracellular signaling and cytokine secretion. The quality of this initial pathogen recognition can have important consequences in both the outcome of infection and the pathogenesis of infectious disease. Two particular PRRs of the C-type lectin receptor family—dendritic cell-specific intercellular adhesion molecule (ICAM)-3 grabbing nonintegrin (DC-SIGN) and dendritic cell-specific ICAM-3 grabbing nonintegrin related (L-SIGN, also known as DC-SIGNR)—have recently been the focus of considerable attention [2–6]. These two lectins, which can act as both cell adhesion receptors and pathogen recognition receptors, are encoded by two genes located on chromosome 19p13.2-3 within a ~26 kb segment [7, 8]. DC-SIGN and L-SIGN exhibit high nucleotide (73%) and

From the CNRS FRE2849, Unit of Human Evolutionary Genetics (L.B.B., L.Q.-M.), and Unité de Génétique Mycobactérienne (L.B.B., O.N., B.G.), Institut Pasteur, Paris, France, and Faculty of Health Sciences (C.L.B., P.D.v.H., E.G.H.), Stellenbosch University, Tygerberg, South Africa.

Address reprint requests to: Lluís Quintana-Murci, CNRS FRE2849, Unit of Human Evolutionary Genetics, Institut Pasteur, 25 rue Dr Roux, 75015 Paris; Tel +33-1-45-68-89-20; Fax: +33-1-45-68-86-39; E-mail: quintana@pasteur.fr

Human Immunology 68, 106–112 (2007)

0198-8859/07/$–see front matter
doi:10.1016/j.humimm.2006.10.020

aminoacid (77%) identity, and identical exon-intron organization [8]. An additional characteristic of both lectins is the presence of a neck region, made up of primarily 7 highly conserved 23 amino acid repeats, that separates the carbohydrate recognition domain (CRD) involved in pathogen binding from the transmembrane region. In regard to expression profiles, DC-SIGN is expressed mainly on endocytic cells, such as dendritic cells and macrophages, whereas L-SIGN is expressed on endothelial cells in liver and lymph nodes, and in cells lining placental capillaries [9–11]. DC-SIGN and L-SIGN share the ability to bind high–mannose oligosaccharides through their CRD, and have been shown to recognize a vast range of microbes, such as HIV-1, Ebola, Hepatitis-C virus, severe acute respiratory syndrome-associated coronavirus (SARS)-coV, and *Mycobacterium tuberculosis* [12, 13].

M. tuberculosis, the causal agent of tuberculosis in humans, remains a leading cause of morbidity and mortality worldwide [14]. Interactions between the tubercle bacillus and host phagocytes are crucial for immunity to mycobacteria and for TB pathogenesis [15]. *M. tuberculosis* can interact with PRRs involved in signal transduction leading to the secretion of cytokines and other mediators of the immune response [16]. In this context, DC-SIGN has been shown to be an important *M. tuberculosis* receptor on the surface of human monocyte-derived DCs [17, 18] and, more recently, it has been shown that L-SIGN can also interact with the tubercle bacillus [19]. Both DC-SIGN and L-SIGN bindings to *M. tuberculosis* are mediated by the mycobacterial cell-wall component mannosylated lipoarabinomannan (ManLAM). In addition, the observation that DC-SIGN may mediate intracellular signaling events leading to cytokine secretion has led some authors to propose that this lectin could be used by pathogens, including *M. tuberculosis*, as a part of an immune evasion strategy to their own advantage [17, 20]. From a genetic perspective, there is increasing evidence that host genetic factors determine differences in host susceptibility to mycobacterial infection and might contribute to the pattern of clinical disease [21, 22]. In this context, we have recently shown that the combination of two DC-SIGN promoter variants (−871G and −336A) is associated with a decreased risk of developing tuberculosis in a South African cohort [23].

However, the extent to which the length of the neck region of both DC-SIGN and L-SIGN might have an impact on the host susceptibility to TB is unclear at present. This tandem-repeat region, which shows a varying degree of length polymorphism [9, 24], is involved in assembling both lectins into a tetrameric protein conformation on the cell surface, and the length of this region can critically influence the pathogen-binding properties of the CRD of these proteins [25–27]. At the population level, the length of the DC-SIGN neck region is highly conserved (mainly 7 repeats), whereas the L-SIGN neck region exhibits an extraordinarily high level of heterozygosity [28]. Furthermore, several studies suggest that the number of DC-SIGN and/or L-SIGN repeat units can contribute to the risk of HIV-1 [29, 30] and SARS infections [31], as well as to HCV replication efficacy [32].

In light of the ability of both DC-SIGN and L-SIGN to bind *M. tuberculosis*, the fact that neck-region length variation may determine the ligand-binding capacities of these lectins and the observation that variation in these regions is associated with a number of infectious diseases, we hypothesized that length variation in the DC-SIGN and L-SIGN neck regions might affect individual susceptibility to TB. To test this hypothesis, we explored the relationship between the DC-SIGN and L-SIGN tandem repeat variation in the neck region and susceptibility to TB in a large cohort of South African Coloured origin.

PATIENTS AND METHODS

Study Cohort

The study was conducted in a cohort of 711 individuals, including 351 TB patients and 360 healthy controls, living in the Cape Town area. Our study population comes from two suburbs of Cape Town that have been extensively studied because of their uniform ethnicity (South African Coloured) and socioeconomic status as well as a high incidence of TB and a low prevalence of HIV [33]. The annual risk of infection (ARI) in these suburbs was estimated at 2.5% in 1987 and at 2.8–3.5% in 1999, and it is therefore highly likely that, in such an environment, the vast majority of controls have been exposed to *M. tuberculosis* [34, 35]. TB patients were bacteriologically-confirmed (smear-positive and/or culture-positive) to present pulmonary tuberculosis (PTB). Their mean age (±standard deviation) was 36.7 (±10.9), and 51.8% were male. Controls were unrelated healthy individuals, from the same community, with the same socioeconomic status, access to health facilities, and chance of diagnosis, with neither signs nor previous history of TB (mean age 34.6 [±12.5], 22% male). All subjects were HIV-negative and older than 18 years. Informed consent was obtained from all participants, and the study was approved by the ethics committee of the Faculty of Health Sciences, Stellenbosch University, South Africa.

Molecular Analyses of the DC-SIGN and L-SIGN Neck-Region Length Polymorphisms

The DC-SIGN and L-SIGN repeat regions in exon 4 were polymerase chain reaction (PCR) amplified from genomic DNA using the following primers: 5′-AGG

CTTGGCACACAGTAGGTG-3′ and 5′-CAACGA CCATCTCAGGCCCAAGA-3′ for DC-SIGN, and 5′-AGGGCTTGGCACACAGTAGGTG-3′ and 5′-ACC CTTGATGTGCAGGAACT-3′ for L-SIGN. PCR amplifications were performed in a final volume of 25 μl using 20 ng of genomic DNA, 0.0016 μg/μl of each primer, 200 μM of dNTP, 1.5 mM of MgCl2, and 0.5 U of BioTaq (Bioline, Randolph, MA, USA). Cycling conditions were as follows: 5 minutes at 94°C, followed by 35cycles of 30 seconds at 94°C, 30 seconds at 60°C, and 1 minute at 72°C. Alleles were distinguished by fragment length after agarose gel electrophoresis and ethidium bromide staining. The difference among alleles is the multiple of 69 nucleotides, which represents the length of each repeat. Because of the high sequence identity (73%) of DC-SIGN and L-SIGN, special care was taken to design primers that specifically amplified the neck region of both genes. Some representative alleles (bands) of both genes were confirmed by direct sequencing to ensure specific amplifications of DC-SIGN and L-SIGN neck-region alleles.

Statistical Analyses

Allele and genotype frequencies were obtained by direct counting. Differences between cases and controls for both allele and genotype frequencies were determined using a two-sided χ^2 test, and a Fisher's exact test when appropriate. Odds ratio (OR) was calculated with 95% confidence intervals (CI). All analyses were performed using STATA 8.2

RESULTS AND DISCUSSION

The allelic and genotype frequencies of the DC-SIGN and L-SIGN neck-region tandem repeats in the 351 TB patients and 360 healthy controls are summarized in Tables 1 and 2, respectively. Both patients and controls, which are ethnically matched, belong to the South African Coloured population. It is worth mentioning that population stratification between the two study groups, a situation that can lead to spurious associations, was excluded in a previous study by analyzing the entire cohort for a panel of 25 independent genomewide single nucleotide polymorphism (SNP) markers [23]. When examining the allelic frequencies of repeat units for DC-SIGN and L-SIGN neck regions, no statistical differences were observed between TB cases and healthy controls (Table 1).

In the case of DC-SIGN, the 7-repeat allele was by far the most frequently observed, with a frequency of more than 98%. As to L-SIGN, the 7-repeat and the 6-repeat alleles account for more than 80% of the overall diversity. We next examined whether the frequency distributions of DC-SIGN and L-SIGN neck-region genotypes were significantly distorted between cases and controls. Again, no significant differences were detected between diseased individuals and healthy controls for both DC-SIGN and L-SIGN. In regard to DC-SIGN, low genotypic variation was observed in accordance with the allelic data. The 7/7 genotype accounted for nearly all genetic variation (more than 96%), and genotypes 7/4, 7/5, 7/6, and 7/8 were observed at very low frequencies (Table 2). For L-SIGN, the genotypes 7/7 and 7/6 were present at similar frequencies (30–36%), followed by 7/5 (~12%), 6/6 (~6%), 6/5 (~5%), and 9/7 (~4%).

The results of the present study indicate that the number of repeats of the DC-SIGN and L-SIGN neck regions does not seem to influence the host susceptibility to develop TB. In the case of DC-SIGN, our results are in agreement with a recent case-control study in a cohort of northwestern Colombian origin [36]. In this report, the authors analyzed DC-SIGN neck-region variation in a cohort of 110 tuberculosis patients and 299 matched controls, and observed no statistical differences between the two study groups. Thus, both studies support the notion that length variation of the DC-SIGN neck region does not influence the host susceptibility to develop TB.

TABLE 1 DC-SIGN and L-SIGN neck-region allelic frequencies (in %) among patients with tuberculosis and healthy controls

	DC-SIGN				*L-SIGN*			
Alleles	Patients (n^a = 702)	Controls (n^a = 720)	*P*	OR (95% CI)	Patients (n^a = 702)	Controls (n^a = 720)	*P*	OR (95% CI)
4	0.14	0	0.49	nc	0.43	0.69	0.50	0.61 (0.15–2.58)
5	0.28	0	0.24	nc	9.40	11.81	0.14	0.78 (0.55–1.09)
6	1.28	0.97	0.58	1.32 (0.49–3.57)	28.49	24.17	0.06	1.25 (0.99–1.59)
7	98.15	98.89	0.25	0.60 (0.25–1.45)	58.40	59.72	0.61	0.95 (0.77–1.17)
8	0.14	0.14	1.00	1.02 (0.06–16.43)	0.43	0.97	0.22	0.44 (0.11–1.70)
9	0	0	—	nc	2.85	2.64	0.81	1.08 (0.57–2.05)

Abbreviation: nc, not computable.

[a] Number of chromosomes analyzed.

TABLE 2 DC-SIGN and L-SIGN neck-region genotype frequencies (in %) among patients with tuberculosis and healthy controls

	DC-SIGN				*L-SIGN*			
Genotypes	Patients (n^a = 351)	Controls (n^a = 360)	*P*	OR (95% CI)	Patients (n^a = 351)	Controls (n^a = 360)	*P*	OR (95% CI)
4\4	0	0	—	nc	0	0.28	1.00	nc
5\4	0	0	—	nc	0.28	0	1.00	nc
5\5	0	0	—	nc	0.57	2.50	0.06	0.22 (0.04–1.04)
6\4	0	0	—	nc	0.28	0.56	1.00	0.51 (0.05–5.67)
6\5	0	0	—	nc	5.13	5.28	0.93	0.97 (0.50–1.88)
6\6	0	0	—	nc	7.12	5.83	0.48	1.24 (0.68–2.26)
7\4	0.28	0	1.00	nc	0.28	0.28	1.00	1.03 (0.06–16.46)
7\5	0.57	0	0.24	nc	12.25	13.33	0.67	0.91 (0.58–1.41)
7\6	2.56	1.94	0.58	1.33 (0.49–3.60)	36.18	29.72	0.07	1.34 (0.98–1.83)
7\7	96.30	97.78	0.24	0.59 (0.24–1.44)	31.34	35.28	0.27	0.84 (0.61–1.14)
8\6	0	0	—	nc	0	0.28	1.00	nc
8\7	0.28	0.28	1.00	1.03 (0.06–16.46)	0.85	1.67	0.51	0.51 (0.13–2.05)
9\6	0	0	—	nc	1.14	0.83	0.72	1.37 (0.30–6.17)
9\7	0	0	—	nc	4.56	3.89	0.66	1.18 (0.57–2.46)
9\9	0	0	—	nc	0	0.28	1.00	nc

Abbreviation: nc, not computable.

[a] Number of individuals.

In the context of other infectious diseases, the only positive association published so far between DC-SIGN neck-region variation and susceptibility to infectious disease is restricted to HIV-1 infection [29]. In this study, the authors observed an excess of heterozygous individuals for DC-SIGN tandem-repeats in a group of repeatedly-exposed seronegative individuals as compared to the groups of HIV-1 seronegative and HIV-1 seropositive individuals. These observations were interpreted as heterozygosity in the DC-SIGN neck region being associated with reduced susceptibility to HIV-1 infection [29]. At the level of the general population, it is of interest that the DC-SIGN neck region exhibits very low levels of polymorphism [9, 28]. Indeed, we recently screened the entire Human Genome Diversity Panel (HGDP-CEPH panel), which is composed of more than 1,000 control individuals from 52 different ethnic groups, for repeat variation in the neck regions of both DC-SIGN and L-SIGN [28]. For DC-SIGN, we observed that the 7-repeat allele accounts for nearly all genetic variation (~99%), and that the other alleles, which range from 2–10 repeats, are present at very low frequencies. In addition, the levels of sequence variation in the entire DC-SIGN coding-region, particularly of those that affect amino-acid identity, were found to be extremely low [28]. The low levels of genetic variation observed in the DC-SIGN coding region are also reflected in the context of disease association studies. Indeed, the different associations published so far between DC-SIGN genetic variation and susceptibility to infectious diseases always involve polymorphisms in the DC-SIGN promoter region, and not in its coding region [23, 37, 38]. For example, we have previously shown that the combination of two DC-SIGN promoter variants (−871G and −336A) is associated with a reduced risk of developing TB in the same South African cohort analyzed here [23]. Furthermore, the genetic variation in DC-SIGN that has been associated with protection against parenteral HIV-1 infection [37] and with the severity of dengue pathogenesis [38] also involves polymorphisms (i.e. −336A/G) restricted to the DC-SIGN promoter region. Taken together, all these studies support the view that it is the variation in the amount of DC-SIGN protein being produced that can influence infectious disease susceptibility, and not differences in the DC-SIGN protein itself or variation in its neck region.

Our study presents the first investigation of the role of L-SIGN neck-region variation in susceptibility to TB. A number of studies have already explored possible correlations between L-SIGN neck-region variation and susceptibility to other infectious diseases [30–32, 39]. For example, the L-SIGN tandem-repeat 7/5 genotype has been recently associated with an increased protection against HIV-1 infection in high risk individuals [30]. However, this association remains controversial because a previous study failed to detect such an association [39]. In the context of Hepatitis C virus (HCV) infection, a study comparing the frequency distribution of L-SIGN neck-region polymorphisms in a group of infected patients with noninfected individuals failed to demonstrate any statistical difference between the two study groups [32]. However, the same authors did observe an association between neck-region polymor-

phisms and individual HCV viral loads, and suggested that length variation in the L-SIGN neck region affects HCV replication efficacy. Finally, a recent study focusing on susceptibility to SARS infection has shown that individuals who are homozygous for L-SIGN neck-region repeats are better protected against SARS infection [31]. However, our results clearly indicate that the L-SIGN neck-region allele/genotype frequencies are not statistically different between TB patients and healthy controls in our large South African cohort (Tables 1 and 2). Thus, our data seem to exclude length variation in the L-SIGN neck region as a factor influencing TB susceptibility.

More generally, our study clearly illustrates the advantages of using admixed populations in the context of disease association studies. Indeed, the South African Coloured population represents a present-day homogenous population who originated from the variable admixture of different populations, such as African Khoisan and Bantu-speakers, Malaysians, Indians, and Europeans [40, 41]. Consequently, the South African Coloured population presents a large degree of genetic diversity, resulting in a high number of alleles or genotype combinations that can be used in association studies. For example, the L-SIGN genotypes 6/5, 7/5, and 9/7 are observed at relatively high frequencies among the South Africans (providing evidence for the genetic input received from European and Asian populations), whereas they are rare or even absent in other subSaharan African populations [28]. The presence of these genotypes in the South African population offers a unique opportunity for testing their association with disease in a single population, a hypothesis that would be difficult to test in other African populations because these genotypes are found at a very low frequency, or are even absent.

In summary, our results show that the length of the neck regions of both DC-SIGN and L-SIGN are not associated with an increased or decreased host susceptibility to develop TB, at least in our South African cohort. These data are in contrast with other disease association studies where the tandem-repeat polymorphisms of DC-SIGN and/or L-SIGN seem to contribute to different susceptibilities to HIV-1 and SARS infections, and to the HCV replication efficacy [29–32].

ACKNOWLEDGMENTS

Luis B. Barreiro was supported by a "Fundação para a Ciência e a Tecnologia" fellowship (SFRH/BD/18580/2004). This research project has been cofinanced by the European Commission, within the 6th Framework Programme (contract no. LSHP-CT-2003–503367), and Institut Pasteur Research funding. We thank M. Kennedy for help with sample collection, the South African Medical Research Council and Department of Science and Technology/National Research Foundation (DST/NRF) Centre of Excellence for Biomedical TB Research for financial assistance, and Bernard Lakowsky for critical reading of the manuscript.

REFERENCES

1. Janeway CA, Medzhitov R: Innate immune recognition. Annu Rev Immunol 20:197, 2002.
2. Curtis BM, Scharnowske S, Watson AJ: Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. Proc Natl Acad Sci USA 89:8356, 1992.
3. Geijtenbeek TB, Torensma R, van Vliet SJ, van Duijnhoven GC, Adema GJ, van Kooyk Y, Figdor CG: Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. Cell 100:575, 2000.
4. Geijtenbeek TB, van Vliet SJ, Engering A, t Hart BA, van Kooyk Y: Self- and nonself-recognition by C-type lectins on dendritic cells. Annu Rev Immunol 22:33, 2004.
5. Soilleux EJ, Barten R, Trowsdale J: DC-SIGN; a related gene, DC-SIGNR; and CD23 form a cluster on 19p13. J Immunol 165:2937, 2000.
6. Pohlmann S, Soilleux EJ, Baribaud F, Leslie GJ, Morris LS, Trowsdale J, Lee B, Coleman N, Doms RW: DC-SIGNR, a DC-SIGN homologue expressed in endothelial cells, binds to human and simian immunodeficiency viruses and activates infection in trans. Proc Natl Acad Sci USA 98:2670, 2001.
7. Bashirova AA, Wu L, Cheng J, Martin TD, Martin MP, Benveniste RE, Lifson JD, KewalRamani VN, Hughes A, Carrington M: Novel member of the CD209 (DC-SIGN) gene family in primates. J Virol 77:217, 2003.
8. Soilleux EJ: DC-SIGN (dendritic cell-specific ICAM-grabbing non-integrin) and DC-SIGN-related (DC-SIGNR): friend or foe? Clin Sci (Lond) 104:437, 2003.
9. Bashirova AA, Geijtenbeek TB, van Duijnhoven GC, van Vliet SJ, Eilering JB, Martin MP, Wu L, Martin TD, Viebig N, Knolle PA, KewalRamani VN, van Kooyk Y, Carrington M: A dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN)-related protein is highly expressed on human liver sinusoidal endothelial cells and promotes HIV-1 infection. J Exp Med 193:671, 2001.
10. Soilleux EJ, Morris LS, Lee B, Pohlmann S, Trowsdale J, Doms RW, Coleman N: Placental expression of DC-SIGN may mediate intrauterine vertical transmission of HIV. J Pathol 195:586, 2001.
11. Soilleux EJ, Morris LS, Leslie G, Chehimi J, Luo Q, Levroney E, Trowsdale J, Montaner LJ, Doms RW, Weissman D, Coleman N, Lee B: Constitutive and induced expression of DC-SIGN on dendritic cell and macrophage subpopulations in situ and in vitro. J Leukoc Biol 71:445, 2002.

12. Koppel EA, Saeland E, de Cooker DJ, van Kooyk Y, Geijtenbeek TB: DC-SIGN specifically recognizes Streptococcus pneumoniae serotypes 3 and 14. Immunobiology 210: 203, 2005.
13. van Kooyk Y, Geijtenbeek TB: DC-SIGN: escape mechanism for pathogens. Nat Rev Immunol 3:697, 2003.
14. Frieden TR, Sterling TR, Munsiff SS, Watt CJ, Dye C: Tuberculosis. Lancet 362:887, 2003.
15. Kaufmann SH: How can immunology contribute to the control of tuberculosis? Nat Rev Immunol 1:20, 2001.
16. Quesniaux V, Fremond C, Jacobs M, Parida S, Nicolle D, Yeremeev V, Bihl F, Erard F, Botha T, Drennan M, Soler MN, Le Bert M, Schnyder B, Ryffel B: Toll-like receptor pathways in the immune responses to mycobacteria. Microbes Infect 6:946, 2004.
17. Geijtenbeek TB, Van Vliet SJ, Koppel EA, Sanchez-Hernandez M, Vandenbroucke-Grauls CM, Appelmelk B, Van Kooyk Y: Mycobacteria target DC-SIGN to suppress dendritic cell function. J Exp Med 197:7, 2003.
18. Tailleux L, Schwartz O, Herrmann JL, Pivert E, Jackson M, Amara A, Legres L, Dreher D, Nicod LP, Gluckman JC, Lagrange PH, Gicquel B, Neyrolles O: DC-SIGN is the major Mycobacterium tuberculosis receptor on human dendritic cells. J Exp Med 197:121, 2003.
19. Koppel EA, Ludwig IS, Hernandez MS, Lowary TL, Gadikota RR, Tuzikov AB, Vandenbroucke-Grauls CM, van Kooyk Y, Appelmelk BJ, Geijtenbeek TB: Identification of the mycobacterial carbohydrate structure that binds the C-type lectins DC-SIGN, L-SIGN and SIGNR1. Immunobiology 209:117, 2004.
20. Tailleux L, Maeda N, Nigou J, Gicquel B, Neyrolles O: How is the phagocyte lectin keyboard played? Master class lesson by Mycobacterium tuberculosis. Trends Microbiol 11:259, 2003.
21. Cooke GS, Hill AV: Genetics of susceptibility to human infectious disease. Nat Rev Genet 2:967, 2001.
22. Casanova JL, Abel L: Genetic dissection of immunity to mycobacteria: the human model. Annu Rev Immunol 20:581, 2002.
23. Barreiro LB, Neyrolles O, Babb CL, Tailleux L, Quach H, McElreavey K, Helden PD, Hoal EG, Gicquel B, Quintana-Murci L: Promoter variation in the DC-SIGN-encoding gene CD209 is associated with tuberculosis. PLoS Med 3:e20, 2006.
24. Mummidi S, Catano G, Lam L, Hoefle A, Telles V, Begum K, Jimenez F, Ahuja SS, Ahuja SK: Extensive repertoire of membrane-bound and soluble dendritic cell-specific ICAM-3-grabbing nonintegrin 1 (DC-SIGN1) and DC-SIGN2 isoforms. Inter-individual variation in expression of DC-SIGN transcripts. J Biol Chem 276: 33196, 2001.
25. Feinberg H, Guo Y, Mitchell DA, Drickamer K, Weis WI: Extended Neck Regions Stabilize Tetramers of the Receptors DC-SIGN and DC-SIGNR. J Biol Chem 280: 1327, 2005.
26. Snyder GA, Colonna M, Sun PD: The structure of DC-SIGNR with a portion of its repeat domain lends insights to modeling of the receptor tetramer. J Mol Biol 347:979, 2005.
27. Guo Y, Atkinson CE, Taylor ME, Drickamer K: All but the shortest polymorphic forms of the viral receptor DC-SIGNR assemble into stable homo- and heterotetramers. J Biol Chem 281:16794, 2006.
28. Barreiro LB, Patin E, Neyrolles O, Cann HM, Gicquel B, Quintana-Murci L: The heritage of pathogen pressures and ancient demography in the human innate-immunity CD209/CD209L region. Am J Hum Genet 77:869, 2005.
29. Liu H, Hwangbo Y, Holte S, Lee J, Wang C, Kaupp N, Zhu H, Celum C, Corey L, McElrath MJ, Zhu T: Analysis of genetic polymorphisms in CCR5, CCR2, stromal cell-derived factor-1, RANTES, and dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin in seronegative individuals repeatedly exposed to HIV-1. J Infect Dis 190:1055, 2004.
30. Liu H, Carrington M, Wang C, Holte S, Lee J, Greene B, Hladik F, Koelle DM, Wald A, Kurosawa K, Rinaldo CR, Celum C, Detels R, Corey L, McElrath MJ, Zhu T: Repeat-region polymorphisms in the gene for the dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin-related molecule: effects on HIV-1 susceptibility. J Infect Dis 193:698, 2006.
31. Chan VS, Chan KY, Chen Y, Poon LL, Cheung AN, Zheng B, Chan KH, Mak W, Ngan HY, Xu X, Screaton G, Tam PK, Austyn JM, Chan LC, Yip SP, Peiris M, Khoo US, Lin CL: Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. Nat Genet 38:38, 2006.
32. Nattermann J, Ahlenstiel G, Berg T, Feldmann G, Nischalke HD, Muller T, Rockstroh J, Woitas R, Sauerbruch T, Spengler U: The tandem-repeat polymorphism of the DC-SIGNR gene in HCV infection. J Viral Hepat 13:42, 2006.
33. Beyers N, Gie RP, Zietsman HL, Kunneke M, Hauman J, Tatley M, Donald PR: The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high-incidence community. S Afr Med J 86:40, 1996.
34. Munch Z, Van Lill SW, Booysen CN, Zietsman HL, Enarson DA, Beyers N: Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. Int J Tuberc Lung Dis 7:271, 2003.
35. Beyers N, Michaelis I, Gie R, Schaaf S, Richardson M, Warren R, Fourie B, van Helden P: Transmission of tuberculosis (TB) to children in a high incidence area. Int J Tuberc Lung Dis 5:S185, 2001.
36. Gómez LM, Anaya JM, Sierra-Filardi E, Cadena J, Corbí A, Martín J: Analysis of DC-SIGN (*CD209*) Functional Variants in Patients with Tuberculosis. Hum Immunol 67:808, 2006, doi: 10.1016/j.humimm.2006.07.003.

37. Martin MP, Lederman MM, Hutcheson HB, Goedert JJ, Nelson GW, van Kooyk Y, Detels R, Buchbinder S, Hoots K, Vlahov D, O'Brien SJ, Carrington M: Association of DC-SIGN promoter polymorphism with increased risk for parenteral, but not mucosal, acquisition of human immunodeficiency virus type 1 infection. J Virol 78: 14053, 2004.

38. Sakuntabhai A, Turbpaiboon C, Casademont I, Chuansumrit A, Lowhnoo T, Kajaste-Rudnitski A, Kalayanarooj SM, Tangnararatchakit K, Tangthawornchaikul N, Vasanawathana S, Chaiyaratana W, Yenchitsomanus PT, Suriyaphol P, Avirutnan P, Chokephaibulkit K, Matsuda F, Yoksan S, Jacob Y, Lathrop GM, Malasit P, Despres P, Julier C: A variant in the CD209 promoter is associated with severity of dengue disease. Nat Genet 37:507, 2005.

39. Lichterfeld M, Nischalke HD, van Lunzen J, Sohne J, Schmeisser N, Woitas R, Sauerbruch T, Rockstroh JK, Spengler U: The tandem-repeat polymorphism of the DC-SIGNR gene does not affect the susceptibility to HIV infection and the progression to AIDS. Clin Immunol 107:55, 2003.

40. van der Ross RE: 100 questions about Coloured South Africans. Cape Town (South Africa). UWC Printing Department, 1993.

41. Nurse GT, Weiner JS, Jenkins T: The peoples of Southern Africa and their affinities. Oxford, Clarendon Press, 1985.

Hum Genet
DOI 10.1007/s00439-007-0335-1

SHORT REPORT

SP110 polymorphisms are not associated with pulmonary tuberculosis in a South African population

C. Babb · E. Hanekom Keet · P. D. van Helden · E. G. Hoal

Received: 14 December 2006 / Accepted: 16 January 2007

Abstract Susceptibility to tuberculosis (TB) in mice has recently been attributed to the *Ipr1* gene. Polymorphisms in the human homologue, SP110, have been investigated in various populations with only one study finding an association with TB susceptibility. We investigated eight SP110 polymorphisms in a South African population, including two novel polymorphisms. No significant association was found with any of the polymorphisms investigated, including two polymorphisms that were previously found to be associated with TB susceptibility in West African populations.

The identification of genes that may confer resistance or susceptibility to tuberculosis is frequently done by hypothesis. However, linkage analysis or animal models may also provide clues leading to genes which can be tested in a case-control approach. The use of murine models has led to the identification of genes such as SLC11A1, where association with human TB has been confirmed in a number of studies (Hoal et al. 2004). It was recently reported that the intracellular pathogen resistance 1 (*Ipr1*) gene, within the "super susceptibility to TB1" (*sst1*) locus, conferred a measure of innate immunity to *Mycobacterium tuberculosis* infection in mice (Pan et al. 2005). The human homologue is SP110 which is believed to be involved in regulation of transcription as it contains a conserved SAND domain which is thought to facilitate DNA-binding in transcription regulation. An association study of the SP110 gene found that there were three polymorphisms that possibly influenced genetic susceptibility to TB in the West African populations of the Gambia, Guinea Bissau and the Republic of Guinea (Tosh et al. 2006). We tested polymorphisms in this gene in a South African population, known as South African Coloured (Lombard et al. 2006). This population set has previously been tested via 25 unlinked SNPs, and found not to be stratified (Barreiro et al. 2006). While our investigation was underway, two large studies in a Ghanaian population (Thye et al. 2006) and more recently a Russian cohort (Szeszko et al. 2006) found no association with SP110 polymorphisms.

Study participants were recruited primarily from the suburbs of Ravensmead and Uitsig in the Western Cape, South Africa. Cases had a microbiologically confirmed diagnosis of pulmonary TB, and were HIV negative (mean age 37, 46% female). Controls were from the same ethnic and socioeconomic group and were healthy, with no history of TB (mean age 32, 85% female). All cases and controls were older than 17 years.

We sequenced exons 4–8, 11, 14, 15, alternative 15, 17 and 18 including surrounding intronic regions in ten individuals and chose eight SNPs to be genotyped by SNaPshot in 381 TB cases and 417 healthy controls (Table 1). SNaPshot is a methodology involving a primer extension genotyping technique that makes use of allele-specific primers and can be performed in a multiplex. Primer sequences are available from the authors. The SNPs included two novel intronic polymorphisms (submitted to NCBI dbSNP November 2006) and two polymorphisms (rs2114592 and rs3948464) that were previously found to be associated with susceptibility to TB (Tosh et al. 2006). A third

C. Babb · E. H. Keet · P. D. van Helden · E. G. Hoal (✉)
Molecular Biology and Human Genetics,
DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, MRC Centre for Molecular and Cellular Biology,
Faculty of Health Sciences, Stellenbosch University,
P.O. Box 19063, Tygerberg, 7505 Cape Town, South Africa
e-mail: egvh@sun.ac.za

Table 1 SP110 genotype distributions in cases with tuberculosis and healthy controls in a South African population

	SNP		Genotype frequencies (%)						Cases versus controls[a]		
			Cases			Controls					
	1	2	11	12	22	11	12	22	P	OR (95% CI)	
rs2114592	C	T	0.526	0.379	0.095	0.533	0.372	0.095	0.886	0.971	(0.733–1.288)
rs1365776	A	G	0.710	0.253	0.037	0.701	0.262	0.037	0.813	1.046	(0.768–1.425)
SP110ins/del	GAAG	–	0.816	0.184	0.000	0.842	0.158	0.000	0.348	0.835	(0.577–1.209)
rs3948464	C	T	0.551	0.373	0.077	0.543	0.374	0.083	0.828	1.033	(0.776–1.374)
rs1135791	T	C	0.500	0.407	0.093	0.490	0.407	0.103	0.825	1.041	(0.780–1.389)
rs3948463	G	A	0.950	0.050	0.000	0.976	0.024	0.000	0.060	0.472	(0.217–1.028)
SP110int48923	C	T	0.869	0.123	0.008	0.886	0.112	0.002	0.516	0.855	(0.559–1.308)
rs35495464	T	C	0.947	0.053	0.000	0.969	0.031	0.000	0.583	0.156	(0.217–1.028)

[a] The homozygotes for the most frequent allele were compared with the sum of the heterozygote and homozygote for the rare allele. Comparing the homozygotes for the rare allele with the sum of the heterozygotes and homozygotes for the most frequent allele gave similar results. OR Odds Ratio CI Confidence Interval

polymorphism, sp110int10, is in linkage disequilibrium (LD) with rs3948464, and therefore does not contribute additional information (Tosh et al. 2006). The sp110int10 polymorphism was not found in the South African population after sequencing 30 samples. In the South African population no significant association with TB susceptibility was found for any of the eight polymorphisms investigated (Table 1). Analysis of the haplotypes by Haploview also revealed no association with any of the inferred haplotypes. All polymorphisms were in Hardy Weinberg Equilibrium. This study had 95% confidence, 80% power to detect an odds ratio of 1.6 for both the previously implicated SNPs.

Replication of the association with TB susceptibility reported in the West African populations was not detected. Since a number of independent populations and SP110 variations have now been investigated, with only one study indicating an association with TB, it is likely that this gene does not contribute to genetic susceptibility to TB in adult humans.

References

Barreiro LB, Neyrolles O, Babb CL, Tailleux L, Quach H, McElreavey K, Helden PD, Hoal EG, Gicquel B, Quintana-Murci L (2006) Promoter variation in the DC-SIGN encoding gene CD209 is associated with tuberculosis. PLoS Med 3:e20

Hoal EG, Lewis L-A, Jamieson SE, Tanzer F, Rossouw M, Victor T, Hillerman R, Beyers N, Blackwell JM, van Helden PD (2004) *SLC11A1 (NRAMP1)* but not *SLC11A2 (NRAMP2)* polymorphisms are associated with susceptibility to tuberculosis in a high-incidence community in South Africa. Int J Tuberc Lung Dis 8:1464–1471

Lombard Z, Brune AE, Hoal EG, Babb C, van Helden PD, Eplen JT, Bornman L (2006) HLA class II disease associations in southern Africa. Tissue Antigens 67:97–110

Pan H, Yan BS, Rojas M, Shebzukhov YV, Zhou H, Kobzik L, Higgins DE, Daly MJ, Bloom BR, Kramnik I (2005) Ipr1 gene mediates innate immunity to tuberculosis. Nature 434:767–772

Szeszko JS, Healy B, Stevens H, Balabanova Y, Drobniewski F, Todd JA, Nejentsev S (2006) Resequencing and association analysis of the SP110 gene in adult pulmonary tuberculosis. Hum Genet DOI 10.1007/s00439-006-0293-z

Thye T, Browne EN, Chinbuah MA, Gyapong J, Osei I, Owusu-Dabo E, Niemann S, Rusch-Gerdes S, Horstmann RD, Meyer CG (2006) No associations of human pulmonary tuberculosis with Sp110 variants. J Med Genet 43:e32

Tosh K, Campbell SJ, Fielding K, Sillah J, Bah B, Gustafson P, Manneh K, Lisse I, Sirugo G, Bennett S, Aaby P, McAdam KP, Bah-Sow O, Lienhardt C, Kramnik I, Hill AV (2006) Variants in the SP110 gene are associated with genetic susceptibility to tuberculosis in West Africa. Proc Natl Acad Sci USA 103:10364–10368

Tuberculosis (2007) 87, 295–302

Available at www.sciencedirect.com

ScienceDirect

journal homepage: http://intl.elsevierhealth.com/journals/tube

Vitamin D receptor gene polymorphisms and sputum conversion time in pulmonary tuberculosis patients

C. Babb[a], L. van der Merwe[b], N. Beyers[c], C. Pheiffer[a], G. Walzl[a], K. Duncan[d], P. van Helden[a], E.G. Hoal[a,*]

[a]*Molecular Biology and Human Genetics, DST/NRF Centre of Excellence in Biomedical Tuberculosis Research and MRC Centre for Molecular and Cellular Biology, Faculty of Health Sciences, Stellenbosch University, PO Box 19063, Tygerberg, 7505, South Africa*
[b]*Biostatistics Unit, Medical Research Council, Tygerberg, South Africa*
[c]*Desmond Tutu TB Centre, Department of Paediatrics and Child Health, Faculty of Health Sciences, Stellenbosch University, South Africa*
[d]*GlaxoSmithKline, Stevenage, UK*

Received 26 October 2006; received in revised form 22 February 2007; accepted 3 March 2007

KEYWORDS
Susceptibility;
Biomarkers;
Surrogate marker;
Extent of TB;
Chest radiograph;
VDR

Summary
A cohort of pulmonary tuberculosis (TB) patients in a South African admixed population was investigated to determine if the vitamin D receptor gene (VDR) polymorphisms *FokI*, *ApaI*, and *TaqI* are associated with TB susceptibility or time to sputum conversion, and to investigate other clinical and demographic factors affecting the rate of response to treatment. Firstly, a case-control association study of 249 TB cases and 352 healthy controls was carried out to investigate association of VDR polymorphisms with TB susceptibility. Secondly, a cohort of pulmonary tuberculosis patients with conversion times for both sputum smear ($n = 220$) and culture ($n = 222$) were analysed to determine factors contributing to mycobacterial resolution in sputum. Age and gender adjusted Cox regression models were constructed. Our results indicate that the extent of disease at diagnosis was predictive of both smear and culture conversion times in the final models. Smoking status and VDR genotype contributed independently to smear conversion time, with *ApaI* 'AA' genotype and *TaqI* 'T'-containing genotypes predictive of a faster response to TB chemotherapy. We did not find an association between VDR genotype and TB in the case-control study. We conclude that the time taken for an individual to convert to sputum negativity while on antituberculosis therapy can be independently predicted by the VDR genotype.

*Corresponding author. Tel.: +27 21 9389412; fax: +27 21 9389476.
E-mail address: egvh@sun.ac.za (E.G. Hoal).

1472-9792/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tube.2007.03.001

Introduction

The Western Cape, South Africa had a reported tuberculosis (TB) incidence of 919/100 000 in 2003,[1] despite a relatively low HIV prevalence.[2] Identification of the genes involved in the response to *Mycobacterium tuberculosis* (*M. tuberculosis*) infection may assist in the development of more effective treatment, and a number of genes have been implicated in susceptibility to TB, including Interferon gamma,[3] Natural Resistance-Associated Macrophage Protein (NRAMP1 or SLC11A1)[4,5] and Vitamin D Receptor.[6–9]

The role of Vitamin D and its receptor (VDR) in innate immunity, and specifically TB, is a subject of intense debate.[7–11] The active metabolite of Vitamin D, 1,25 dihydroxyvitaminD$_3$, suppresses growth of *M. tuberculosis* in vitro[12,13] and this effect may be facilitated by Toll-Like Receptors in vivo.[14] Case-control studies to assess the importance of VDR polymorphisms in TB have produced varying results in different populations.[10,15]

Recently Roth et al. (2004) reported an association of VDR polymorphisms with the time to sputum conversion in 78 TB patients,[16] although no association with susceptibility to TB was found in a case-control analysis. Validation of this preliminary report of TB treatment response being investigated in relation to a genetic component is vital, preferably in a different population.

We analysed three VDR polymorphisms, *FokI*, *ApaI* and *TaqI*, and a number of other clinical and demographic factors in a large, well characterised group of first time pulmonary TB cases from the population group known as South African Colored in the Western Cape, South Africa. Although admixed, this population has been shown in a previous study not to be stratified.[17]

The *FokI* (rs10735810) polymorphism is a functional polymorphism[18] and can alter the amount of VDR produced.[19,20] *ApaI* (rs7975232) in intron VIII was analysed because it appears to form an important haplotype.[8,19] *TaqI* (rs731236) is a silent polymorphism (T/C) located in exon IX, in the 3′ region. Strong linkage between the *ApaI* and *TaqI* is observed, and they are in linkage disequilibrium (LD) with a polyA variable number of tandem repeats which has been shown to affect transcription.[21]

A number of variables have been indicated as potentially being involved in the response of TB patients to treatment, and this requires validation. In this study, we have extended the number of variables and examined their effect on response during TB chemotherapy in a large, well-characterised cohort in a South African population.

Methods

Patients and surveillance

All TB cases were enrolled for a longitudinal study evaluating pulmonary TB cases during and after treatment.[22,23] Ethical approval for this study was obtained from the Institutional Review Board, Faculty of Health Sciences, Stellenbosch University and from the City of Cape Town. Informed consent was obtained from all subjects. Cases were all first time pulmonary TB patients, not non-tuberculosis mycobacteria or MOTTs infected, HIV negative, 18–65 years old, not pregnant, had no chronic diseases, not multi-drug resistant (MDR) TB and received direct observed therapy, short course (DOTS) as prescribed by the South African National Tuberculosis Programme, based on World Health Organisation guidelines. All cases were diagnosed by having a sputum sample positive on auramine staining, a second positive sputum sample or a posterior anterior and lateral chest radiograph typical of pulmonary TB. Blood was collected at diagnosis and total peripheral white blood cell (WBC) counts, and absolute neutrophil, monocyte, and lymphocyte numbers were measured.

Cases were followed up with regular clinic visits from diagnosis (day 0), including day 1 and day 2, then weekly visits till month 2 (day 55) followed by monthly visits till month 6 (day 182) and subsequent visits in month 9 (day 273) and month 12 (day 364). For research purposes, sputum was collected at all time points for smear microscopy using Ziehl–Nielsen (ZN) staining and for *M. tuberculosis* culture using the BACTEC method (Beckton-Dickinson, USA). Restriction fragment length polymorphism (RFLP) of the *M. tuberculosis* strain and isoniazid and rifampicin susceptibility testing was performed. Serum was not available for the majority of samples, therefore the Vitamin D levels were not measured.

A questionnaire was used to collect data on education level (completion of high school, yes/no), smoking of tobacco or cigarettes (yes/no), cannabis use (yes/no), alcohol consumption (yes/no) and limited data on income, measured as either above or below ZAR1000/$\pm$£80/$\pm$\$150 per month. The following data were collected for all cases: age, gender, number of self reported TB symptoms at first clinical examination, weight, height, body mass index, and Mantoux skin test size at diagnosis and month 2. The chest radiographs were interpreted, based on a previous method,[24] by a single clinician who used a standardised reading form to evaluate the number of cavities visible on the chest radiograph, and the extent of pulmonary involvement ($<$right upper lobe (RUL), $=$RUL, $>$RUL and $>$1 lung) as an indication of disease severity.

Case-control

The 249 TB cases were those enrolled in the longitudinal study as first time pulmonary TB patients. The 352 controls were enrolled from a variety of sources such as clinics, households and places of work. They were all from the same high incidence suburb, socio-economic status and population group, and were enrolled to serve as controls in our case-control studies. They had no clinical history or symptoms of TB and were older than 17 years.

Cohort

A cohort from the longitudinal study was selected for conversion time analysis. Inclusion required the case to have a conversion time for smear and/or culture. A conversion time in days (from positive to negative *M. tuberculosis* in sputum) was estimated from the day of diagnosis (day 0) and a date midway between the last positive result and the first of two successive negative results.[16] When more than 91 days had elapsed between the last positive and the first of

two successive negative results, samples were censored and the last positive day was used as the day of conversion. Defaulters (a case who did not take the prescribed medication for at least two consecutive months) and cases that transferred outside of the study area were excluded if there was no conversion time available before this event. This resulted in a total of 220 cases with smear conversion times and 222 cases with culture conversion times.

The cohort was also divided into slow and fast responders for both smear and culture conversion time. Fast responders were defined as cases that converted to negativity before day 55 (month 2) and slow responders on or after day 55. Month 2 was chosen as the cutoff as this corresponds to the International Union Against Tuberculosis and Lung Disease recommendations for the time at which sputum conversion is determined in order to establish treatment effectiveness.

SNP analysis

Genomic DNA was isolated using the Nucleon BACC3 Kit for blood (Amersham Biosciences, UK). Primers used were as in Wilkinson et al.[10] Samples were amplified using Supertherm Gold DNA Polymerase (JMR Holdings, UK) according to the following program: 1 cycle at 94 °C for 15 min followed by 35 cycles at 94 °C for 30 s, 67 °C for 30 s and 72 °C for 45 s, then 1 cycle at 72 °C for 10 min and finally a 15 °C hold step. All fragments were digested overnight at 65, 37 and 55 °C for *TaqI*, *ApaI* and *BseGI*/*FokI* restriction enzymes (Fermentas, Canada), respectively. Resulting fragment sizes were 270 base pairs (bp), 208 and 62 bp for *FokI*; 716, 485 and 231 bp for *ApaI*; and 716, 512, 311, 204 and 201 bp for *TaqI* digestion. A lower case ('f', 'a' or 't') was used to indicate the presence of an endonuclease site. Fragments were separated on a 2% agarose gel.

Statistical analysis

All analyses were done in R: A language and environment for statistical computing, Base R,[25] as well as packages survival,[26] genetics,[27] and haplo.stats.[28] Genotype, allele and diplotype frequencies of the polymorphisms were calculated. As unrelated cases and controls were analysed, haplotypes could only be inferred. However, the diplotype is a true representation of the genotype and was constructed by combining the observed genotypes in the order in which they occur in the gene: *FokI*, *ApaI* and *TaqI*. LD and Hardy–Weinberg equilibrium were determined.

Case-control

A total of 249 cases and 352 controls were analysed to investigate associations with susceptibility to TB using Fisher's exact test. Haplotypes were inferred and analysed with the R package haplo.stats.[29] The diplotypes were constructed manually and their association with TB analysed by logistic regression.

Cohort

Kaplan Meier survival curves were constructed for all available variables on both smear and culture conversion times. The association of each of the variables with conversion time was assessed using log rank tests.

The fast and slow responders to treatment were compared by a Fisher's exact test.

An individual, age and gender adjusted, Cox regression analysis (univariate) was done on all available variables for both smear and culture conversion times.

A multivariate analysis was done by means of the selection of an optimal Cox model for each of the methods (smear and culture) for measuring conversion times. A set of candidate variables were identified from which the optimal selection was made. They included those that have previously been found to be important for conversion in TB by Roth et al.[16] i.e. high school completion, *TaqI* and *FokI*. MDR TB, a previously noted important predictor, was one of the exclusion criteria for our cases. Isoniazid mono-resistance was investigated instead, as no rifampicin mono-resistance was seen in this cohort. We also added variables that were significantly associated ($p<0.050$) with conversion time in smear or culture according to the age and gender adjusted individual Cox analyses. A basic set of variables consisting of age, gender and the genotypes were included in the model irrespective of their effect. The variables included in the final models were selected with a backwards procedure.

Results

Case-control study

In the controls, there was strong linkage disequilibrium between *ApaI* and *TaqI* ($D' = 0.999$). *FokI* was not linked with either *TaqI* or *ApaI*, ($D' = 0.048$ and 0.005, respectively). All polymorphisms were in Hardy Weinberg equilibrium.

No association was found between genotype or allele counts for the individual VDR polymorphisms and the presence or absence of pulmonary TB (Table 1). A non-significant association was seen between the inferred *FokI-ApaI-TaqI* haplotype and TB (global p-value $= 0.078$), with the 'FaT' haplotype overrepresented in controls ($p = 0.063$) and the 'FAT' haplotype more frequent in cases ($p = 0.062$) (Table 2). Of the 27 possible diplotypes only 18 were observed. There were no significant differences between the diplotype frequencies in cases and controls after adjusting for age and gender using logistic regression ($p = 0.24$) (Table 3).

Cohort and conversion times

Fast/slow analysis: When the cases were divided into either fast ($n = 160$) or slow ($n = 60$) responders according to smear conversion before/after day 55, i.e. month 2 (Table 4), there was a strong trend to slower conversion in the 'F'-containing genotypes of *FokI* ($p = 0.055$). Dividing cases into fast ($n = 96$) and slow ($n = 126$) responders with respect to culture conversion showed a significant association with the *ApaI* genotypes ($p = 0.029$). These results were not adjusted for multiple comparisons. Fewer fast responders (10%) had an *ApaI* 'aa' genotype than slow responders (23%). None of the diplotypes were significantly associated with a fast or slow response.

Log rank analysis: In the log rank analyses there were significant associations with faster smear conversion time and no smoking, low number of cavities at diagnosis, low extent of TB at diagnosis, low absolute neutrophil numbers and completion of high school. Non-smokers ($n = 12$) converted sooner than smokers, but this association was seen only with smear conversion time and not with culture conversion time. In smokers ($n = 202$), an association between smear conversion time and VDR genotype was observed ($p = 0.031$), where individuals with a *TaqI* 'tt' genotype took longer to become sputum negative. Faster culture conversion time was associated with females, low extent of TB at diagnosis, high absolute lymphocyte numbers and low absolute neutrophil numbers.

Table 1 Case-control analysis of VDR genotype and pulmonary tuberculosis in a South African Colored population.

Genotype		Control *n* (%)	Cases *n* (%)	*p*-value*
FokI	ff	20 (6)	13 (5)	0.447
	Ff	129 (37)	104 (42)	
	FF	203 (58)	132 (53)	
ApaI	aa	63 (18)	40 (16)	0.160
	Aa	173 (49)	108 (43)	
	AA	116 (33)	101 (41)	
TaqI	Tt	22 (6)	19 (8)	0.751
	Tt	140 (40)	94 (38)	
	TT	190 (54)	136 (55)	

*p-values were derived from Fisher's exact test.

Table 3 Observed diplotypes of VDR gene polymorphisms *FokI*, *ApaI* and *TaqI*, in pulmonary tuberculosis cases and controls from a South African Colored population.

Diplotype	Controls *n* (%)	Cases *n* (%)	Total *n* (%)
ffaaTT	3 (1)	2 (1)	5 (1)
ffAaTt	5 (1)	5 (1)	10 (2)
ffAaTT	5 (1)	0 (0)	5 (1)
ffAAtt	2 (1)	3 (1)	5 (1)
ffAATt	4 (1)	2 (1)	6 (1)
ffAATT	1 (<1)	1 (<1)	2 (<1)
FfaaTT	25 (7)	17 (7)	42 (7)
FfAaTt	23 (7)	25 (10)	48 (8)
FfAaTT	39 (11)	27 (11)	66 (11)
FfAAtt	8 (2)	7 (3)	15 (2)
FfAATt	21 (6)	17 (7)	38 (6)
FfAATT	13 (4)	11 (4)	24 (4)
FFaaTT	35 (10)	21 (8)	56 (9)
FFAaTt	50 (14)	20 (8)	70 (12)
FFAaTT	51 (14)	31 (12)	82 (14)
FFAAtt	12 (3)	9 (4)	21 (3)
FFAATt	37 (11)	25 (10)	62 (10)
FFAATT	18 (5)	26 (10)	44 (7)
Total *n*	352	249	601

Note: Logistic regression analysis of cases and controls, adjusting for age and gender, gave a *p*-value = 0.24.

Table 2 Two and three locus inferred haplotypes for the VDR gene polymorphisms *FokI*, *ApaI* and *TaqI* for pulmonary tuberculosis cases and controls in a South African Colored population.

Haplotype	% of Controls ($n = 352$)	% of Cases ($n = 249$)	*p*-value	Global *p*-value
FokI, *ApaI*, *TaqI*				
faT	10	11	0.913	0.078
fAt	6	9	0.117	
fAT	8	6	0.826	
FaT	32	27	0.063	
FAt	20	17	0.386	
FAT	24	30	0.062	
ApaI, *TaqI*				
aT	43	38	0.099	0.206
At	26	27	0.896	
AT	31	36	0.111	
FokI, *ApaI*				
fa	10	11	0.866	0.279
fA	14	15	0.327	
Fa	32	27	0.069	
FA	45	44	0.285	
FokI, *TaqI*				
ft	6	10	0.107	0.247
fT	18	17	0.950	
Ft	20	17	0.388	
FT	56	57	0.947	

Note: Haplotypes for the polymorphisms were inferred and tested for association with pulmonary TB using the R package: haplo.stats.

Table 4 VDR genotypes of fast versus slow responders to tuberculosis treatment.

Polymorphism, genotype	Smear ($n = 220$)		Culture ($n = 222$)	
	Fast responders[a] n (%)	Slow responders[b] n (%)	Fast responders[a] n (%)	Slow responders[b] n (%)
FokI				
ff	10 (6)	1 (2)	5 (5)	6 (5)
Ff	61 (38)	33 (55)	38 (40)	58 (46)
FF	89 (56)	26 (43)	53 (55)	62 (49)
		$p = 0.055$[c]		$p = 0.649$[c]
ApaI				
aa	28 (18)	11 (18)	10 (10)	29 (23)
Aa	67 (42)	24 (40)	47 (49)	46 (37)
AA	65 (41)	25 (42)	39 (41)	51 (40)
		$p = 0.959$[c]		$p = 0.029$[c]
TaqI				
tt	9 (6)	6 (10)	7 (7)	8 (6)
Tt	63 (39)	21 (35)	40 (42)	45 (36)
TT	88 (55)	33 (55)	49 (51)	73 (58)
		$p = 0.490$[c]		$p = 0.573$[c]

[a]Pulmonary tuberculosis cases that responded to treatment before day 55 (month 2).
[b]Pulmonary tuberculosis cases that responded to treatment on or after day 55 (month 2).
[c]A Fisher exact test was done. The p-values were not adjusted for multiple comparisons.

Individual Cox regression analysis: In the age and gender adjusted individual Cox analyses of smear conversion time, only the following factors were found to have a significant ($p < 0.050$) effect on faster conversion: not smoking, high school completion, low extent of TB at diagnosis, low total WBC and absolute neutrophil numbers, and high absolute lymphocyte numbers. The number of cavities present at diagnosis was weakly associated with smear conversion time ($p = 0.054$). In the individual, age and gender adjusted, Cox analyses of culture conversion time, only the following factors significantly affected faster conversion: low extent of TB at diagnosis, short height, high absolute lymphocyte numbers and low absolute neutrophil numbers.

Optimal Cox model: Variables that were independently predictive of faster smear conversion time in the final, optimal Cox model (Table 5), were the genotypes *ApaI* 'AA' versus 'aa' and *TaqI* 'Tt' and 'TT' versus 'tt'. Low extent of TB at diagnosis, not smoking and low total WBC count at diagnosis were also predictive. Variables that were significantly associated and independent predictors in the optimal Cox model for faster culture conversion time were low extent of TB at diagnosis, short height and low absolute neutrophil numbers. The magnitude of the effect of the variables can be seen by referring to the estimated coefficient in Table 5.

The final optimal Cox models constructed for time to smear and culture conversion, were both highly significant ($p = 0.0004$) (Table 5). These models obviate the need to adjust for multiple testing.

Discussion

Host genetic susceptibility to the development of TB after infection, can be determined via case-control studies comparing genotypes between the groups. More subtle effects of genotype are those impacting on the severity of disease, or the ability of the treated TB cases to recover i.e. undergo sputum conversion to negative culture or smear stain. We investigated both of the above aspects with respect to the VDR gene in a large study of a South African population, as VDR polymorphisms have been implicated both in susceptibility to pulmonary TB disease,[6–9] and recently, in the time to mycobacterial resolution of TB in a Peruvian population.[16]

No significant association was found in our case-control analysis between pulmonary TB and the VDR polymorphisms. Diplotypes and haplotypes were also analysed because they are believed to be more informative than individual polymorphisms,[8] but we found only weak associations between the VDR haplotype and TB susceptibility. In our study the *FokI-ApaI-TaqI* 'FAT' haplotype tended to be associated with TB and may be a risk factor, whereas the 'FaT' haplotype is possibly protective. In a TDT study of West African families, Bornman et al. found an association between TB and the *FokI-ApaI* 'FA' haplotype where the *ApaI* 'A' allele showed increased transmission to affected offspring.[8] However, no association between the *ApaI* genotype and TB susceptibility was found in a Gambian case-control study.[6] In our case-control study 41% of cases had the *ApaI* 'AA' genotype compared with 33% of controls, but this difference was non-significant, as was the association of the *FokI-ApaI* 'Fa' haplotype with controls ($p = 0.070$). There was no association between the VDR diplotype and TB susceptibility. A larger sample set might provide more information although the present case-control study provided 80% power to detect an odds ratio of 1.7 or higher. The polymorphisms investigated here may be in LD with another polymorphism which determines susceptibility, possibly the polyA repeat region which affects

Table 5 Optimal Cox regression analysis model of factors influencing time to sputum conversion in tuberculosis patients.

Variables	Smear ($n = 220$)		Culture ($n = 222$)	
	p-value[b]	Est. coef.[c]	p-value[b]	Est. coef.[c]
Age	0.910	−0.001	0.560	−0.004
Gender	0.940	0.011	0.440	−0.145
Smoking Yes/No	**0.047**	−0.675	–	–
High school completion Yes/No	0.150	0.272	0.078	0.318
Extent of TB at diagnosis from chest radiograph (<RUL[a], = RUL, >RUL, >1lung)	**0.010**	−0.186	**0.024**	−0.164
Total white blood cell count at diagnosis	**0.039**	−0.043	–	–
Absolute neutrophil numbers at diagnosis	–	–	**0.023**	−0.026
Height (cm)	–	–	**0.003**	−0.029
FokI Ff versus *ff*	0.094	−0.576	0.470	0.262
FokI FF versus *ff*	0.480	−0.235	0.180	0.476
ApaI Aa versus *aa*	0.130	0.338	0.960	−0.011
ApaI AA versus *aa*	**0.040**	0.480	0.920	0.024
TaqI Tt versus *tt*	**0.047**	0.600	0.740	0.105
TaqI TT versus *tt*	**0.026**	0.679	0.960	−0.014
Final model Global p-value	**0.0004**		**0.0004**	

[a]Right upper lobe.
[b]A significant p-value (in bold) indicates that a variable in the model contributes independently to conversion time.
[c]Estimated coefficient for modeled conversion time in days.

transcription.[21] The patterns of LD vary between populations, possibly accounting for the inconsistency in results from different investigations.[9,30]

Comparison between the fast and slow responders to therapy indicated that individuals with an *ApaI* 'A'-containing genotype (culture conversion time $p = 0.029$) had a faster response to treatment, and tended to convert before day 55 (month 2) when on standard DOTS treatment. *FokI* 'F'-containing genotypes tended to show delayed conversion, but the frequency of *FokI* 'ff' was too low in our population to show a significant effect (smear conversion time $p = 0.055$). Roth et al. ($n = 78$) found a significant difference at an earlier time point of 30 days (month 1), where the homozygous *FokI* genotype of the less frequent allele was associated with faster conversion time by culture or smear.[16] Their study also found an association between *TaqI* genotype and culture conversion time at 30 days although no *TaqI* 'tt' genotypes were present in the Peruvian cases. We found no association with *TaqI* and the response to treatment at day 30 or day 55, but *TaqI* was an independently significant predictor of smear conversion time in our final, optimal Cox model where *TaqI* 'tt' genotype was associated with delayed smear conversion time. Such associations could be used to identify people who would benefit from more intensive treatment or follow up.

In the cohort study of pulmonary TB cases, the only variable to contribute independently to the final optimal model of both smear and culture conversion time during treatment, was the extent of TB at diagnosis as determined by chest radiograph. It could be expected that the extent of TB at diagnosis would be associated with the length of time to sputum conversion, with severely affected cases taking longer, as has been shown to occur with MDR disease.[31] However, this is the first report, to the best of our knowledge, of a correlation between the severity of TB disease and the rate of treatment response.

In the final, optimal model of variables contributing to smear conversion time, VDR genotypes contributed independently to the model. Both the *TaqI* and *ApaI* genotypes were independently significant, as were smoking status and total WBC count at diagnosis. In the culture conversion optimal model, additional independently significant variables were absolute neutrophil numbers at diagnosis and height. Although not independently predictive of faster conversion, high school completion contributed to both optimal models, and lower educational level is a known risk factor for TB.[32,33] The final models were both highly significant ($p = 0.0004$).

The final model of the Cox regression analysis therefore indicates that the following variables are independent predictors of a faster conversion time in smear and/or culture: not smoking, low extent of TB at diagnosis, low total WBC count and low absolute neutrophil number at diagnosis, shorter height, an *ApaI* 'AA' genotype, and a *TaqI* 'T' containing VDR genotype. In the optimal model for smear conversion (Table 5), it can be seen from the estimated coefficient that VDR genotype and smoking had a greater clinical effect than the other variables. All effects were fairly modest.

The function of the *TaqI* polymorphism is not clear. Previously, the *TaqI* 'tt' genotype was found to protect against TB in a Gambian population[6] and the combination of the *TaqI* 'TT/Tt' genotype and 25-hydroxycholecalciferol deficiency was associated with TB in Gujarati Asians.[10] Contrary to these indications that *TaqI* 't' may be advantageous, the *TaqI* 'tt' genotype is associated with decreased levels of VDR mRNA and protein levels in peripheral blood mononuclear cells.[34] Roy et al. found the *TaqI* 'tt' genotype to be associated with susceptibility to

leprosy *per se*,[35] and *TaqI* 'tt' predisposed Tamil-speaking females to TB.[7] Our results correspond with the latter views of the function of *TaqI* in that cases with a 'tt' genotype take longer to respond to TB chemotherapy, and therefore may have a poorer immune response. Even though a patient is undergoing chemotherapy for TB, an efficiently functioning immune system is still necessary for optimum response and early sputum conversion.

The negative correlation between total WBC count at diagnosis or absolute neutrophil number at diagnosis and time to conversion in the final models for smear or culture, respectively, may reflect the involvement of both of these cell types in the acute response to inflammation. Our results extend to the human model, the evidence recently presented by Keller et al.[36] that an early influx of granulocytes contributes to susceptibility to *M. tuberculosis* in mice.

An association between smoking and TB infection status (a positive tuberculin skin test) has been reported in our population[37] and smokers have a higher rate of progression from infection to clinical TB in India.[38] We have shown here that smoking is also a significant factor in delaying sputum conversion time while on DOTS, as measured by smear. However, the number of non-smokers in this cohort was low and a further study would be needed to confirm this. The association between height and culture conversion time could be an artifact due to the fact that there were a small number of people of small stature and high mass in whom culture conversion time was unusually fast. As can be seen from the estimated coefficient in Table 5, the actual effect of height was negligible.

The lack of concordance in the results between smear (ZN) and culture (BACTEC) conversion illustrates the complexity of this disease and the subtly different end points measured by the two techniques. At month 2, 31% of patients were smear negative but still culture positive. Roth et al. found an association between VDR genotypes and culture conversion times (as measured by Micro Observation Direct Susceptibility test)[16] whereas in our large study of patients with strict inclusion criteria and intensive follow-up, associations were found with smear conversion times. Culture is more sensitive, but as a very low number of bacteria can give rise to a positive culture, it may be a less robust tool for measuring an endpoint of sputum conversion time.

Currently, in most TB programmes, the first measure of effective therapy is sputum conversion at 2 months. The discovery of a genetic or immunological marker indicating how efficiently a patient responds to treatment could have a major impact on clinical trials and shorten the time period necessary for drug testing. In the coming era of personalised medicine, genetic markers have the potential to be developed into simple, affordable tests applicable even in high-burden countries. Estimates of disease transmission in the community studied here are extremely high, and the percentage of TB due to transmission could be in excess of 70%.[39] The extended time to microbial resolution of TB may be a contributory factor in the continuation of the TB epidemic.

Acknowledgements

We thank the following: The Ravensmead and Uitsig community of the Western Cape, South Africa for their willingness in participating in the project; the nurses in the Desmond Tutu TB Centre, Stellenbosch University for their dedication to the project; Katherine Lawrence for database management; Gillian Black for database input; Nora Carroll and Erica Engelke for bacteriology testing and Robert Gie for evaluating chest radiographs.

Funding: Financial support from GlaxoSmithKline Action TB and a European and Developing Countries Clinical Trials Partnership (EDCTP code 2004.01.T.d1) grant to Prof. P. van Helden.

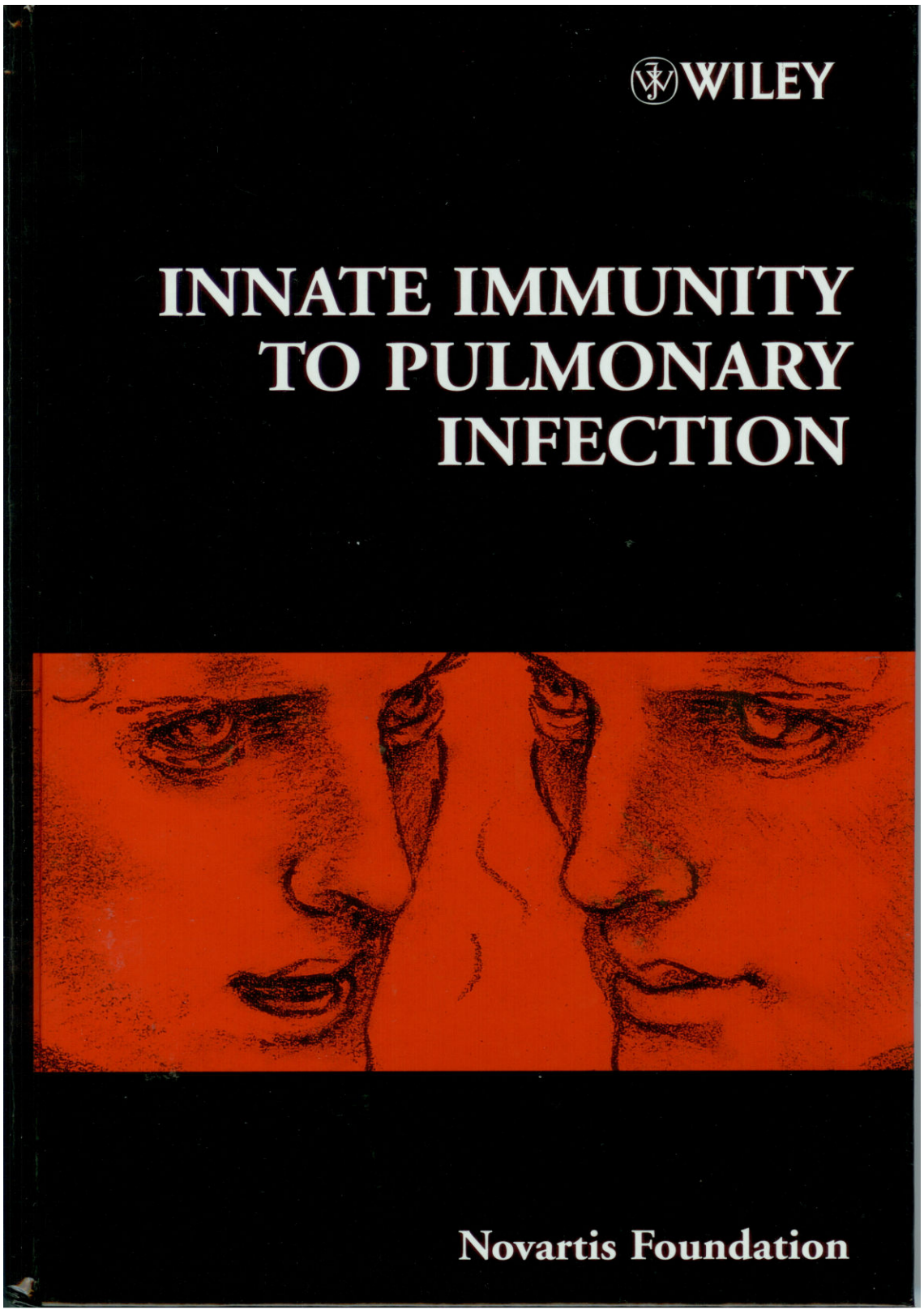
Competing interests: None declared

Ethical approval: Obtained from Stellenbosch University and the City of Cape Town (see Methods).

References

1. Incidence of Tuberculosis South Africa Department of Health. Pretoria: Department of Health . 2004.
2. National HIV sero-prevelance survey of women attending public antenatla clinics in South Africa 1998. Pretoria: Department of Health, 1999.
3. Rossouw M, Nel HJ, Cooke GS, van Helden PD, Hoal EG. Association between tuberculosis and a polymorphic NFkappaB binding site in the interferon gamma gene. *Lancet* 2003;**361**: 1871–2.
4. Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV. Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. *N Engl J Med* 1998;**338**:640–4.
5. Hoal EG, Lewis L-A, Jamieson SE, Tanzer F, Rossouw M, Victor T, et al. *SLC11A1* (*NRAMP1*) but not *SLC11A2* (*NRAMP2*) polymorphisms are associated with susceptibility to tuberculosis in a high-incidence community in South Africa. *Int J Tuberc Lung Dis* 2004;**8**:1464–71.
6. Bellamy R, Ruwende C, Corrah T, McAdam KP, Thursz M, Whittle HC, et al. Tuberculosis and chronic hepatitis B virus infection in Africans and variation in the vitamin D receptor gene. *J Infect Dis* 1999;**179**:721–4.
7. Selvaraj P, Narayanan PR, Reetha AM. Association of vitamin D receptor genotypes with the susceptibility to pulmonary tuberculosis in female patients & resistance in female contacts. *Indian J Med Res* 2000;**111**:172–9.
8. Bornman L, Campbell SJ, Fielding K, Bah B, Sillah J, Gustafson P, et al. Vitamin D receptor polymorphisms and susceptibility to tuberculosis in West Africa: a case-control and family study. *J Infect Dis* 2004;**190**:1631–41.
9. Lewis SJ, Baker I, Davey SG. Meta-analysis of vitamin D receptor polymorphisms and pulmonary tuberculosis risk. *Int J Tuberc Lung Dis* 2005;**9**:1174–7.
10. Wilkinson RJ, Llewelyn M, Toossi Z, Patel P, Pasvol G, Lalvani A, et al. Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London: a case-control study. *Lancet* 2000;**355**:618–21.
11. Fitness J, Floyd S, Warndorff DK, Sichali L, Malema S, Crampin AC, et al. Large-scale candidate gene study of tuberculosis susceptibility in the Karonga district of northern Malawi. *Am J Trop Med Hyg* 2004;**71**:341–9.
12. Denis M. Killing of Mycobacterium tuberculosis within human monocytes: activation by cytokines and calcitriol. *Clin Exp Immunol* 1991;**84**:200–6.
13. Rockett KA, Brookes R, Udalova I, Vidal V, Hill AV, Kwiatkowski D. 1,25-Dihydroxyvitamin D3 induces nitric oxide synthase and

suppresses growth of Mycobacterium tuberculosis in a human macrophage-like cell line. *Infect Immun* 1998;**66**:5314–21.
14. Liu PT, Stenger S, Li H, Wenzel L, Tan BH, Krutzik SR, et al. Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. *Science* 2006;**311**:1770–3.
15. Liu W, Cao WC, Zhang CY, Tian L, Wu XM, et al. VDR and NRAMP1 gene polymorphisms in susceptibility to pulmonary tuberculosis among the Chinese Han population: a case-control study. *Int J Tuberc Lung Dis* 2004;**8**:428–34.
16. Roth DE, Soto G, Arenas F, Bautista CT, Ortiz J, Rodriguez R, et al. Association between Vitamin D Receptor Gene Polymorphisms and Response to Treatment of Pulmonary Tuberculosis. *J Infect Dis* 2004;**190**:920–7.
17. Barreiro LB, Neyrolles O, Babb CL, Tailleux L, Quach H, McElreavey K, et al. Promoter Variation in the DC-SIGN Encoding Gene CD209 Is Associated with Tuberculosis. *PLoS Medicine* 2006;**3**:e20.
18. Baker AR, McDonnell DP, Hughes M, Crisp TM, Mangelsdorf DJ, Haussler MR, et al. Cloning and expression of full-length cDNA encoding human vitamin D receptor. *Proc Natl Acad Sci USA* 1988;**85**:3294–8.
19. Nejentsev S, Godfrey L, Snook H, Rance H, Nutland S, Walker NM, et al. Comparative high-resolution analysis of linkage disequilibrium and tag single nucleotide polymorphisms between populations in the vitamin D receptor gene. *Hum Mol Genet* 2004;**13**:1633–9.
20. Uitterlinden AG, Fang Y, van Meurs JBJ, van Leeuwen H, Pols HAP. Vitamin D receptor gene polymorphisms in relation to vitamin D related disease states. *J Steroid Biochem Mol Biol* 2004;**89–90**:187–93.
21. Whitfield GK, Remus LS, Jurutka PW, Zitzer H, Oza AK, Dang HT, et al. Functionally relevant polymorphisms in the human nuclear vitamin D receptor gene. *Mol Cell Endocrinol* 2001;**177**:145–59.
22. Walzl G, Beyers N, van Helden P. TB: a partnership for the benefit of research and community. *Tran R Soc Tropical Med Hygiene* 2005;**99**:S15–9.
23. Brahmbhatt S, Black GF, Carroll NM, Beyers N, Salker F, Kidd M, et al. Immune markers measured before treatment predict outcome of intensive phase tuberculosis therapy. *Clin Exp Immunol* 2006;**146**:243–52.
24. Simon G. Radiology in epidemiological studies and some therapeutic trials. *Br Med J* 1966;**2**:491–4.
25. R: A language and environment for statistical computing. R Development Core Team. R foundation for statistical computing, Vienna, Austria. 2006. 〈www.r-project.org〉.
26. Therneau T, Lumey T. Survival analysis including penalised likelihood. R package version 2.20, 2006.
27. Warnes G, Leisch F. Genetics: population genetics. R package version 1.2.0. 2005.
28. Sinnwell JP, Schaid DJ. Statistical analysis of haplotypes with traits and covariates when linkage phase is ambiguous. R package version 1.2.2, 2006.
29. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Human Genetics* 2002;**70**:425–34.
30. Delgado JC, Baena A, Thim S, Goldfeld AE. Ethnic-specific genetic associations with pulmonary tuberculosis. *J Infect Dis* 2002;**186**:1463–8.
31. Holtz TH, Sternberg M, Kammerer S, Laserson KF, Riekstina V, Zarovska E, et al. Time to sputum culture conversion in multidrug-resistant tuberculosis: predictors and relationship to treatment outcome. *Ann Intern Med* 2006;**144**:650–9.
32. Tekkel M, Rahu M, Loit HM, Baburin A. Risk factors for pulmonary tuberculosis in Estonia. *Int J Tuberc Lung Dis* 2002;**6**:887–94.
33. Shetty N, Shemko M, Vaz M, D'Souza G. An epidemiological evaluation of risk factors for tuberculosis in South India: a matched case control study. *Int J Tuberc Lung Dis* 2006;**10**:80–6.
34. Ogunkolade BW, Boucher BJ, Prahl JM, Bustin SA, Burrin JM, Noonan K, et al. Vitamin D receptor (VDR) mRNA and VDR protein levels in relation to vitamin D status, insulin secretory capacity, and VDR genotype in Bangladeshi Asians. *Diabetes* 2002;**51**:2294–300.
35. Roy S, Frodsham A, Saha B, Hazra SK, Mascie-Taylor CGN, Hill AVS. Association of vitamin D receptor genotype with leprosy type. *J Infect Dis* 1999;**179**:187–91.
36. Keller C, Hoffmann R, Lang R, Brandau S, Hermann C, Ehlers S. Genetically determined susceptibility to tuberculosis in mice causally involves accelerated and enhanced recruitment of granulocytes. *Infect Immun* 2006;**74**:4295–309.
37. den Boon S, Van Lill SWP, Borgdorff MW, Verver S, Bateman ED, Lombard CJ, et al. Association between smoking and tuberculosis infection: a population survey in a high tuberculosis incidence area. *Thorax* 2005;**60**:555–7.
38. Gajalakshmi V, Peto R, Kanaka TS, Jha P. Smoking and mortality from tuberculosis and other diseases in India: retrospective study of 43 000 adult male deaths and 35 000 controls. *Lancet* 2003;**362**:507–15.
39. van der Spuy GD, Warren RM, Richardson M, Beyers N, Behr MA, van Helden PD. Use of genetic distance as a measure of ongoing transmission of Mycobacterium tuberculosis. *J Clin Microbiol* 2003;**41**:5640–4.



Published in 2006 by John Wiley & Sons Ltd,
The Atrium, Southern Gate,
Chichester PO19 8SQ, UK

National 01243 779777
International (+44) 1243 779777
e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on http://www.wileyeurope.com
or http://www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, Ontario, Canada L5R 4J3

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Novartis Foundation Symposium 279
x + 238 pages, 23 figures, 5 tables

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13 978-0-470-02656-4 (HB)

Typeset in 10½ on 12½ pt Garamond by SNP Best-set Typesetter Ltd., Hong Kong
Printed and bound in Great Britain by T. J. International Ltd, Padstow, Cornwall.
This book is printed on acid-free paper responsibly manufactured from sustainable forestry, in which at least two trees are planted for each one used for paper production.

Contents

TB epidemiology and human genetics

Paul D van Helden, Marlo Möller, Chantal Babb, Robin Warren, Gerhard Walzl, Pieter Uys and Eileen Hoal

Department of Medical Biochemistry/(US/MRC) Centre for Molecular and Cellular Biology and DST/NRF Centre of Excellence for Biomedical TB Research, Faculty of Health Sciences, University of Stellenbosch, PO Box 19063, Tygerberg 7505, South Africa

Abstract. The impact of tuberculosis (TB) is considerably lower than one may expect, since in the absence of immunosuppression, fewer than 10% of infected individuals will develop active disease. The relatively low proportion of individuals who progress to active disease after infection can probably be ascribed to innate resistance in most infected individuals, since vaccination using BCG or a previous episode of TB does not work reliably or effectively to confer protection in high burden parts of the world. Innate factors affecting resistance or susceptibility can be modulated by the environment and such external influences cannot be ignored. Specifically, we will address bacterial variability as well as environmental factors such as diet, smoking, helminths and hormones. We will also discuss host genes that may be involved in susceptibility or resistance at various stages of infection or disease. The discovery of as yet unknown genes impacting on TB susceptibility or disease course may lead to new insights into mechanisms of disease and novel therapies. With adaptive immunity being of little value and good TB control programmes being rare, innate resistance is still our best defence against this disease.

2006 Innate immunity to pulmonary infection. Wiley, Chichester (Novartis Foundation Symposium 279) p 17–41

It is estimated that approximately one third of the global population is infected by tuberculosis (TB). Fortunately, far fewer than 10% of those infected eventually become ill (in the absence of immunosuppression). Studies on families, twins and adoptees have suggested that there is clearly a genetic component involved in susceptibility or resistance to TB. On the other hand, recent research has shown that there are many different strains of *Mycobacterium tuberculosis (Mtb)* and that they can influence the course of disease. Clearly we have to deal with on the one hand a successful pathogen, but on the other hand a host that has learnt to live with this pathogen. It has been proposed that the bacterium is perhaps 10 000 years old and evolved as humans domesticated cattle and settled in villages. Recently, however, it has been suggested that the organism may be as old as 3 million years, which implies a long period of co-evolution and thus adaptation on both sides (Gutierrez

et al 2005). At least four successive epidemics of tuberculosis are thought to have occurred in the last 4000 years, viz. in the Nile Valley, Greece, the Americas (approximately 1000 years ago) and Europe. Each wave probably spans centuries, but incidence peaks probably last a few decades only.

In general, when *Mtb* is introduced into a naïve population living under harsh conditions, it may spread rapidly. Thus, in the UK with the massive migration into cities that occurred during the industrial revolution and the poor living conditions at that time, an ideal opportunity for an epidemic was created. In London, it was estimated that 20% of all deaths in 1667 were due to TB. TB peaked (possibly) in the UK around 1780 (early industrial revolution) at about 1120/100 000 p.a., or, it is estimated that 1.25% of the entire population died each year from TB. Thereafter, in England, TB started to decline years before other infectious diseases and long before the introduction of control programmes or antibiotics.

A recent mirror of this epidemic was seen in the Inuit, where, after introduction of TB to a naïve settlement, most individuals died and the epidemic rapidly waned. The waning of any epidemic may occur once the living conditions change or the population becomes more resistant, as susceptible individuals disappear. We argue that this is not due to adaptive immunity, but innate immunity which is likely to increase with exposure to the organism and subsequent removal (by death) of susceptible individuals.

Infection by *Mtb* is a complex and multistage process proceeding from the initial encounter with the pathogen. For this reason we need to imagine a multistep process (Fig. 1).

At each stage in this process, innate factors may play an important role. While there is a body of evidence that suggests there may be some immunity acquired from prior exposure to Mycobacterial species (e.g. BCG vaccination), there is also much evidence to suggest that prior infection does not necessarily confer any protection against further infection or progression to disease (Rook et al 2005, Cosma et al 2004).

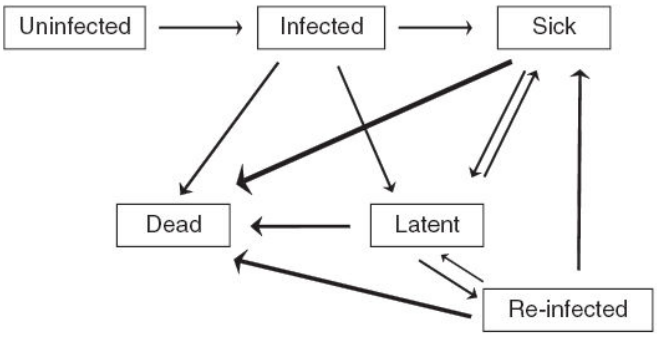


FIG. 1. Transitions in tuberculosis.

Apart from living conditions and an increased (genetic) innate resistance due to death of susceptibles, other factors can play a role: in England and much of the developed world, in the 1800s for the first time food production exceeded population growth and real wages allowed the purchase of adequate food. This is very likely one of the factors that contributed to the decline in TB after 1830, since it is known that a person who is 10% underweight has a threefold increased risk for developing TB after infection.

Excluding sociological determinants, the biological factors likely to be involved in innate resistance to *Mtb* are: (1) the bacterium, (2) the environment (nutrition and other infections), (3) the host. These will be considered below.

The bacterium

Infectious disease studies are complicated by the fact that two genomes, one prokaryotic and one eukaryotic, are interacting in an age-old contest. Genotyping studies have shown that there are thousands of different *Mtb* strains in circulation and comparative genomics has shown that the genome of *Mtb* has evolved through single nucleotide polymorphisms (SNPs), insertions and deletions. This has prompted researchers to investigate the relationship between genome variation and phenotype. A study by Tsolaki et al (2004) suggested a correlation between deletions and the severity of disease while Manca et al (2001) showed that a deletion in the pks 1–15 gene encoding production of phenolic glycolipid was responsible for an altered immune response.

Recent studies have concluded that the 'Beijing' strain is more pathogenic, causes a febrile response on infection and has a higher propensity to develop drug resistance. Furthermore, this strain induces a Th2 immune response on infection allowing for progression towards disease. This differs from the principle genetic group 2 strain CDC1551 which induces a strong Th1 response and less progression towards disease (Manca et al 2001). When the above strains were tested in an *in vivo* mouse model it was shown that the Beijing strain was more pathogenic and could outgrow the CDC1551 strain. There is ample evidence that shows that *Mtb* strains have different growth rates and prompt variable host responses, e.g. cytokine and T cell responses (Manca et al 2001, Janulionis et al 2005, Hoal-van Helden et al 2001a and 2001b). This evidence suggests that these effects are also host dependent. Despite these advances in defining different levels of pathogenicity, many mechanisms underlying these differences in the bacterium remain to be elucidated. It is hypothesised that two evolutionary scenarios may explain these observations: (1) distant evolutionary events which induce an inherited trait that is characteristic of the evolutionary lineage, and (2) recent evolutionary events which induce an inherited trait that is characteristic of a sub-population within a defined lineage.

Thus, signature polymorphisms in *Mtb* isolates may be associated with specific innate (and adaptive) reactions. Some of these also provide a growth advantage and explain the abundance of certain strain types regionally or globally.

The environment

Environmental factors that could be involved in innate resistance to TB include the infection pressure from *Mtb* in the immediate environment, nutrition, other infectious organisms, hormones (e.g. steroid hormones) and substance abuse, such as tobacco smoke (den Boon et al 2005) which suppresses macrophage activity. It is well known that TB can be associated with poverty, which in turn is associated with malnutrition, not only calorie deficit, but more importantly in the case of infectious diseases, with micronutrient imbalances.

Recent work has provided evidence that the different behaviour of omega-3 versus -6 lipids seen at the level of cells and organisms can also be detected *in vitro* in the membrane of the phagosome enclosing mycobacteria. Using phagosomal membrane actin assembly as a functional, *in vitro* readout, these studies showed that the omega-6 lipid, arachidonic acid, as well as six other pro-inflammatory lipids, could stimulate phagosome actin assembly, fusion with lysosomes and a significant increase in pathogen killing. In contrast, the addition of the omega-3 lipids, especially eicosapentanoic acid, suppresses phagosomal actin assembly and induces a significant increase in the growth of pathogenic mycobacteria (*Mtb* and *M. avium*) in macrophages (Anes et al 2003). The ability of these lipids to increase pathogen growth has also been shown to operate at the level of mice and guinea pigs, in the case of both *Mtb* and Salmonella (Paul et al 1997, Chang et al 1992). Our prediction is that dietary manipulation of omega-6 and other pro-inflammatory lipids should help to restrict the growth of pathogens within macrophage phagosomes. Even short term dietary intervention can have dramatic effects on the above-mentioned processes in animal models (Kris-Etherton et al 2002).

The gender bias in tuberculosis has never been satisfactorily explained, but may yield clues to innate resistance of susceptibility factors. Population or gender-based dietary consumption habits would influence disease prevalence. Worldwide, the same gender bias in TB disease is seen. In childhood, no significant differences are noted, but during adolescence girls experience an initially higher rate of TB, whereas in adulthood, males experience a considerably higher disease incidence. Some of this difference in adulthood may be ascribed to behavioural and cultural differences, but other factors are likely to be important. An example of this is the food consumption preferences seen between males and females (see Table 1 below). However, the gender bias may also be linked to steroid hormones, such as DHEA (dehydroepiandrosterone), which have been shown to influence the course of TB

TABLE 1 Gender-based nutritional bias of TB patients

	Mean daily amount	
	Males (n = 23)	*Females (n = 7)*
Fe (mg)	15 (6)	8.7 (8.1)
Mg (mg)	341 (346)	233 (258)
Zn (mg)	14.2 (9.4)	10.1 (6.8)
Se (μg)	71 (45)	30 (45)
Vitamin A (μg)	688 (625)	893 (500)
β-carotene (μg)	2389 (3000–6000)	3522 (3000–6000)
Vitamin B12 (μg)	12.6 (2)	4 (2)
Vitamin C (mg)	67 (75)	108 (60)
Folate (μg)	232 (320)	193 (320)

These figures based on actual food consumed while in hospital (recommended amounts are shown in brackets); see Roberts et al (2005).

disease, specifically at higher concentrations exacerbating pathology (Rook et al 1997).

The net effect of increased iron is to increase risk for active TB, as is the lower levels of some key antioxidant vitamins. The overall effect of these micronutrients and the proteins (such as NRAMP1) involved in their homeostasis (contributing to the 'ionome', Eide et al 2005) is clearly critical. For example, it is known that *Mtb* has an absolute requirement for iron, and that iron supplements should be avoided during TB disease. Furthermore, it has been shown that the total antioxidant profile (could be regarded as a general measure of ion and vitamin status) is significantly lower in TB patients than controls, however, a causative relationship has not yet been established (Wiid et al 2004).

Finally, the effect of multiple infections needs to be considered. Mathematical modelling suggests that in an area of high TB incidence and ARI (annual risk of infection), multiple infection (or super-infection) would be common (Fig. 2).

The simple probability that a particular individual will experience exactly k infection events during a stay of n years (n may be fractional) in a community where the ARI is given by p is:

$$ {}_{n}\mathrm{P}_{k} = n^{k} p^{k} e^{-np} / k! \quad (n = 0, 1, 2, \ldots, k \leq n) $$

Recently, superinfection or reinfection has been proven to occur frequently as predicted (van Rie et al 1999, Warren et al 2004). In an elegant experiment done in a zebrafish model with *M. marinum*, Cosma et al (2004) showed that newly infecting mycobacteria track directly to an existing granuloma harbouring bacteria from a prior infection. The net effect of this is not known, but superinfection may drive

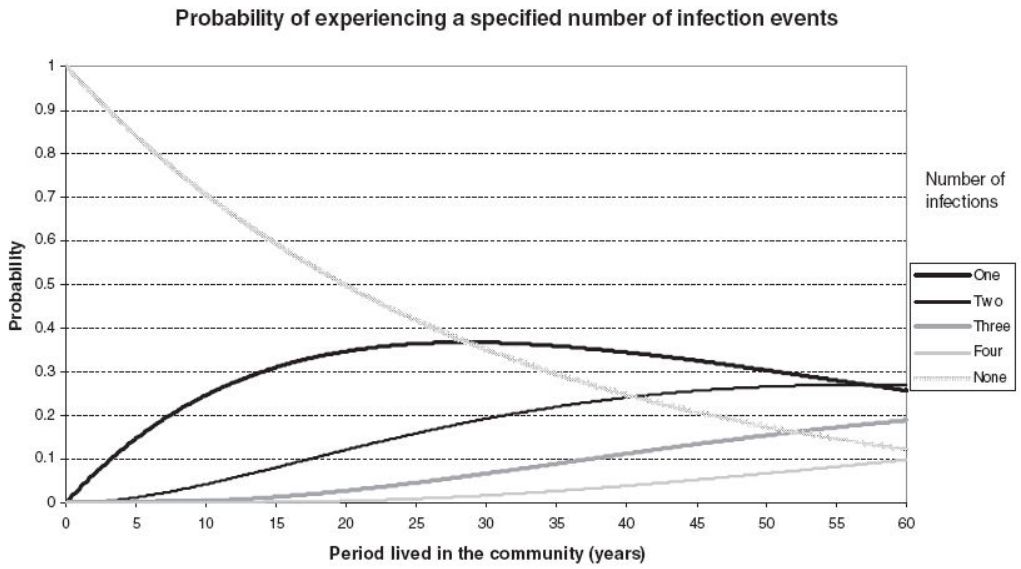


FIG. 2. Muliplicity of Infection. Graphs of ${}_nP_k$ for various k (0, 1 . . . 4), with n along horizontal axis (units of years) for the case of $P = 3.5\%$.

the progression from infection to active disease by either activating latent bacteria (e.g. by means of resuscitation factors, rpf) or by simply overwhelming the innate and adaptive immunity of the host. This effect may be linked to the observation that infection by high doses of mycobacteria will induce a Th2-type response (stimulation of IL4 secretion) rather than Th1, such as may occur on low dose exposure (Rook et al 2005).

Finally, the reaction to infection by *Mtb* is complicated by the intimate connection between the innate and adaptive immune systems and that in reality, most newborn humans are vaccinated with *M. bovis* BCG. Such vaccination may provide for up to 80% protection in developed countries, but far less or none in developing countries, particularly those in the tropics (Rook et al 2005). This may be linked to the mixed Th1/Th2 response in countries of the tropics, which may well be a consequence of exposure to helminths. Thus, the apparent innate response to *Mtb* infection is de facto a 'primed' response and differs in individuals according to exposure to other infecting organisms, such as environmental (myco) bacteria and helminths (Rook et al 2005).

The host

The host defences against intracellular bacteria are mainly cell-mediated but also humoral and therefore any genetic deficiencies in components that play a role in

these systems can lead to susceptibility. While there is a definite role for environmental factors, studies have indicated that genetic factors may be even more important than the environment in determining the outcome of infection. The macrophage is usually the first important cell encountering the invading pathogen, and many of the genes governing macrophage function can be expected to influence this essential first step in the innate defence system.

Investigations of TB, a multifactorial disease, have to take into account that there is likely to be an interaction between environmental factors and common polymorphisms in a number of genes. A large body of evidence points to the major role of genetic factors in the human response to a number of infectious pathogens, and these genes could also impact on treatment and vaccine efficacy. The approaches that have been used to identify the genetic component include segregation analysis, animal models and linkage analysis. Understanding the immune responses of individuals with more resistant genotypes, particularly where this can be replicated in a number of different populations, could suggest novel therapies to combat this highly successful pathogen. Complex disease, unlike monogenic conditions, can be influenced by several genes, with each gene making a small contribution to the overall susceptibility to the disease. Tuberculosis is perhaps more complex than most in that the different phenotypes or forms of the disease such as cavitatory TB, pleural effusion, TB meningitis, etc. may be influenced by different genes.

Identification of common TB susceptibility genes

Complex traits such as TB can be investigated via two general designs. Firstly, family-based linkage analysis via genome-wide scanning, and secondly, population-based association studies of candidate genes.

Genome scans

The major advantage of the model-free genome scan is that novel genes may be identified. Although the phenotype is usually TB, it is possible that using intermediate phenotypes in other immune pathways could indicate as yet unsuspected genes. The first genome scan in TB was conducted on two samples of affected sibling pairs from The Gambia and South Africa, and identified two regions, on chromosome 15q and Xq (Bellamy et al 2000). The gene UBE3A in the 15q11–q13 region which encodes a ubiquitin ligase in macrophages, was subsequently associated with TB (Cervino et al 2002). A recent genome-wide scan for tuberculosis and leprosy *per se*, conducted in Brazil, found a cluster of susceptibility genes across chromosome 17q11.2 (Jamieson et al 2004) and indicated that four separate candidate genes, NOS2A, CCL18, CCL4 and STAT5B may contribute to this region of linkage.

Candidate gene association studies

Association studies can suffer from lack of reproducibility of results, and it is important that studies be done with large numbers and repeated, preferably in ethnically diverse populations. However, many of these studies have indicated genes and pathways that are important in the pathogenesis of TB (Fig. 3).

A candidate gene approach: current genes

Human leukocyte antigen

HLA-DR2 is most consistently associated with TB in many populations, including Indian, Polish, Thai, Indonesian and Russian (Lombard et al 2006). Nonetheless, inter-population variations in HLA/TB associations have been reported. HLA-DQB1*0503 was found to influence TB progression in the Cambodian population (Goldfeld et al 1998), but not in the people of the Western Cape (Goldfeld & Hoal, unpublished results). DQB1*0601 was associated with TB susceptibility in the Thai and South Indian population and the HLA haplotype DRB1*08032-DQB1*0601 was associated with genetic susceptibility to multidrug-resistant TB in Korean patients. A study of the Venda population showed an association of DRB1*1302 with TB susceptibility (Lombard et al 2006), whereas Boshoff et al (unpublished data) have shown a marginal association of DRB1*03 with TB in the South African Coloured population.

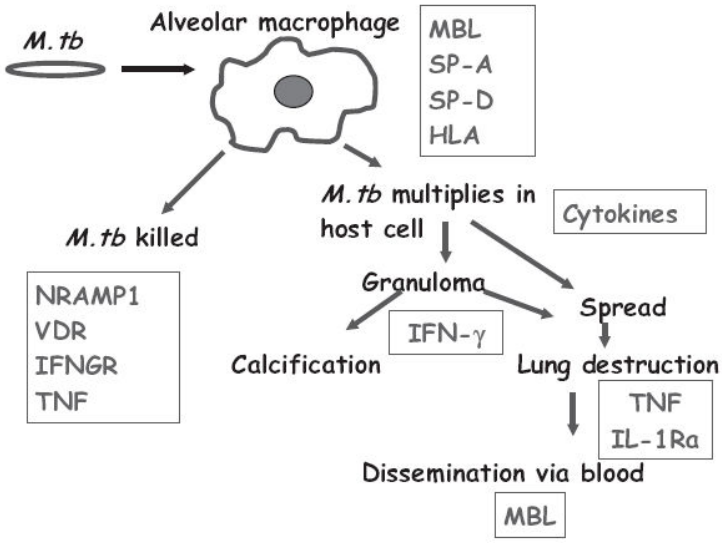


FIG. 3. A simplified representation of the TB disease process and some of the genes that may be involved at different stages.

In this context, it may be worthwhile to explore the relationship between mycobacterial strain type and genetics, e.g. HLA type. For example, in Cambodia, strain diversity is likely to be low and dominated (estimated 60–80%) by the Beijing/W strain type. In the Western Cape, Beijing type strains occur at possibly approximately 20% of total isolates (unpublished data). Therefore, it may be that HLA type is closely linked to *Mtb* strain type in a given locality, and that after extensive exposure, a skewing of HLA type may occur with concomitant resistance to certain strain types. Introduction of new strain types with new epitopes to that locality or ethnic group would then be expected to generate a new epidemic.

Natural resistance-associated macrophage protein

The Natural resistance-associated macrophage protein 1 (*NRAMP1* or *SLC11A1*) gene is a major determinant of natural resistance to intracellular infections, and was originally identified in the mouse model. It is an integral membrane protein expressed only in the lysosome of macrophages and monocytes. After phagocytosis of bacteria, NRAMP1 is targeted to the membrane of the phagosome containing the bacterium, where it may modify the environment to affect the replication of the bacterium, acting as a divalent cation pump which could remove iron or other divalent cations from the phagosome (Blackwell et al 2000).

Associations of *NRAMP1* with TB have been found in Japan, Canada, Korea, Guinea-Conakry, Vietnam, the Gambia and South Africa (Hoal et al 2004) and in most instances the allele over-represented in controls is thought to drive the highest rate of transcription of the protein. Stepwise logistic regression analysis of the South African results indicated that the 5′ and 3′ polymorphisms contribute separate main effects (Hoal et al 2004). More recently, it has been suggested that NRAMP1 may influence only the speed of progression from infection to disease (Malik et al 2005).

Vitamin D receptor

Vitamin D receptor (VDR) is synthesized in monocytes and activated T and B lymphocytes. Its ligand, the active metabolite of vitamin D, calcitriol, is produced in the kidney and by activated monocytes and macrophages, particularly in granulomas. Through its interaction with vitamin D, the retinoid X receptor (RXR) and the vitamin D response element (VDRE), VDR exerts several immunomodulatory effects (Selvaraj et al 2004). These include the activation of monocytes and cell mediated immunity, modulation of the Th1–Th2 host immune response, suppression of lymphocyte proliferation and restriction of *Mtb* survival in macrophages. Vitamin D deficiency is linked to TB by epidemiological evidence. It was found

that the prevalence of both vitamin D deficiency and TB was high in Asians because of their vegetarian diet and thus VDR polymorphisms should not necessarily be viewed in isolation. A particular allele was associated with female patients in an Indian population, and was found to increase susceptibility to pulmonary TB in the Gujerati population, but only in patients with a lack of serum vitamin D (Wilkinson et al 2000). The *Fok1* polymorphism was associated with TB in the Chinese Han population. A large study in West Africa found no association in a case control analysis but an association was found with a particular haplotype in the transmission disequilibrium test family data (Bornman et al 2004). This dependence on the haplotype could explain many of the divergent findings on this and other genes.

Evidence of the subtlety of the effect of VDR polymorphisms in the immune response was found by Roth et al (2004) in Peru, who detected an association with time to sputum conversion in TB patients after diagnosis, but did not find a significant association with susceptibility to TB disease.

Collectins

Mannose-binding lectin. Mannose-binding lectin (MBL) is a serum lectin which acts as an opsonin to promote phagocytosis. Intracellular microorganisms may increase their infectivity by using this system, as it promotes the uptake of bacteria into macrophages where they survive. Low functional MBL-serum levels can occur because of the presence of three variant alleles which lead to an unstable protein. Low MBL levels can protect against infection with *Mtb*. This was found in case-control studies where heterozygosity for the *MBL* variant alleles was associated with protection against the disease and the B allele has also been associated with protection against TB and particularly tuberculous meningitis in South Africa (Hoal-van Helden et al 1999). Conversely, an increased susceptibility to pulmonary TB was found in homozygous carriers of the variant alleles in India and a study in Texas gave equivocal results.

Surfactant proteins (SP)-A and SP-D. Uptake of *Mtb* appears to be facilitated by SP-A and inhibited by SP-D. A Mexican population was typed for polymorphisms in both SP-A and SP-D (Floros et al 2000) and TB cases were compared with two control groups. Using multiple logistic regression analysis, an allele of SP-D was found to be associated with susceptibility to TB only when compared with the skin-test positive control group and an allele each of SP-A1 and SP-A2 was associated with TB susceptibility only when compared with the general control group. This illustrates the extreme sensitivity of association studies to definition of phenotype.

Interferon γ/IL12 pathway

Interleukin (IL)12 stimulates interferon (IFN)γ production by lymphocytes, induces type 1 helper T cell responses and is essential for resistance against infection with intracellular bacteria. It is produced by macrophages particularly when infection with intracellular microorganisms occurs. IL12 is a cytokine composed of a heavy chain (IL12B) and a light chain (IL12A). The functional response of lymphocytes to IL12 is dependent on the expression of the IL12 receptor. Any deficiency in these genes will cause a decrease in IFNγ production.

This pathway has been implicated in TB susceptibility by a wide variety of methods. In the mouse model, gene knockout experiments have indicated the importance of IFNγ, IFNγ receptor 1, and IL12 in susceptibility to mycobacterial infection. In this way, we have a gain of knowledge by loss of function. In the human equivalent, specific gene defects have been found to cause rare familial susceptibility to normally non-pathogenic mycobacteria (Casanova & Abel 2002). The first defect identified in a gene in the IFNγ pathway was the autosomal recessive IFNγ receptor ligand binding (IFNγR1) deficiency. This resulted in an overexpression of a dominant form of the IFNγR1, which binds IFNγ, but lacks the intracellular signalling domain. Detailed investigation of patients has led to the identification of mutations in several autosomal genes in the IFNγ pathway that can increase susceptibility to these atypical mycobacterial infections, including complete IFNγ receptor signal transduction chain (IFNγR2) deficiency, autosomal-dominant partial deficiency of the signal transducer and activator of transcription, and autosomal-dominant partial deficiency of the IFNγR1. An autosomal recessive IL12 deficiency was associated with BCG and *S. enteritidis* infection, as were IL12 receptor deficiencies.

At the population level, association with TB was found with a haplotype of *IL12RB1* in Japan, with an intron 2 allele and a specific haplotype in a large study in Hong Kong, and with 2 promoter polymorphisms in a family-based study in Morocco (Remus et al 2004). In a Croatian population an allele of a polymorphic microsatellite of IFNγR1 was associated with protection against pulmonary TB, but not in a Gambian population (Newport et al 2003).

A promoter polymorphism (+874 A→T) in the IFNγ gene itself, which appears to result in lower NF-κB binding and lower transcription levels of IFNγ, was demonstrated in case-control studies to be associated with susceptibility to TB in Sicily, Spain and South Africa (Rossouw et al 2003). The finding in the South African population was replicated in an independent TDT study, confirming the importance of this gene in tuberculosis at the population level.

DC-SIGN

The transmembrane C-type lectin DC-SIGN (Dendritic Cell Specific Intercellular adhesion molecule [ICAM]-Grabbing Nonintegrin), or CD209, is known to be the major *Mtb* receptor on human dendritic cells. DC-SIGN was found to interact with HIV in 1992, and is now also known to be a pathogen receptor for Cytomegalovirus, Ebola, *Helicobacter pylori*, *Leishmania* and *Shistosoma mansoni*. Unifying features of all these pathogens is that they cause chronic infections that can last a lifetime, and their persistence depends on the manipulation of the Th1/Th2 balance. DC-SIGN binds strongly to mycobacteria such as *Mtb* and *M. bovis* BCG via the mannose capped cell wall component (ManLAM) of the pathogen, but does not bind to LAM that lacks the mannose cap (AraLAM). This is intriguing as ManLAM is abundant in slow growing virulent mycobacteria, such as *Mtb* and *M. leprae*, whereas AraLAM is abundant in fast growing atypical, avirulent mycobacteria, such as *M. smegmatis* and *M. chelona*. It has been suggested that *Mtb* targets DC-SIGN both to infect dendritic cells and to down-regulate the dendritic cell mediated immune response.

We tested whether polymorphisms in DC-SIGN are associated with susceptibility to tuberculosis, and found an association between DC-SIGN promoter variation and risk of developing tuberculosis in our South African cohort. The −871G and −336A allelic combination is significantly overrepresented among healthy controls ($P = 1.6 \times 10^{-3}$) and population stratification was excluded (Barreiro et al 2006). The above allelic combination is usually confined to Eurasian populations, and it is possible that these two variants may have increased in frequency in non-African populations as a result of host genetic adaptation to a longer history of exposure to tuberculosis.

Conclusions

Genetic studies in infectious disease are usually complicated because of the presence of two different genomes and the influence their interaction can have on the disease. Although several genes have been identified as susceptibility genes for a number of intracellular bacteria, it is necessary to bear in mind that other genes and the environment can have an influence on the development of the disease, which is the reason that no single major susceptibility gene has been identified in any infectious human disease. The results from strategies used to identify candidate genes or to associate the candidate genes with infectious disease are not the final word on the subject of susceptibility, but provide important evidence on the pathways involved. A greater understanding of the immune response to TB could provide insights into novel treatments that target genetically based susceptibility, such as aerosolised IFNγ, TNF modulation, or even simple supplementation of

vitamin D. These therapies could specifically target the more vulnerable individuals in a population and lead to improved health in the entire community. It is likely that each individual has a spectrum of risk factors, which will include genes and environmental factors that will confer a risk profile on that individual. Evidence for this is work that shows that individuals who have had a prior episode have a fourfold higher risk for developing another episode of active TB than those who have never had active TB (Verver et al 2005).

Thus, it may be that innate immunity is the most important process protecting individuals against tuberculosis and by understanding this process; we may develop new ways to combat this ancient scourge.

References

Anes E, Kühnel MP, Bos E, Moniz-Pereira J, Habermann A, Griffiths G 2003 Selected lipids activate phagosome actin assembly and maturation resulting in killing of pathogenic mycobacteria. Nat Cell Biol 5:793–802

Barreiro LB, Neyrolles O, Babb CL et al 2006 Promoter variation in the DC-SIGN-encoding gene CD209 is associated with tuberculosis. PLoS Medicine 3:e20

Bellamy R, Beyers N, McAdam KP et al 2000 Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. Proc Natl Acad Sci USA 97:8005–8009

Blackwell JM, Searle S, Goswami T, Miller EN 2000 Understanding the multiple functions of Nramp1. Microbes Infect 2:317–321

Bornman L, Campbell SJ, Fielding K et al 2004 Vitamin D receptor polymorphisms and susceptibility to tuberculosis in West Africa: 1 case-control and family study. J Infect Dis 190:1631–1641

Casanova JL, Abel L 2002 Genetic dissection of immunity to mycobacteria: The human model. Annu Rev Immunol 20:581–620

Cervino ACL, Lakiss S, Sow O et al 2002 Fine mapping of a putative tuberculosis susceptibility locus on chromosome 15q11-13 in African families. Hum Mol Genet 11:1599–1603

Chang HR, Dulloo AG, Vladoianu IR et al 1992 Fish oil decreases natural resistance of mice to infection with Salmonella typhimurium. Metabolism 41:1–2

Cosma CC, Humbert O, Ramakrishnan L 2004 Superinfecting mycobacteria home to established tuberculous granulomas. Nat Immunol 5:828–835

den Boon S, van Lill SWP, Borgdorff MW et al 2005 Association between smoking and tuberculosis infection: a population survey in a high tuberculosis incidence area. Thorax 60: 555–557

Eide DJ, Clark S, Nair TM et al 2005 Characterization of the yeast ionome: a genome-wide analysis of nutrient mineral and trace element homeostasis in Saccharomyces cerevisiae. Genome Biology 6:R77

Floros J, Lin HM, Garcia A et al 2000 Surfactant protein genetic marker alleles identify a subgroup of tuberculosis in a Mexican population. J Infect Dis 182:1473–1478

Goldfeld AE, Delgado JC, Thim S et al 1998 Association of an HLA-DQ allele with clinical tuberculosis. JAMA 279:226–228

Gutierrez MC, Brisse S, Brosch R et al 2005 Ancient origin and gene mosaicism of the progenitor of mycobacteriumtuberculosis. PLoS Pathog 1:e5

Hoal-van Helden EG, Epstein J, Victor TC et al 1999 Mannose-binding protein B allele confers protection against tuberculous meningitis. Pediatr Res 45:459–464

Hoal-van Helden EG, Hon D, Lewis L-A, Beyers N, van Helden PD 2001a Mycobacterial growth in human macrophages: Variation according to donor, inoculum and bacterial strain. Cell Biol Int 25:77–81

Hoal-van Helden EG, Stanton L-A, van Helden PD 2001b Diversity of *in vitro* cytokine responses by human macrophages to infection by *Mycobacterium tuberculosis* strains. Cell Biol Int 25:83–90

Hoal-van Helden EG, Lewis LA, Jamieson S et al 2004 SLC11A1 (NRAMP1) but not SLC11A2 (NRAMP2) polymorphisms are associated with susceptibility to tuberculosis in a high incidence community in South Africa. Int J Tuberc Lung Dis 8:1464–1471

Jamieson SE, Miller EN, Black GF et al 2004 Evidence for a cluster of genes on chromosome 17q11–q21 controlling susceptibility to tuberculosis and leprosy in Brazilians. Genes Immun 5:46–57

Janulionis E, Sofer C, Schwander SK et al 2005 Survival and replication of clinical mycobacterium tuberculosis isolates in the context of human innate immunity. Infect Immun 2595–2601

Kris-Etherton PM, Harris WS, Appel LJ 2002 Fish consumption, fish oil, omega-3 fatty acids and cardiovascular disease. Circulation 106:2747–2757

Lombard Z, Brune AE, Hoal EG et al 2006 HLA class II disease associations in southern Africa. Tissue Antigens 67:97–110

Malik S, Abel L, Tooker H et al 2005 Alleles of the NRAMP1 gene are risk factors for pediatric tuberculosis disease. PNAS 34:12183–12188

Manca C, Tsenova L, Bergtold A et al 2001 Virulence of a Mycobacterium tuberculosis clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha/beta. Proc Natl Acad Sci USA 98:5752–5757

Newport MJ, Awomoyi AA, Blackwell JM 2003 Polymorphism in the interferon-gamma receptor-1 gene and susceptibility to pulmonary tuberculosis in The Gambia. Scand J Immunol 58:383–385

Paul KP, Leichsenring M, Pfisterer M et al 1997 Influence of n-6 and n-3 polyunsaturated fatty acids on the resistance to experimental tuberculosis. Metabolism 46:619–624

Remus N, El Baghdadi J, Fieschi C et al 2004 Association of IL12RB1 polymorphisms with pulmonary tuberculosis in adults in Morocco. J Infect Dis 190:580–587

Roberts T, Herselman M, Marais D, Labadarios D 2005 Served versus actual nutrient intake of hospitalised patients with tuberculosis as compared with energy and nutrient requirements. South African Journal of Clinical Nutrition 18:78–93

Rook GA, Hernandez-Pando R 1997 Pathogenetic role, in human and murine tuberculosis, of changes in the peripheral metabolism of glucocorticoids and antiglucocorticoids. Psychoneuroendocrinology 22 Suppl 1:S109–113

Rook GAW, Dheda K, Zumla A 2005 Immune responses to tuberculosis in developing countries: implications for new vaccines. Nat Rev Immunol 5:661–667

Rossouw M, Nel HJ, Cooke GS, van Helden PD, Hoal EG 2003 Association between tuberculosis and a polymorphic NF-κB binding site in the interferon γ gene. Lancet 361: 1871–1872

Roth DE, Soto G, Arenas F et al 2004 Association between vitamin D receptor gene polymorphisms and response to treatment of pulmonary tuberculosis. J Infect Dis 190:920–927

Selvaraj P, Chandra G, Jawahar MS, Vadya Rani M, Nisha Rajeshwari D, Narayanan PR 2004 Regulatory role of vitamin D receptor gene variants of BsmI, ApaI, TaqI, and FokI polymorphisms on macrophage phagocytosis and lymphoproliferative response to mycobacterium tuberculosis antigen in pulmonary tuberculosis. J Clin Immunol 24:523–532

Tsolaki AG, Hirsch AE, DeRiemer K et al 2004 Functional and evolutionary genomics of Mycobacterium tuberculosis: Insights from genomic deletions in 100 strains. Proc Natl Acad Sci USA 101:4865–4870

Van Rie A, Warren RM, Richardson M et al 1999 Exogenous reinfection is a common cause of tuberculosis recurrence after cure. New Eng J Medicine 341:1174–1179

Verver S, Warren RM, Beyers N et al 2005 Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. Am J Respir Crit Care Med 171: 1430–1435

Warren RM, Victor TC, Streicher EM et al 2004 Patients with active tuberculosis often have different strains in the same sputum specimen. Am J Respir Crit Care Med 169:610–614

Wiid I, Seaman T, Hoal EG, Benade AJS, Paul D van Helden 2004 Total antioxidant levels are low during active TB and rise with anti-tuberculosis therapy. IUBMB Life 56:101–106

Wilkinson RJ, Llewelyn M, Toossi Z et al 2000 Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London: a case-control study. Lancet 355:618–621

DISCUSSION

Segal: What is known about the cellular molecular mechanisms by which the TB bacillus is killed?

van Helden: I don't know where to start in answering this question. For example, the whole notion of nitric oxide (NO) is controversial. Acidification of the phagosome is important, and *Mtb* stops that. Your question to me addresses something I have often said: we shouldn't be looking at TB patients but rather those people who are infected and don't become ill.

Segal: This is the key for understanding the effects of nutrition and IFNγ and so on. We have been doing a study on Crohn's disease, which is similar in that it is often a granulomatous lesion. We did a range of studies in humans rather than in models. We did two things relevant to TB. First, we made skin windows. Gordon Brown had one of these on his arm. You sandpaper a bit of skin off and see what cells come out. You can also apply things onto the skin and see how that can modulate what happens. In Crohn's we found there is a major defect of the accumulation of neutrophils at skin windows. We believe this to be the primary lesion. Second, we injected killed bacteria to see what effect this had on the acute inflammatory response at the site of infection. We used *Escherichia coli*, but you could use TB that had been irradiated. You might find big differences. Many of the immunological effects people are looking at in these diseases are secondary effects to the failure to clear antigen, for example. Then it is assumed that they have some causal relationship to the disease when actually they don't.

van Helden: In TB there are so many different kinds of responses. The first category of infected individuals don't even develop a granuloma and the bacilli are killed. In the next category the granuloma will form, but the bacilli will remain latent for the rest of the person's life. In the third category, the primary focus will form, and there will then be secondary foci and active disease.

Segal: In the first group, I guess one would say that the organism has been killed, digested and removed. In the second, it has been killed but not digested. In

the third it hasn't been killed at all. It would be nice to translate that into actual measurements, either in the body or test-tube.

Quesniaux: When patients are treated with anti-TNF antibodies, one of the adverse effects is the occurrence of infections (Mohan et al 2003) and TB is one of the most prominent of these (Keane et al 2001, Mohan et al 2004, Keane 2005). Here we are really acting on the second category: reactivation of previous infections.

van Helden: Perhaps the best evidence for reactivation disease was work done by Troels Lillebaek (Lillebaek et al 2002). The son developed TB 33 years after the father had had a case. It transpired that they both had exactly the same isolate and no one else in that country had the same isolate. There is no way he could have got it from anyone else, indicating that the TB is held inactivated but not dead.

Quesniaux: TB is kept under active control from the immune response.

Brown: I'd like to go back to the point Tony Segal raised about the molecular mechanism of killing. One of the big holes in TB research concerns the receptors that are involved in uptake, and how this influences the resultant response. Many of the proposed receptors don't seem to fit. For example, CR3 is not expressed in alveolar macrophages. And a recent paper shows that in a normal, uninfected individual DC-SIGN is not expressed in macrophages—it is only induced after infection (Tailleux 2005).

Quesniaux: Are there differences in DC-SIGN expression after infection by different strains of TB?

van Helden: We've never looked.

Speert: I agree completely about CR3 not being expressed in alveolar macrophages. My understanding is that cigarette smoking does up-regulate CR3. Could this be a way of tying together some of the things that we have heard? I have a comment about babies infected with BCG. The ones that get seriously ill and die are the ones with profound immunodeficiency. The other group where BCG immunization is contraindicated is chronic granulomatous disease. Reactive oxygen radicals appear to play some role in protection against tuberculosis.

van Helden: We have now found plenty of HIV-positive children with BCGosis.

Mantovani: I was intrigued by the dietary data that you showed. I remember in my medical education that the suggestion was to overfeed patients with TB. The numbers presented here were low. Were these data borne out by larger numbers?

van Helden: The hospital is very small, and there are no other data. This is a major hole in our research. The other problem with TB patients is that they don't feel well, so they don't eat well. If you measure anything in a newly diagnosed TB patient, it may not be trustworthy, because they haven't been eating well because they haven't been feeling well. But in this country, in general, they are also poor,

so they may be malnourished from poverty as well. The work done in mice involves much larger numbers and this is well substantiated. In humans nutritional work is poor.

Mantovani: One of the genetic associations you mentioned is with the IL1RA. Is this with infection or disease manifestations? In other infectious disorders, such as *H. pylori* gastritis, the manifestation is not really related to the IL1RA genotype but to the haplotype, and the balance between the pro and anti-inflammatory stimulant molecules in the IL1 system. Are there any data along these lines?

Hoal: There have been a couple of studies on IL1RA (Bellamy et al 1998, Wilkinson et al 1999). As with a number of association studies, they show slightly different things. Some show the haplotype to be associated but some show there is an effect just with the polymorphism.

Wilkinson: We showed that an extended haplotype was associated with higher delayed type hypersensitivity (DTH). The low producing IL1RA allele linked to the high-producing IL1β associated with DTH (Wilkinson et al 1999).

Bekker: We have been doing some work in a small impoverished community with high HIV and TB rates in Cape Town, looking at Mtb strains in patients with and without HIV. We are seeing W Beijing strain much more in young people who are HIV infected. Has anyone looked at the immune competency of people infected with the W Beijing strain?

van Helden: I would predict that the Beijing in your younger age group is a reflection of an emerging strain. It has been relatively recently introduced into our society. Because it is a more able strain it spreads faster. At a Novartis Foundation Symposium in 1997 I proposed a square box scheme, where I hypothesized that you regard your strains as being represented in different parts of this box. They spread well but don't cause much disease, or they cause lots of disease but don't necessarily spread well (van Helden 1998). Beijing is going to be a position in the box. I think it is a very successful emerging strain that is spreading quickly.

Bekker: The counter to that is a set of recent data again from the community we have investigated in the south of Cape Town. In a cross sectional prevalence study in the community where we have sampled randomly, 10% of the community, it seems that people who have HIV and TB co-infection **are** presenting and **are** in the TB clinic and on treatment. People who are sitting quietly with their disease as yet undiagnosed are the older population who are not HIV infected. Immune competent people live with their disease for a long time in this population, before symptoms push them to seek healthcare, and I guess are able to spread their disease. It may be that W Beijing is the predominant strain in this scenario. The immune incompetent (HIV infected) people pick this particular strain up easily; they are the 'canaries' in the population.

van Helden: The question that we have asked is whether Beijing can go into a latency phase. I don't know. This would be interesting to address.

Wilkinson: We have looked at the proportion of strains that are Beijing in isolates from Red Cross Childrens' Hospital in Cape Town. Even within the last few years the proportion of Beijing has tended to increase. It is difficult to know what the denominator is because this is a hospital rather than a community. The second thing is that some of these strains won't grow very well in the laboratory: they appear to have a resuscitation defect. We speculate that if these strains are unable to resuscitate *in vivo*, teleologically speaking they must continue reinfecting people and therefore do best in environments where there is intense transmission.

Mizrahi: What we see depends crucially on what we can culture. Our tools are blunt, with culturability of *Mtb* strains being one of the limiting factors. The standard genotyping tool, which is based on an insertion element, is also relatively crude. However, with the introduction of higher-resolution tools such as DNA microarrays (Tsolaki et al 2004), we are now in a position to start looking at the association between the ability to reactivate and the genotype of the strain. These are studies that need to be done.

Schoub: I want to comment on the interaction between viral infections and TB. We have tried to mine data from the South African National Health Laboratory Service databank to look at seasonality, and correlate it with, for example, reactivation disease. One can speculate on the mechanism but do you have any comment on the actual observation?

van Helden: We don't have any data.

McGreal: What do we actually know at a structural and functional level when we talk about strain difference in *Mtb*? And also at a functional level what impact does strain difference have in terms of the immune response and possible immune evasion strategies. You indicated some associations with different strains, but how much do we know about that?

van Helden: Very little. We are only now developing microarray technology to answer those questions. Part of the problem is choosing the genome to base the microarray on. If you take your standard H37Rv, it has lots of deletions compared to many of our clinical strains. We need to take multiple strains and fully sequence them, and there are now five *Mtb* genomes available. The Broad Institute in Cambridge, MA, may sequence another 10. We should develop a microarray based on everything we can find.

McGreal: When you said that certain strains are prevalent in certain areas and populations, is that a result of local susceptibility? If you were to take a different genetic scheme would you see the same susceptibility?

van Helden: That's something we want to look at.

Mizrahi: It is important to mention that certain studies have revealed that the associations between strain families of *Mtb* and their human host populations are stable (Hirsh et al 2004, Baker et al 2004). This finding speaks to the issue of TB being an ancestral pathogen of humankind. The associations deduced from studies

of tuberculosis in immigrant populations imply that particular lineages of *Mtb* may be highly adapted to certain human populations, and less well adapted to others. This notion is supported by the results of a new and larger study published recently by Peter Small's group (Gagneux et al 2006).

van Helden: On the other hand it is to some extent because the immigrant populations stick together. It is a big confound.

Steinman: Can we have a little more discussion about DC-SIGN? First of all, we find it difficult to detect on most human blood leukocytes. Does anyone have experience otherwise and does it change in patients? We find that DC-SIGN is expressed on the small subset of myeloid dendritic cells (DCs) in blood following culture in IL4 (Granelli-Piperno et al 2006). It would be striking if you suddenly saw DC-SIGN increase in TB, since we are all influenced by the papers (Geijtenbeek et al 2003) saying that ligation of DC-SIGN by lipoarabinans from mycobacteria can block the maturation of DCs. The second thing is, DC-SIGN is seen on the monocyte-derived DCs. This is where it was described, and this seems to be due to induction by IL4. If you add IL4 to a monocyte, in a day you will have lots of DC-SIGN. It is not yet clear what monocyte-derived DCs correspond to *in vivo*. The third thing is what happens in the lymphoid organ. We have found that DC-SIGN is abundant on the macrophages in the lymph node medulla. It is not detectable among most DCs in the T cell area in apparently normal lymph nodes. Because of what has been shown *in vitro* with DC-SIGN we should look at it much more assiduously in disease states.

Lambrecht: We have looked at lung DCs from human lavage samples and DC-SIGN is not found on these either.

Gordon: There are other mannose recognition lectins on macrophages.

Brown: The ability of DC-SIGN to modulate the DC function is very likely. We have shown that signalling through another lectin, Dectin-1 can recruit Syk resulting in the induction of IL2 and IL10.

Steinman: That was the message of Geijtenbeek et al (2003). However as mentioned, the sites of DC-SIGN expression *in vivo* are not clear. We are all assuming that it is present on all DCs and exclusively DCs, but neither seems to be the case.

Brown: A lot of these DC receptors are not DC specific. This is another growing theme.

van Helden: How did you look for the expression?

Steinman: By monoclonal antibodies.

Gordon: What about the other genetic factors? You went over them quite quickly as though none of them really mattered, or all of them mattered a little.

van Helden: I don't think anyone has found a major genetic factor.

Mayosi: Do we know the population-attributable risk for any of the genetic factors?

Hoal: It is always small. This is the issue with genetic studies of infectious diseases or other complex diseases. We will find a number of genes that are important but only have a small contribution. I don't know whether you could have the situation where someone's susceptibility rises significantly if they have a combination of a number of these susceptibility alleles. We just don't know that yet. The field is still in the phase of finding the genes, replicating them in a second population, and working out what sort of significance they have.

Gordon: Do you think it is fair to talk about the Bronte family and genetics? How do you separate out the environmental factors and the genetic ones?

van Helden: Apparently, in the Bronte family the father had chronic TB. He died in his 80s, but I think he infected the rest and they all died young.

Gordon: Why does this indicate genetic susceptibility?

van Helden: Other families have a case but not all succumb to disease.

Lambrecht: One clinical problem we run into occasionally is elderly women with atypical mycobacteria in the middle lobe. People always claimed that there were anatomical problems with the middle lobe, but isn't there also a genetic predisposition to develop atypical mycobacterial disease in the lungs?

van Helden: I have no idea. The problem with that sort of thing is power. We looked at our first 2000 isolates, and out of these we found 60 that were attributable to MOTTs. Only now have we put some effort into trying to find this out. But if we only have 60 cases out of 2000, we don't have enough genetic power to really examine this.

Lambrecht: It would be easy to do a candidate gene approach with 60 patients.

van Helden: The populational attributable risk is quite small per gene. It won't be a monogenic effect like the IFNγ receptor defects in the Maltese kindred, for example. You could be right, but I doubt it.

Wilkinson: There was a study done by the British Thoracic Society that looked at susceptibility to *Mycobacterium malmoense* and this showed a weak association with VDR promoter polymorphisms (Gelder et al 2000). However, the predominant factors that predispose people to atypical mycobacterioses are damaged lungs and immunosuppression.

Finn: I know that in this meeting we are focused on the innate immune system. You mentioned several times that this genetic difference may signify that someone's innate immunity is stronger than other people's. One of the important roles of the innate immune system is to jump-start the adaptive immune system and generate good memory responses that will protect you for the rest of your life. The question is, how good is the memory response to the bug in those people who do not get reinfected and those who are protected under high risk conditions? And what types of effector mechanisms are involved that would instruct us in terms of generating that type of immunity through vaccination to protect the rest of the population?

van Helden: This is an important issue, but we haven't done any work on this. There is old work which suggests that the partial protective effect of infection varies between 16 and 18%. Unfortunately, I am not qualified to comment directly on how good that work is. It is based on mathematical modelling and epidemiology. I am sceptical about the interpretation. The tool that is used is generally the skin test, which is extremely blunt. Half of the T cell researchers would say that there is partial immunity. Then there are others who say that it is impossible, because reinfection is occurring and if you are reinfected you have a four times higher likelihood to progress to disease, so where is your partial immunity. To reconcile these two schools of thought I would like to suggest that you increase your categories of individuals. You have categories of individuals that will develop partial immunity and another category that won't.

Finn: Among those that develop partial immunity there will be some that will still be susceptible.

Brown: Is it going to be possible to make vaccine for TB, given the occurrence of reinfection?

van Helden: Many think so, since it could be cell-mediated immunity.

Finn: With regard to the comment you made about a low dose of BCG maintaining better protection, to an immunologist this immediately says that it generates higher affinity and avidity T cells and so on.

van Helden: The New Zealanders say that the dose and timing of repeat BCG is critical. The Irish say that a high-dose BCG vaccine given once is just as good.

Walzl: I can't think of any pathogen where a vaccine provides better protection than the natural infection. But there is always a first time.

Steinman: Cowpox isn't natural for humans, but it worked as a smallpox vaccine. Immunologists just haven't got onto the vaccine scene, but when we do, I think that we can do better than complex microbial vaccines!

Finn: As you mentioned, some pathogens have evolved a symbiosis with the host, and a vaccine might do a better job eliciting immunity than the bugs.

Ryffel: Going back to Valerie's comment on the anti-TNF antibody induced reactivation of chronic/latent infection, we are able to model reactivation or tuberculosis infection in mice (Botha & Ryffel 2003). By administration of neutralizing TNF antibodies or soluble TNF-R we are able to reactivate chronic infection (unpublished). Further we demonstrated that membrane TNF provides a partial protection to infection (Fremond et al 2005) suggesting that neutralising exclusively soluble but not membrane TNF may reduce the risk of reactivation of TB infection.

Quesniaux: An individual who has had a first infection with TB and has had this under control for 20 years will have mounted a very efficient T cell memory response. If you now come along with anti-TNF, this could all go wrong with the TB being reactivated.

Finn: I disagree with your statement that this person has mounted a good memory response. This is an effector memory-type response, something that is continuously protective. It is not that good deep central memory that results from eliminating the pathogen. In the complete absence of antigen you deeply bury your central memory response that can then be reactivated on subsequent infection. You cannot say that reactivation of the bug by TNF means that it has now defeated a well established memory response.

Quesniaux: How do we explain the effect of anti-TNF?

Steinman: TNF has many effects on DCs. In the literature on NOD mice, people have tried to manipulate the onset of that spontaneous autoimmune disease by manipulating TNF levels. There is now evidence that the DC is responsible. The approach is to block TNF early in life, which then reduces the severity of the disease. If you then target antigens to the DCs in the TNF blocked mice, you can see changes in antigen presentation. Therefore, the DCs may be changing as a result of TNF manipulation.

Finn: There is constant cross-talk between the innate and the adaptive immune system. It doesn't just go from the innate to the adaptive and then stop, it goes back as well. You have a wonderfully stimulated memory CD4 response specific for the pathogen. You can arm your macrophage that is infected with a pathogen to make more IFNγ by the interaction with the $CD4^+$ T cell. If you have a good adaptive immune memory, it can then stimulate much higher activity of your innate system when the new infection comes. If you have a very good $CD4^+$ T cell memory response, that macrophage will be much more effective in destroying the pathogen.

van Helden: In thinking about vaccination and protective immunity, we have to remember that there is evidence that BCG does work. My feeling is that if there is no immunosuppression, 90–95% of people infected with TB will not become ill. BCG is not going to do anything for them anyway. It is a small part of the population in which vaccination can have an effect.

Finn: I would like to argue against this point. The 95% of the people will not develop disease unless they are immunosuppressed in some way, unless they live a long life and their immune systems become old. But if you generate a strong immune memory early in life that immune memory is a bit more protective from what happens to that person later in life, who therefore, if healthy and well-nourished will not be susceptible to the disease. Generating a strong and effective immune memory early in life will protect many more people than the 5% that you think will be protected by the vaccine.

Steinman: The point is how you assay memory. This is an evolving field. We used to say that HIV-infected people have good memory, because you could easily detect $CD8^+$ T cells in them that make IFNγ in response to HIV antigens. By current criteria, though, these CD8 cells are dysfunctional. They don't grow in response

to antigen. From the same individuals, CMV, EBV and influenza specific cells grow fine (Arrode et al 2005). There is a dysfunction here of the HIV specific $CD8^+$ T cells, and they are termed 'helpless' memory cells. How we assay memory is very important.

Finn: The bottom line is that we know a lot more than last year and 10 years ago. What we know speaks much more in favour of the vaccine being able to do this better than the natural disease.

Steinman: Then there are suppressor cells. What you described in terms of reinfection would fit what we know about suppressor cells. These cells are triggered by antigen, but then they can suppress other immune responses to other antigens, particularly when they are presented by the same presenting cell. If you respond to mycobacteria A and make a suppressor cell specific for peptides from protein A, it will block the immune response to other mycobacteria proteins, B. That is, as long as the antigen presenting cells are presenting both A and B, the suppressor cells for A will block presentation of B. However, it is still not straightforward to measure suppressor cells currently in human. This is a big gap.

Hoal: I want to return to the vaccination and susceptibility story. This is mouse work that I don't know very well, but I heard that some mouse strains that are more susceptible to mycobacteria are also less likely to be protected by vaccination. This is quite worrying if the same were to occur in the human population. We have to be careful when we do vaccination trials in humans to look at this susceptible portion of the population.

Gordon: Bernard Ryffel, do the mouse models allow you to get at some of these questions?

Ryffel: It is well established that the susceptibility to TB infection differs among mouse strains, B6 mice are typically resistant, while Balb/c mice are more susceptible; the genomic analysis will certainly provide polygenic resistance loci, which are currently unknown. What is really amazing is that mice with complete ablation of Toll-like receptor/MyD88 signalling are able to have preserved T cell response to TB antigens. But this T cell response is not protective, as the innate immune response is profoundly defective (Fremond et al 2004). However, BCG vaccination of MyD88 deficient mice provides a short-term, but not long-term protection.

Walzl: There are mouse models concerning regulatory cells. Strangely enough, people examined induction of regulatory cells by *Mtb* in conjunction with the hygiene hypothesis and the allergic response. There are several studies showing *Mtb* in circulatory cells that suppress allergic airway responses. It is strange that people look at it that way; the logical conclusion would be that those responses are there because *Mtb* wants to promote itself.

Lambrecht: This could explain why the low dose BCG works better than the high dose. If you give high doses of freeze-inactivated BCG it is a strong inducer of T_{regs}. There are moieties within the BCG which have the ability to induce

T_{regs}. If they are to the advantage of the bug, this would explain why high levels of BCG would lead to greater susceptibility. It is a balance between protecting from an over-zealous response and on the other side inducing immunity.

E Sim: Am I right in thinking that you can distinguish between reactivation and reinfection by the same strain?

van Helden: No, we can't.

Ryffel: Do you have more evidence that there is coinfection? If this is the case, with two strains, what is the prevalence? What is the contribution of each?

van Helden: We have no quantitative data.

Wilkinson: There is an interesting recent paper showing three cases of TB, all of which are pulmonary, and have another extrapulmonary site (Garcia de Viedma et al 2005). For each of the three extrapulmonary sites there is a different strain. The authors suggest that the extrapulmonary strain has greater virulence than the pulmonary strains when they infect macrophages together.

References

Arrode G, Finke JS, Zebroski H, Siegal FP, Steinman RM 2005 CD8+ T cells from most HIV-1 infected patients, even when challenged with mature dendritic cells, lack functional recall memory to HIV gag but not other viruses. Eur J Immunol 35:159–170

Baker L, Brown T, Maiden MC, Drobniewski F 2004 Silent nucleotide polymorphisms and a phylogeny for Mycobacterium tuberculosis. Emerg Infect Dis 10:1568–1577

Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV 1998 Assessment of the interleukin 1 gene cluster and other candidate gene polymorphisms in host susceptibility to tuberculosis. Tuber Lung Dis 79:83–89

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

Fremond CM, Yeremeev V, Nicolle DM, Jacobs M, Quesniaux VF, Ryffel B 2004 Fatal Mycobacterium tuberculosis infection despite adaptive immune response in the absence of MyD88. J Clin Invest 114:1790–1799

Fremond C, Allie N, Dambuza I et al 2005 Membrane TNF confers protection to acute mycobacterial infection. Respir Res 6:136

Gagneux S, DeRiemer K, Van T, Kato-Maeda M et al 2006 Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci USA 103:2869–2873

Garcia de Viedma D, Lorenzo G, Cardona PJ et al 2005 Association between the infectivity of Mycobacterium tuberculosis strains and their efficiency for extrarespiratory infection. J Infect Dis 192:2059–2065

Geijtenbeek TB, Van Vliet SJ, Koppel EA et al 2003 Mycobacteria target DC-SIGN to suppress dendritic cell function. J Exp Med 197:7–17

Gelder CM, Hart KW, Williams OM et al 2000 Vitamin D receptor gene polymorphisms and susceptibility to Mycobacterium malmoense pulmonary disease. J Infect Dis 181:2099–2102

Granelli-Piperno A, Shimeliovich I, Pack M, Trumpfheller C, Steinman RM 2006 HIV-1 selectively infects a subset of nonmaturing BDCA1-positive dendritic cells in human blood. J Immunol 176:991–998

Hirsh AE, Tsolaki AG, DeRiemer K, Feldman MW, Small PM 2004 Stable association between strains of Mycobacterium tuberculosis and their human host populations. Proc Natl Acad Sci USA 101:4871–4876

Keane J 2005 TNF-blocking agents and tuberculosis: new drugs illuminate an old topic. Rheumatology (Oxford) 44:714–720

Keane J, Gershon S, Wise RP et al 2001 Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. N Engl J Med 345:1098–1104

Lillebaek T, Dirksen A, Baess I, Strunge B, Thomsen VO, Andersen AB 2002 Molecular evidence of endogenous reactivation of Mycobacterium tuberculosis after 33 years of latent infection. J Infect Dis 185:401–404

Mohan AK, Cote TR, Siegel JN, Braun MM 2003 Infectious complications of biologic treatments of rheumatoid arthritis. Curr Opin Rheumatol 15:179–184

Mohan AK, Cote TR, Block JA, Manadan AM, Siegel JN, Braun MM 2004 Tuberculosis following the use of etanercept, a tumor necrosis factor inhibitor. Clin Infect Dis 39:295–299

Tailleux L, Pham-Thi N, Bergeron-Lafaurie A et al 2005 DC-SIGN induction in alveolar macrophages defines privileged target host cells for mycobacteria in patients with tuberculosis. PLoS Med 2:e381

Tsolaki AG, Hirsh AE, DeRiemer K et al 2004 Functional and evolutionary genomics of Mycobacterium tuberculosis: insights from genomic deletions in 100 strains. Proc Natl Acad Sci USA 101:4721–4722

van Helden PD 1998 Bacterial genetics and strain variation. In: Genetics and tuberculosis. Wiley, Chichester (Novartis Found Symp 217) p 178–194

Wilkinson RJ, Patel P, Llewelyn M et al 1999 Influence of polymorphism in the genes for the Interleukin 1 Receptor Antagonist and Interleukin IL-1β on tuberculosis. J Exp Med 189:1863–1874

Reference List

1. (2006) Mid-year population estimates, South Africa. Statistics South Africa, Pretoria,

2. Abe T, Iinuma Y, Ando M, Yokoyama T, Yamamoto T, Nakashima K, Takagi N, Baba H, Hasegawa Y, Shimokata K (2003) NRAMP1 polymorphisms, susceptibility and clinical features of tuberculosis. J Infect 46:215-220

3. Abecasis GR, Cookson WOC (2000) GOLD - Graphical Overview of Linkage Disequilibrium. Bioinformatics 16:182-183

4. Abecasis GR, Ghosh D, Nichols TE (2005) Linkage disequilibrium: ancient history drives the new genetics. Hum Hered 59:118-124

5. Abel L, Casanova JL (2000) Genetic predisposition to clinical tuberculosis: bridging the gap between simple and complex inheritance. Am J Hum Genet 67:274-277

6. Adams CH, Werely CJ, Victor TC, Hoal EG, Rossouw G, van Helden PD (2003) Allele frequencies for glutathione S-transferase and N-acetyltransferase 2 differ in African population groups and may be associated with oesophageal cancer or tuberculosis incidence. Clin Chem Lab Med 41:600-605

7. Algood HM, Flynn JL (2004) CCR5-deficient mice control Mycobacterium tuberculosis infection despite increased pulmonary lymphocytic infiltration. J Immunol 173:3287-3296

8. Alvarez CP, Lasala F, Carrillo J, Muniz O, Corbi AL, Delgado R (2002) C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans. J Virol 76:6841-6844

9. Andersen P, Doherty TM (2005) The success and failure of BCG - implications for a novel tuberculosis vaccine. Nat Rev Microbiol 3:656-662

10. Andersson DI, Levin BR (1999) The biological cost of antibiotic resistance. Curr Opin Microbiol 2:489-493

11. Anzala AO, Ball TB, Rostron T, O'Brien SJ, Plummer FA, Rowland-Jones SL (1998) CCR2-64I allele and genotype association with delayed AIDS progression in African women. Lancet 351:1632-1633

12. Appelmelk BJ, van D, I, van Vliet SJ, Vandenbroucke-Grauls CM, Geijtenbeek TB, van Kooyk Y (2003) Cutting edge: carbohydrate profiling identifies new pathogens that interact with dendritic cell-specific ICAM-3-grabbing nonintegrin on dendritic cells. J Immunol 170:1635-1639

13. Arnadottir T, Rieder HL, Trebucq A, Waaler HT (1996) Guidelines for conducting tuberculin skin test surveys in high prevalence countries. Tuber Lung Dis 77 Suppl 1:1-19

14. Awomoyi AA, Marchant A, Howson JM, McAdam KP, Blackwell JM, Newport MJ (2002) Interleukin-10, polymorphism in SLC11A1 (formerly NRAMP1), and susceptibility to tuberculosis. J Infect Dis 186:1808-1814

15. Babb C, Keet EH, van Helden PD, Hoal EG (2007) SP110 polymorphisms are not associated with pulmonary tuberculosis in a South African population. Hum Genet 121:521-522

16. Babb C, van der Merwe L, Beyers N, Pheiffer C, Walzl G, Duncan K, van Helden P, Hoal EG (2007) Vitamin D receptor gene polymorphisms and sputum conversion time in pulmonary tuberculosis patients. Tuberculosis 87:295-302

17. Baghdadi JE, Orlova M, Alter A, Ranque B, Chentoufi M, Lazrak F, Archane MI, Casanova JL, Benslimane A, Schurr E, Abel L (2006) An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. J Exp Med 203:1679-1684

18. Balding DJ (2006) A tutorial on statistical methods for population association studies. Nat Rev Genet 7:781-791

19. Bamshad M, Wooding S, Salisbury BA, Stephens JC (2004) Deconstructing the relationship between genetics and race. Nat Rev Genet 5:598-609

20. Barreiro LB, Patin E, Neyrolles O, Cann HM, Gicquel B, Quintana-Murci L (2005) The heritage of pathogen pressures and ancient demography in the human innate-immunity CD209/CD209L region. Am J Hum Genet 77:869-886

21. Barreiro LB, Neyrolles O, Babb CL, Tailleux L, Quach H, McElreavey K, Helden PD, Hoal EG, Gicquel B, Quintana-Murci L (2006) Promoter Variation in the DC-SIGN Encoding Gene CD209 Is Associated with Tuberculosis. PLoS Medicine 3:e20

22. Barreto ML, Pereira SM, Ferreira AA (2006) BCG vaccine: efficacy and indications for vaccination and revaccination. J Pediatr (Rio J) 82:S45-S54

23. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265

24. Bashirova AA, Geijtenbeek TB, van Duijnhoven GC, van Vliet SJ, Eilering JB, Martin MP, Wu L, Martin TD, Viebig N, Knolle PA, KewalRamani VN, van Kooyk Y, Carrington M (2001) A dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN)-related protein is highly expressed on human liver sinusoidal endothelial cells and promotes HIV-1 infection. J Exp Med 193:671-678

25. Bashirova AA, Wu L, Cheng J, Martin TD, Martin MP, Benveniste RE, Lifson JD, KewalRamani VN, Hughes A, Carrington M (2003) Novel member of the CD209 (DC-SIGN) gene family in primates. Journal of Virology217

26. Bellamy R (2003) Susceptibility to mycobacterial infections: the importance of host genetics. Genes Immun 4:4-11

27. Bellamy R, Beyers N, McAdam KP, Ruwende C, Gie R, Samaai P, Bester D, Meyer M, Corrah T, Collin M, Camidge DR, Wilkinson D, Hoal-Van Helden E, Whittle HC, Amos W, van Helden P, Hill AV (2000) Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. Proc Natl Acad Sci U S A 97:8005-8009

28. Bellamy R, Ruwende C, Corrah T, McAdam KP, Thursz M, Whittle HC, Hill AV (1999) Tuberculosis and chronic hepatitis B virus infection in Africans and variation in the vitamin D receptor gene. J Infect Dis 179:721-724

29. Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV (1998) Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. N Engl J Med 338:640-644

30. Bergman MP, Engering A, Smits HH, van Vliet SJ, van Bodegraven AA, Wirth HP, Kapsenberg ML, Vandenbroucke-Grauls CM, van Kooyk Y, Appelmelk BJ (2004) Helicobacter pylori modulates the T helper cell 1/T helper cell 2 balance through phase-variable interaction between lipopolysaccharide and DC-SIGN. J Exp Med 200:979-990

31. Beyers N, Gie RP, Schaaf HS, Van Zyl S, Talent JM, Nel ED, Donald PR (1997) A prospective evaluation of children under the age of 5 years living in the same household as adults with recently diagnosed pulmonary tuberculosis. Int J Tuberc Lung Dis 1:38-43

32. Beyers N, Gie RP, Zietsman HL, Kunneke M, Hauman J, Tatley M, Donald PR (1996) The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high-incidence community. S Afr Med J 86:40-1, 44

33. Bifani PJ, Plikaytis BB, Kapur V, Stockbauer K, Pan X, Lutfey ML, Moghazeh SL, Eisner W, Daniel TM, Kaplan MH, Crawford JT, Musser JM, Kreiswirth BN (1996) Origin and interstate spread of a New York City multidrug-resistant Mycobacterium tuberculosis clone family. JAMA 275:452-457

34. Blackwell JM (2001) Genetics and genomics in infectious disease susceptibility. Trends Mol Med 7:521-526

35. Bloch DB, Nakajima A, Gulick T, Chiche JD, Orth D, de La Monte SM, Bloch KD (2000) Sp110 localizes to the PML-Sp100 nuclear body and may function as a nuclear hormone receptor transcriptional coactivator. Mol Cell Biol 20:6138-6146

36. Bloom BR, Fine PE (1994) The BCG experience: Implications for future vaccines against Tuberculosis. In: Bloom BR (ed) Tuberculosis: Pathogenesis, Protection, and Control. ASM Press, Washington, DC, pp 531-557

37. Bornman L, Campbell SJ, Fielding K, Bah B, Sillah J, Gustafson P, Manneh K, Lisse I, Allen A, Sirugo G, Sylla A, Aaby P, McAdam KP, Bah-Sow O, Bennett S, Lienhardt C, Hill AV (2004) Vitamin D receptor polymorphisms and susceptibility to tuberculosis in West Africa: a case-control and family study. J Infect Dis 190:1631-1641

38. Bottomley MJ, Collard MW, Huggenvik JI, Liu Z, Gibson TJ, Sattler M (2001) The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. Nat Struct Biol 8:626-633

39. Brahmbhatt S, Black GF, Carroll NM, Beyers N, Salker F, Kidd M, Lukey PT, Duncan K, van Helden P, Walzl G (2006) Immune markers measured before treatment predict outcome of intensive phase tuberculosis therapy. Clin Exp Immunol 146:243-252

40. Brandstatter A, Parsons TJ, Parson W (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. Int J Legal Med 117:291-298

41. Brodie MJ, Boobis AR, Hillyard CJ, Abeyasekera G, Stevenson JC, MacIntyre I, Park BK (1982) Effect of rifampicin and isoniazid on vitamin D metabolism. Clin Pharmacol Ther 32:525-530

42. Brody JR, Kern SE (2004) Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. Biotechniques 36:214-216

43. Burgner D, Jamieson SE, Blackwell JM (2006) Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? Lancet Infect Dis 6:653-663

44. Buyru N, Tezol A, Yosunkaya-Fenerci E, Dalay N (2003) Vitamin D receptor gene polymorphisms in breast cancer. Exp Mol Med 35:550-555

45. Cao K, Moormann AM, Lyke KE, Masaberg C, Sumba OP, Doumbo OK, Koech D, Lancaster A, Nelson M, Meyer D, Single R, Hartzman RJ, Plowe CV, Kazura J, Mann DL, Sztein MB, Thomson G, Fernandez-Vina MA (2004) Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. Tissue Antigens 63:293-325

46. Cardon LR, Abecasis GR (2003) Using haplotype blocks to map human complex trait loci. Trends Genet 19:135-140

47. Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. Lancet 361:598-604

48. Carrington M, Dean M, Martin MP, O'Brien SJ (1999) Genetics of HIV-1 infection: chemokine receptor CCR5 polymorphism and its consequences. Human Molecular Genetics 8:1939-1945

49. Casanova JL, Abel L (2002) Genetic Dissection of Immunity to Mycobacteria: The Human Model. Annu Rev Immunol 20:581-620

50. Cervino AC, Lakiss S, Sow O, Bellamy R, Beyers N, Hoal-Van Helden E, van Helden P, McAdam KP, Hill AV (2002) Fine mapping of a putative tuberculosis-susceptibility locus on chromosome 15q11-13 in African families. Hum Mol Genet 11:1599-1603

51. Chan VS, Chan KY, Chen Y, Poon LL, Cheung AN, Zheng B, Chan KH, Mak W, Ngan HY, Xu X, Screaton G, Tam PK, Austyn JM, Chan LC, Yip SP, Peiris M, Khoo US, Lin CL (2006) Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. Nat Genet 38:38-46

52. Chang TJ, Lei HH, Yeh JI, Chiu KC, Lee KC, Chen MC, Tai TY, Chuang LM (2000) Vitamin D receptor gene polymorphisms influence susceptibility to type 1 diabetes mellitus in the Taiwanese population. Clin Endocrinol (Oxf) 52:575-580

53. Chen HY, Chen WC, Hsu CD, Tsai FJ, Tsai CH (2002) Relation of vitamin D receptor FokI start codon polymorphism to bone mineral density and occurrence of osteoporosis in postmenopausal women in Taiwan. Acta Obstet Gynecol Scand 81:93-98

54. Chensue SW, Warmington KS, Allenspach EJ, Lu B, Gerard C, Kunkel SL, Lukacs NW (1999) Differential expression and cross-regulatory function of RANTES during mycobacterial (type 1) and schistosomal (type 2) antigen-elicited granulomatous inflammation. J Immunol 163:165-173

55. Cilliers SP (1963) The Coloureds of South Africa; A Factual Survey. Banier Publishers (Pty) Ltd, Cape Town

56. Classen CN, Warren R, Richardson M, Hauman JH, Gie RP, Ellis JH, van Helden PD, Beyers N (1999) Impact of social interactions in the community on the transmission of tuberculosis in a high incidence area. Thorax 54:136-140

57. Clegg AO, Ashton LJ, Biti RA, Badhwar P, Williamson P, Kaldor JM, Stewart GJ (2000) CCR5 promoter polymorphisms, CCR5 59029A and CCR5 59353C, are under represented in HIV-1-infected long-term non-progressors. AIDS 14:103-108

58. Colhoun HM, McKeigue PM, Davey SG (2003) Problems of reporting genetic associations with complex outcomes. Lancet 361:865-872

59. Collins HL, Kaufmann SHE (2001) The many faces of host responses to tuberculosis. Immunology 103:1-9

60. Colmenares M, Puig-Kroger A, Pello OM, Corbi AL, Rivas L (2002) Dendritic cell (DC)-specific intercellular adhesion molecule 3 (ICAM-3)-grabbing nonintegrin (DC-SIGN, CD209), a C-type surface lectin in human DCs, is a receptor for Leishmania amastigotes. J Biol Chem 277:36766-36769

61. Comstock GW (1978) Tuberculosis in twins: a re-analysis of the Prophit survey. Am Rev Respir Dis 117:621-624

62. Cooke GS, Hill AV (2001) Genetics of susceptibility to human infectious disease. Nat Rev Genet 2:967-977

63. Creery D, Weiss W, Lim WT, Aziz Z, Angel JB, Kumar A (2004) Down-regulation of CXCR-4 and CCR-5 expression by interferon-gamma is associated with inhibition of chemotaxis and human immunodeficiency virus (HIV) replication but not HIV entry into human monocytes. Clinical and Experimental Immunology 137:156-165

64. Crowle AJ, May MH (1990) Inhibition of tubercle bacilli in cultured human macrophages by chloroquine used alone and in combination with streptomycin, isoniazid, pyrazinamide, and two metabolites of vitamin D3. Antimicrob Agents Chemother 34:2217-2222

65. Curtis BM, Scharnowske S, Watson AJ (1992) Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. Proc Natl Acad Sci U S A 89:8356-8360

66. Daniel TM (2006) The history of tuberculosis. Respir Med 100:1862-1870

67. Dean M, Carrington M, O'Brien SJ (2002) Balanced polymorphism selected by genetic versus infectious human disease. Annual Review of Genomics and Human Genetics 3:263-292

68. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, Goedert JJ, Buchbinder SP, Vittinghoff E, Gomperts E, Donfield S, Vlahov D, Kaslow R, Saah A, Rinaldo C, Detels R, Obrien SJ (1996) Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Science 273:1856-1862

69. Delgado JC, Baena A, Thim S, Goldfeld AE (2002) Ethnic-specific genetic associations with pulmonary tuberculosis. J Infect Dis 186:1463-1468

70. Deluca HF, Cantorna MT (2001) Vitamin D: its role and uses in immunology. FASEB J 15:2579-2585

71. den Boon S, Van Lill SWP, Borgdorff MW, Verver S, Bateman ED, Lombard CJ, Enarson DA, Beyers N (2005) Association between smoking and tuberculosis infection: a population survey in a high tuberculosis incidence area. Thorax 60:555-557

72. den Boon S, Verver S, Marais BJ, Enarson DA, Lombard CJ, Bateman ED, Irusen E, Jithoo A, Gie RP, Borgdorff MW, Beyers N (2007) Association between passive smoking and infection with Mycobacterium tuberculosis in children. Pediatrics 119:734-739

73. Denis M (1991) Killing of Mycobacterium tuberculosis within human monocytes: activation by cytokines and calcitriol. Clin Exp Immunol 84:200-206

74. Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997-1004

75. Donald PR, Schoeman JF (2004) Tuberculous meningitis. N Engl J Med 351:1719-1720

76. Doolan DL, Southwood S, Chesnut R, Appella E, Gomez E, Richards A, Higashimoto YI, Maewal A, Sidney J, Gramzinski RA, Mason C, Koech D, Hoffman SL, Sette A (2000) HLA-DR-promiscuous T cell epitopes from Plasmodium falciparum pre-erythrocytic-stage antigens restricted by multiple HLA class II alleles. J Immunol 165:1123-1137

77. Du Toit ED, Emmanuel JC, West G, Taljaard DG, Oudshoorn M (1990) HLA Frequencies in Black Zimbabweans. Tissue Antigens 36:122-124

78. Du Toit ED, MacGregor KJ, Taljaard DG, Oudshoorn M (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

79. Duan HF, Zhou XH, Ma Y, Li CY, Chen XY, Gao WW, Zheng SH (2003) [A study on the association of 3'UTR polymorphisms of NRAMP1 gene with susceptibility to tuberculosis in Hans]. Zhonghua Jie He He Hu Xi Za Zhi 26:286-289

80. Ducati RG, Ruffino-Netto A, Basso LA, Santos DS (2006) The resumption of consumption-- a review on tuberculosis. Mem Inst Oswaldo Cruz 101:697-714

81. Dudbridge F (2003) Pedigree disequilibrium tests for multilocus haplotypes. Genet Epidemiol 25:115-121

82. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barmada MM, Rotter JI, Nicolae DL, Cho JH (2006) A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 314:1461-1463

83. Dusso AS, Brown AJ, Slatopolsky E (2005) Vitamin D. Am J Physiol Renal Physiol 289:F8-28

84. Dye C, Garnett GP, Sleeman K, Williams BG (1998) Prospects for worldwide tuberculosis control under the WHO DOTS strategy. Directly observed short-course therapy. Lancet 352:1886-1891

85. Easterbrook PJ, Rostron T, Ives N, Troop M, Gazzard BG, Rowland-Jones SL (1999) Chemokine receptor polymorphisms and human immunodeficiency virus disease progression. Journal of Infectious Diseases 180:1096-1105

86. Ebert R, Schutze N, Adamski J, Jakob F (2006) Vitamin D signaling is modulated on multiple levels in health and disease. Mol Cell Endocrinol

87. Feinberg H, Guo Y, Mitchell DA, Drickamer K, Weis WI (2005) Extended neck regions stabilize tetramers of the receptors DC-SIGN and DC-SIGNR. J Biol Chem 280:1327-1335

88. Ferguson RG (1934) The Indian Tuberculosis Problem And Some Preventitive Measures. The Canadian Medical Association Journal May:544-547

89. Ferrie RM, Schwarz MJ, Robertson NH, Vaudin S, Super M, Malone G, Little S (1992) Development, multiplexing, and application of ARMS tests for common mutations in the CFTR gene. Am J Hum Genet 51:251-262

90. Fewster ME, Kies B (1984) HLA antigens in multiple sclerosis in coloured South Africans. J Neurol Sci 66:175-181

91. Fine PE (2001) BCG: the challenge continues. Scand J Infect Dis 33:243-245

92. Fitness J, Floyd S, Warndorff DK, Sichali L, Malema S, Crampin AC, Fine PE, Hill AV (2004) Large-scale candidate gene study of tuberculosis susceptibility in the Karonga district of northern Malawi. Am J Trop Med Hyg 71:341-349

93. Floto RA, MacAry PA, Boname JM, Mien TS, Kampmann B, Hair JR, Huey OS, Houben EN, Pieters J, Day C, Oehlmann W, Singh M, Smith KG, Lehner PJ (2006) Dendritic cell stimulation by mycobacterial Hsp70 is mediated through CCR5. Science 314:454-458

94. Flynn JL, Chan J (2001) Immunology of tuberculosis. Annu Rev Immunol 19:93-129

95. Fraziano M, Cappelli G, Santucci M, Mariani F, Amicosante M, Casarini M, Giosue S, Bisetti A, Colizzi V (1999) Expression of CCR5 is increased in human monocyte-derived macrophages and alveolar macrophages in the course of in vivo and in vitro Mycobacterium tuberculosis infection. Aids Research and Human Retroviruses 15:869-874

96. Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D (2004) Assessing the impact of population stratification on genetic association studies. Nat Genet 36:388-393

97. Freeman BD, Buchman TG, McGrath S, Tabrizi AR, Zehnbauer BA (2002) Template-directed dye-terminator incorporation with fluorescence polarization detection for analysis of single nucleotide polymorphisms implicated in sepsis. J Mol Diagn 4:209-215

98. Frieden TR, Munsiff SS (2005) The DOTS strategy for controlling the global tuberculosis epidemic. Clin Chest Med 26:197-205, v

99. Frieden TR, Sterling TR, Munsiff SS, Watt CJ, Dye C (2003) Tuberculosis. Lancet 362:887-899

100. Gagneux S, Deriemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, Small PM (2006) Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 103:2869-2873

101. Gajalakshmi V, Peto R, Kanaka TS, Jha P (2003) Smoking and mortality from tuberculosis and other diseases in India: retrospective study of 43 000 adult male deaths and 35 000 controls. Lancet 362:507-515

102. Galvani AP, Slatkin M (2003) Evaluating plague and smallpox as historical selective pressures for the CCR5-Delta 32 HIV-resistance allele. Proceedings of the National Academy of Sciences of the United States of America 100:15276-15279

103. Gandhi NR, Moll A, Sturm AW, Pawinski R, Govender T, Lalloo U, Zeller K, Andrews J, Friedland G (2006) Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet 368:1575-1580

104. Gao PS, Fujishima S, Mao XQ, Remus N, Kanda M, Enomoto T, Dake Y, Bottini N, Tabuchi M, Hasegawa N, Yamaguchi K, Tiemessen C, Hopkin JM, Shirakawa T, Kishi F (2000) Genetic variants of NRAMP1 and active tuberculosis in Japanese populations. International Tuberculosis Genetics Team. Clin Genet 58:74-76

105. Geijtenbeek TB, Krooshoop DJ, Bleijs DA, van Vliet SJ, van Duijnhoven GC, Grabovsky V, Alon R, Figdor CG, van Kooyk Y (2000) DC-SIGN-ICAM-2 interaction mediates dendritic cell trafficking. Nat Immunol 1:353-357

106. Geijtenbeek TB, Krooshoop DJ, Bleijs DA, van Vliet SJ, van Duijnhoven GC, Grabovsky V, Alon R, Figdor CG, van Kooyk Y (2000) DC-SIGN-ICAM-2 interaction mediates dendritic cell trafficking. Nat Immunol 1:353-357

107. Geijtenbeek TB, Torensma R, van Vliet SJ, van Duijnhoven GC, Adema GJ, van Kooyk Y, Figdor CG (2000) Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. Cell 100:575-585

108. Geijtenbeek TB, van Vliet SJ, Engering A, 't Hart BA, van Kooyk Y (2004) Self- and nonself-recognition by C-type lectins on dendritic cells. Annu Rev Immunol 22:33-54

109. Geijtenbeek TB, van Vliet SJ, Koppel EA, Sanchez-Hernandez M, Vandenbroucke-Grauls CM, Appelmelk B, van Kooyk Y (2003) Mycobacteria target DC-SIGN to suppress dendritic cell function. J Exp Med 197:7-17

110. Gelder CM, Hart KW, Williams OM, Lyons E, Welsh KI, Campbell IA, Marshall SE (2000) Vitamin D receptor gene polymorphisms and susceptibility to Mycobacterium malmoense pulmonary disease. Journal of Infectious Diseases 181:2099-2102

111. Golden MP, Vikram HR (2005) Extrapulmonary tuberculosis: an overview. Am Fam Physician 72:1761-1768

112. Goldstein DB, Hirschhorn JN (2004) In genetic control of disease, does 'race' matter? Nat Genet 36:1243-1244

113. Gomez LM, Anaya JM, Sierra-Filardi E, Cadena J, Corbi A, Martin J (2006) Analysis of DC-SIGN (CD209) functional variants in patients with tuberculosis. Hum Immunol 67:808-811

114. Gonzalez E, Bamshad M, Sato N, Mummidi S, Dhanda R, Catano G, Cabrera S, McBride M, Cao XH, Merrill G, O'Connell P, Bowden DW, Freedman BI, Anderson SA, Walter EA, Evans JS, Stephan KT, Clark RA, Tyagi S, Ahuja SS, Dolan MJ, Ahuja SK (1999) Race-specific HIV-1 disease-modifying effects associated with CCR5 haplotypes. Proceedings of the National Academy of Sciences of the United States of America 96:12004-12009

115. Gonzalez E, Dhanda R, Bamshad M, Mummidi S, Geevarghese R, Catano G, Anderson SA, Walter EA, Stephan KT, Hammer MF, Mangano A, Sen L, Clark RA, Ahuja SS, Dolan MJ, Ahuja SK (2001) Global survey of genetic variation in CCR5, RANTES, and MIP-1alpha: impact on the epidemiology of the HIV-1 pandemic. Proc Natl Acad Sci U S A 98:5199-5204

116. Gonzalez-Neira A, Ke X, Lao O, Calafell F, Navarro A, Comas D, Cann H, Bumpstead S, Ghori J, Hunt S, Deloukas P, Dunham I, Cardon LR, Bertranpetit J (2006) The portability of tagSNPs across populations: a worldwide survey. Genome Res 16:323-330

117. Guo Y, Atkinson CE, Taylor ME, Drickamer K (2006) All but the shortest polymorphic forms of the viral receptor DC-SIGNR assemble into stable homo- and heterotetramers. J Biol Chem 281:16794-16798

118. Halary F, Amara A, Lortat-Jacob H, Messerle M, Delaunay T, Houles C, Fieschi F, Arenzana-Seisdedos F, Moreau JF, Dechanet-Merville J (2002) Human cytomegalovirus binding to DC-SIGN is required for dendritic cell infection and target cell trans-infection. Immunity 17:653-664

119. Haldane JBS (1949) Disease and evolution. La Ricerca Scientifica 19 (suppl):68-76

120. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98

121. Halldorsson BV, Bafna V, Lippert R, Schwartz R, De la Vega FM, Clark AG, Istrail S (2004) Optimal haplotype block-free selection of tagging SNPs for genome-wide association studies. Genome Res 14:1633-1640

122. Halldorsson BV, Istrail S, De la Vega FM (2004) Optimal selection of SNP markers for disease association studies. Hum Hered 58:190-202

123. Hammond MG, Du Toit ED, Sanchez-Mazas A, Andrien M, Coluzzi M, de Pablo MR, de Stefano G, Kaplan C (1997) HLA in sub-Saharan Africa. In: Charon D (ed) EDK, Paris, France, pp 345-353

124. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, Mayr G, De la Vega FM, Briggs J, Gunther S, Prescott NJ, Onnie CM, Hasler R, Sipos B, Folsch UR, Lengauer T, Platzer M, Mathew CG, Krawczak M, Schreiber S (2007) A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet 39:207-211

125. Hanekom M, van der Spuy GD, Streicher E, Ndabambi SL, McEvoy CR, Kidd M, Beyers N, Victor TC, van Helden PD, Warren RM (2007) A recently evolved sublineage of the Mycobacterium tuberculosis Beijing strain family is associated with an increased ability to spread and cause disease. J Clin Microbiol 45:1483-1490

126. Hayes CE, Nashold FE, Spach KM, Pedersen LB (2003) The immunological functions of the vitamin D endocrine system. Cell Mol Biol 49:277-300

127. Hayes VM, Petersen DC, Scriba TJ, Zeier M, Grimwood A, van Rensburg EJ (2002) African-based CCR5 single-nucleotide polymorphism associated with HIV-1 disease progression. AIDS 16:2229-2231

128. Hesseling AC, Rabie H, Marais BJ, Manders M, Lips M, Schaaf HS, Gie RP, Cotton MF, van Helden PD, Warren RM, Beyers N (2006) Bacille Calmette-Guerin vaccine-induced disease in HIV-infected and HIV-uninfected children. Clin Infect Dis 42:548-558

129. Hingley-Wilson SM, Sambandamurthy VK, Jacobs WR, Jr. (2003) Survival perspectives from the world's most successful pathogen, Mycobacterium tuberculosis. Nat Immunol 4:949-955

130. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95-108

131. Hoal EG (2002) Human genetic susceptibility to tuberculosis and other mycobacterial diseases. IUBMB Life 53:225-229

132. Hoal EG, Lewis L-A, Jamieson SE, Tanzer F, Rossouw M, Victor T, Hillerman R, Beyers N, Blackwell JM, van Helden PD (2004) *SLC11A1 (NRAMP1)* but not *SLC11A2 (NRAMP2)* polymorphisms are associated with susceptibility to tuberculosis in a high-incidence community in South Africa. Int J Tuberc Lung Dis 8:1464-1471

133. Hoal-van Helden EG, Stanton LA, Warren R, Richardson M, van Helden PD (2001) Diversity of in vitro cytokine responses by human macrophages to infection by mycobacterium tuberculosis strains. Cell Biol Int 25:83-90

134. Holtz TH, Sternberg M, Kammerer S, Laserson KF, Riekstina V, Zarovska E, Skripconoka V, Wells CD, Leimane V (2006) Time to sputum culture conversion in multidrug-resistant tuberculosis: predictors and relationship to treatment outcome. Ann Intern Med 144:650-659

135. Hou MF, Tien YC, Lin GT, Chen CJ, Liu CS, Lin SY, Huang TJ (2002) Association of vitamin D receptor gene polymorphism with sporadic breast cancer in Taiwanese patients. Breast Cancer Res Treat 74:1-7

136. Hughes MR, Malloy PJ, Kieback DG, Kesterson RA, Pike JW, Feldman D, O'Malley BW (1988) Point mutations in the human vitamin D receptor gene associated with hypocalcemic rickets. Science 242:1702-1705

137. Hull J, Rowlands K, Lockhart E, Moore C, Sharland M, Kwiatkowski D (2003) Variants of the chemokine receptor CCR5 are associated with severe bronchiolitis caused by respiratory syncytial virus. Journal of Infectious Diseases 188:904-907

138. Imanishi T, Akaza T, Kimura A, Tokunaga T, Gojobori T (1992) Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. In: Tsuji K, Aizawa M, Saszuki T (eds) HLA 1991. Proceedings of the Eleventh International Histocompatibility Workshop and Conference. Oxford University Press, Oxford, pp 1065-1074

139. Ingles SA, Haile RW, Henderson BE, Kolonel LN, Nakaichi G, Shi CY, Yu MC, Ross RK, Coetzee GA (1997) Strength of linkage disequilibrium between two vitamin D receptor markers in five ethnic groups: implications for association studies. Cancer Epidemiol Biomarkers Prev 6:93-98

140. Ingles SA, Haile RW, Henderson BE, Kolonel LN, Nakaichi G, Shi CY, Yu MC, Ross RK, Coetzee GA (1997) Strength of linkage disequilibrium between two vitamin D receptor markers in five ethnic groups: implications for association studies. Cancer Epidemiol Biomarkers Prev 6:93-98

141. Jamieson SE, Miller EN, Black GF, Peacock CS, Cordell HJ, Howson JM, Shaw MA, Burgner D, Xu W, Lins-Lainson Z, Shaw JJ, Ramos F, Silveira F, Blackwell JM (2004) Evidence for a cluster of genes on chromosome 17q11-q21 controlling susceptibility to tuberculosis and leprosy in Brazilians. Genes Immun 5:46-57

142. Janeway CA, Jr., Medzhitov R (2002) Innate immune recognition. Annu Rev Immunol 20:197-216

143. Jepson A, Fowler A, Banya W, Singh M, Bennett S, Whittle H, Hill AV (2001) Genetic regulation of acquired immune responses to antigens of Mycobacterium tuberculosis: a study of twins in West Africa. Infect Immun 69:3989-3994

144. Jiang YH, Bressler J, Beaudet AL (2004) Epigenetics and human disease. Annu Rev Genomics Hum Genet 5:479-510

145. Juffermans NP, Paxton WA, Dekkers PEP, Verbon A, de Jonge E, Speelman P, van Deventer SJH, van der Poll T (2000) Up-regulation of HIV coreceptors CXCR4 and CCR5 on CD4(+) T cells during human endotoxemia and after stimulation with (myco)bacterial antigens: the role of cytokines. Blood 96:2649-2654

146. Kadereit S, Gewert DR, Galabru J, Hovanessian AG, Meurs EF (1993) Molecular cloning of two new interferon-induced, highly related nuclear phosphoproteins. J Biol Chem 268:24432-24441

147. Kallmann FJ, Reisner D (1942) Twin studies on the significance of genetic factors in tuberculosis. Am Rev Tuberc 47:549-547

148. Kamath AB, Alt J, Debbabi H, Behar SM (2003) Toll-like receptor 4-defective C3H/HeJ mice are not more susceptible than other C3H substrains to infection with Mycobacterium tuberculosis. Infect Immun 71:4112-4118

149. Kaufmann SH (2001) How can immunology contribute to the control of tuberculosis? Nat Rev Immunol 1:20-30

150. Keller C, Hoffmann R, Lang R, Brandau S, Hermann C, Ehlers S (2006) Genetically determined susceptibility to tuberculosis in mice causally involves accelerated and enhanced recruitment of granulocytes. Infect Immun 74:4295-4309

151. Kim JG, Lim KS, Kim EK, Choi YM, Lee JY (2001) Association of vitamin D receptor and estrogen receptor gene polymorphisms with bone mass in postmenopausal Korean women. Menopause 8:222-228

152. Kitanaka S, Takeyama K, Murayama A, Kato S (2001) The molecular basis of vitamin D-dependent rickets type I. Endocr J 48:427-432

153. Kochi A (1991) The global tuberculosis situation and the new control strategy of the World Health Organization. Tubercle 72:1-6

154. Koppel EA, Ludwig IS, Hernandez MS, Lowary TL, Gadikota RR, Tuzikov AB, Vandenbroucke-Grauls CM, van Kooyk Y, Appelmelk BJ, Geijtenbeek TB (2004) Identification of the mycobacterial carbohydrate structure that binds the C-type lectins DC-SIGN, L-SIGN and SIGNR1. Immunobiology 209:117-127

155. Kostrikis LG, Huang YX, Moore JP, Wolinsky SM, Zhang LQ, Guo Y, Deutsch L, Phair J, Neumann AU, Ho DD (1998) A chemokine receptor CCR2 allele delays HIV-1 disease progression and is associated with a CCR5 promoter mutation. Nature Medicine 4:350-353

156. Kostrikis LG, Neumann AU, Thomson B, Korber BT, McHardy P, Karanicolas R, Deutsch L, Huang YX, Lew JF, McIntosh K, Pollack H, Borkowsky W, Spiegel HML, Palumbo P, Oleske J, Bardeguez A, Luzuriaga K, Sullivan J, Wolinsky SM, Koup RA, Ho DD, Moore JP (1999) A polymorphism in the regulatory region of the CC-chemokine receptor 5 gene influences perinatal transmission of human immunodeficiency virus type 1 to African-American infants. Journal of Virology 73:10264-10271

157. Koul A, Herget T, Klebl B, Ullrich A (2004) Interplay between mycobacteria and host signalling pathways. Nat Rev Microbiol 2:189-202

158. Kramnik I, Demant P, Bloom BB (1998) Susceptibility to tuberculosis as a complex genetic trait: analysis using recombinant congenic strains of mice. Novartis Found Symp 217:120-131

159. Kramnik I, Dietrich WF, Demant P, Bloom BR (2000) Genetic control of resistance to experimental infection with virulent Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 97:8560-8565

160. Kristiansen TB, Knudsen TB, Ohlendorff S, Eugen-Olsen J (2001) A new multiplex PCR strategy for the simultaneous determination of four genetic polymorphisms affecting HIV-1 disease progression. Journal of Immunological Methods 252:147-151

161. Kritzinger F.E, den Boon S, Enarson D, Verver S, Lombard C, Borgdorff MW, Gie R, Beyers N (2007) Increase in high tuberculosis transmission rate asssociated with high burden of previously treated cases in an endemic region. In preparation

162. Kurashima K, Mukaida N, Fujimura M, Yasui M, Nakazumi Y, Matsuda T, Matsushima K (1997) Elevated chemokine levels in bronchoalveolar lavage fluid of tuberculosis patients. American Journal of Respiratory and Critical Care Medicine 155:1474-1477

163. Kusner DJ (2005) Mechanisms of mycobacterial persistence in tuberculosis. Clin Immunol 114:239-247

164. Kwok PY (2002) SNP genotyping with fluorescence polarization detection. Hum Mutat 19:315-323

165. Laird NM, Lange C (2006) Family-based designs in the age of large-scale gene-association studies. Nat Rev Genet 7:385-394

166. Lavebratt C, Apt AS, Nikonenko BV, Schalling M, Schurr E (1999) Severity of tuberculosis in mice is linked to distal chromosome 3 and proximal chromosome 9. J Infect Dis 180:150-155

167. Lee B, Doranz BJ, Rana S, Yi YJ, Mellado M, Frade JMR, Martinez C, O'Brien SJ, Dean M, Collman RG, Doms RW (1998) Influence of the CCR2-V64I polymorphism on human immunodeficiency virus type 1 coreceptor activity and on chemokine receptor function of CCR2b, CCR3, CCR5, and CXCR4. Journal of Virology 72:7450-7458

168. Lewis CM (2002) Genetic association studies: design, analysis and interpretation. Brief Bioinform 3:146-153

169. Lewis SJ, Baker I, Davey SG (2005) Meta-analysis of vitamin D receptor polymorphisms and pulmonary tuberculosis risk. Int J Tuberc Lung Dis 9:1174-1177

170. Li HT, Zhang TT, Zhou YQ, Huang QH, Huang J (2006) SLC11A1 (formerly NRAMP1) gene polymorphisms and tuberculosis susceptibility: a meta-analysis. Int J Tuberc Lung Dis 10:3-12

171. Lichterfeld M, Nischalke HD, van Lunzen J, Sohne J, Schmeisser N, Woitas R, Sauerbruch T, Rockstroh JK, Spengler U (2003) The tandem-repeat polymorphism of the DC-SIGNR gene does not affect the susceptibility to HIV infection and the progression to AIDS. Clin Immunol 107:55-59

172. Lienhardt C (2001) From exposure to disease: the role of environmental factors in susceptibility to and development of tuberculosis. Epidemiol Rev 23:288-301

173. Link DC (2005) Neutrophil homeostasis - A new role for stromal cell-derived factor-1. Immunologic Research 32:169-178

174. Lipsitch M, Sousa AO (2002) Historical intensity of natural selection for resistance to tuberculosis. Genetics 161:1599-1607

175. Liu H, Carrington M, Wang C, Holte S, Lee J, Greene B, Hladik F, Koelle DM, Wald A, Kurosawa K, Rinaldo CR, Celum C, Detels R, Corey L, McElrath MJ, Zhu T (2006) Repeat-region polymorphisms in the gene for the dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin-related molecule: effects on HIV-1 susceptibility. J Infect Dis 193:698-702

176. Liu H, Chao D, Nakayama EE, Taguchi H, Goto M, Xin X, Takamatsu JK, Saito H, Ishikawa Y, Akaza T, Juji T, Takebe Y, Ohishi T, Fukutake K, Maruyama Y, Yashiki S, Sonoda S, Nakamura T, Nagai Y, Iwamoto A, Shioda T (1999) Polymorphism in RANTES chemokine promoter affects HIV-1 disease progression. Proc Natl Acad Sci U S A 96:4581-4585

177. Liu H, Hwangbo Y, Holte S, Lee J, Wang C, Kaupp N, Zhu H, Celum C, Corey L, McElrath MJ, Zhu T (2004) Analysis of genetic polymorphisms in CCR5, CCR2, stromal cell-derived factor-1, RANTES, and dendritic cell-

specific intercellular adhesion molecule-3-grabbing nonintegrin in seronegative individuals repeatedly exposed to HIV-1. J Infect Dis 190:1055-1058

178. Liu PT, Stenger S, Li H, Wenzel L, Tan BH, Krutzik SR, Ochoa MT, Schauber J, Wu K, Meinken C, Kamen DL, Wagner M, Bals R, Steinmeyer A, Zugel U, Gallo RL, Eisenberg D, Hewison M, Hollis BW, Adams JS, Bloom BR, Modlin RL (2006) Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. Science 311:1770-1773

179. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, MacDonald ME, Stuhlmann H, Koup RA, Landau NR (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 86:367-377

180. Liu W, Cao WC, Zhang CY, Tian L, Wu XM, Habbema JD, Zhao QM, Zhang PH, Xin ZT, Li CZ, Yang H (2004) VDR and NRAMP1 gene polymorphisms in susceptibility to pulmonary tuberculosis among the Chinese Han population: a case-control study. Int J Tuberc Lung Dis 8:428-434

181. Liu W, Zhang CY, Tian L, Li CZ, Wu XM, Zhao QM, Zhang PH, Yang SM, Yang H, Cao WC (2003) [A case-control study on natural-resistance-associated macrophage protein 1 gene polymorphisms and susceptibility to pulmonary tuberculosis]. Zhonghua Yu Fang Yi Xue Za Zhi 37:408-411

182. Lombard Z, Brune AE, Hoal EG, Babb C, van Helden PD, Epplen JT, Bornman L (2006) HLA class II disease associations in southern Africa. Tissue Antigens 67:97-110

183. Lombard Z, Dalton DL, Venter PA, Williams RC, Bornman L (2006) Association of HLA-DR, -DQ, and vitamin D receptor alleles and haplotypes with tuberculosis in the Venda of South Africa. Hum Immunol 67:643-654

184. Lopez B, Aguilar D, Orozco H, Burger M, Espitia C, Ritacco V, Barrera L, Kremer K, Hernandez-Pando R, Huygen K, Van Soolingen D (2003) A marked difference in pathogenesis and immune response induced by different Mycobacterium tuberculosis genotypes. Clinical and Experimental Immunology 133:30-37

185. Lozach PY, Lortat-Jacob H, de Lacroix dL, Staropoli I, Foung S, Amara A, Houles C, Fieschi F, Schwartz O, Virelizier JL, Arenzana-Seisdedos F, Altmeyer R (2003) DC-SIGN and L-SIGN are high affinity binding receptors for hepatitis C virus glycoprotein E2. J Biol Chem 278:20358-20366

186. Lux M (1998) Perfect subjects: race, tuberculosis, and the Qu'Appelle BCG Vaccine Trial. Can Bull Med Hist 15:277-295

187. Ma X, Dou S, Wright JA, Reich RA, Teeter LD, El Sahly HM, Awe RJ, Musser JM, Graviss EA (2002) 5' dinucleotide repeat polymorphism of NRAMP1 and susceptibility to tuberculosis among Caucasian patients in Houston, Texas. Int J Tuberc Lung Dis 6:818-823

188. Macgregor S, Khan IA (2006) GAIA: an easy-to-use web-based application for interaction analysis of case-control data. BMC Med Genet 7:34

189. Maeda N, Nigou J, Herrmann JL, Jackson M, Amara A, Lagrange PH, Puzo G, Gicquel B, Neyrolles O (2003) The cell surface receptor DC-SIGN discriminates between Mycobacterium species through selective recognition of the mannose caps on lipoarabinomannan. Journal of Biological Chemistry 278:5513-5516

190. Maharaj B, Hammond MG, Appadoo B, Leary WP, Pudifin DJ (1987) HLA-A, B, DR, and DQ antigens in black patients with severe chronic rheumatic heart disease. Circulation 76:259-261

191. Makino Y, Cook DN, Smithies O, Hwang OY, Neilson EG, Turka LA, Sato H, Wells AD, Danoff TM (2002) Impaired T cell function in RANTES-deficient mice. Clin Immunol 102:302-309

192. Malik AN, Godfrey-Faussett P (2005) Effects of genetic variability of Mycobacterium tuberculosis strains on the presentation of disease. Lancet Infect Dis 5:174-183

193. Mansouri D, Adimi P, Mirsaeidi M, Mansouri N, Khalilzadeh S, Masjedi MR, Adimi P, Tabarsi P, Naderi M, Filipe-Santos O, Vogt G, de Beaucoudrey L, Bustamante J, Chapgier A, Feinberg J, Velayati AA, Casanova JL (2005) Inherited disorders of the IL-12-IFN-gamma axis in patients with disseminated BCG infection. Eur J Pediatr 164:753-757

194. Marais BJ, Gie RP, Hesseling AC, Schaaf HS, Lombard C, Enarson DA, Beyers N (2006) A refined symptom-based approach to diagnose pulmonary tuberculosis in children. Pediatrics 118:e1350-e1359

195. Marais BJ, Hesseling AC, Gie RP, Schaaf HS, Beyers N (2006) The burden of childhood tuberculosis and the accuracy of community-based surveillance data. Int J Tuberc Lung Dis 10:259-263

196. Marais BJ, Pai M (2006) Specimen collection methods in the diagnosis of childhood tuberculosis. Indian J Med Microbiol 24:249-251

197. Marais BJ, Victor TC, Hesseling AC, Barnard M, Jordaan A, Brittle W, Reuter H, Beyers N, van Helden PD, Warren RM, Schaaf HS (2006) Beijing and Haarlem genotypes are overrepresented among children with drug-resistant tuberculosis in the Western Cape Province of South Africa. J Clin Microbiol 44:3539-3543

198. Martell RW, Du Toit ED, Kalla AA, Meyers OL (1989) Association of rheumatoid arthritis with HLA in three South African populations--whites, blacks and a population of mixed ancestry. S Afr Med J 76:189-190

199. Martell RW, van Eeden PJ, Madden MV, Du Toit ED, Dent DM (1988) Lack of an association between carcinoma of the stomach and the major histocompatibility complex (HLA) in Cape Coloureds. Tissue Antigens 32:150-154

200. Martin MP, Lederman MM, Hutcheson HB, Goedert JJ, Nelson GW, van Kooyk Y, Detels R, Buchbinder S, Hoots K, Vlahov D, O'Brien SJ, Carrington M (2004) Association of DC-SIGN promoter polymorphism with increased risk for parenteral, but not mucosal, acquisition of human immunodeficiency virus type 1 infection. J Virol 78:14053-14056

201. Martineau AR, Honecker FU, Wilkinson RJ, Griffiths CJ (2007) Vitamin D in the treatment of pulmonary tuberculosis. J Steroid Biochem Mol Biol 103:793-798

202. Martineau AR, Wilkinson RJ, Wilkinson KA, Newton SM, Kampmann B, Hall BM, Packe GE, Davidson RN, Eldridge SM, Maunsell ZJ, Rainbow SJ, Berry JL, Griffiths CJ (2007) A Single Dose of Vitamin D Enhances Immunity to Mycobacteria. Am J Respir Crit Care Med

203. Marzi A, Gramberg T, Simmons G, Moller P, Rennekamp AJ, Krumbiegel M, Geier M, Eisemann J, Turza N, Saunier B, Steinkasserer A, Becker S, Bates P, Hofmann H, Pohlmann S (2004) DC-SIGN and DC-SIGNR interact with the glycoprotein of Marburg virus and the S protein of severe acute respiratory syndrome coronavirus. J Virol 78:12090-12095

204. Mathieu C, Adorini L (2002) The coming of age of 1,25-dihydroxyvitamin D(3) analogs as immunomodulatory agents. Trends Mol Med 8:174-179

205. McDermott DH, Beecroft MJ, Kleeberger CA, Al Sharif FM, Ollier WE, Zimmerman PA, Boatin BA, Leitman SF, Detels R, Hajeer AH, Murphy PM (2000) Chemokine RANTES promoter polymorphism affects risk of both HIV infection and disease progression in the Multicenter AIDS Cohort Study. AIDS 14:2671-2678

206. McDermott DH, Zimmerman PA, Guignard F, Kleeberger CA, Leitman SF, Murphy PM (1998) CCR5 promoter polymorphism and HIV-1 disease progression. Lancet 352:866-870

207. McGreal EP, Miller JL, Gordon S (2005) Ligand recognition by antigen-presenting cell C-type lectin receptors. Curr Opin Immunol 17:18-24

208. Medzhitov R (2001) Toll-like receptors and innate immunity. Nat Rev Immunol 1:135-145

209. Michael NL, Chang G, Louie LG, Mascola JR, Dondero D, Birx DL, Sheppard HW (1997) The role of viral phenotype and CCR-5 gene defects in HIV-1 transmission and disease progression. Nature Medicine 3:338-340

210. Michael NL, Louie LG, Rohrbaugh AL, Schultz KA, Dayhoff DE, Wang CE, Sheppard HW (1997) The role of CCR5 and CCR2 polymorphisms in HTV-1 transmission and disease progression. Nature Medicine 3:1160-1162

211. Miller EN, Jamieson SE, Joberty C, Fakiola M, Hudson D, Peacock CS, Cordell HJ, Shaw MA, Lins-Lainson Z, Shaw JJ, Ramos F, Silveira F, Blackwell JM (2004) Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians. Genes Immun 5:63-67

212. Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, Mai CP, Nguyen TH, Nguyen NB, Pham XK, Sarno EN, Alter A, Montpetit A, Moraes ME, Moraes JR, Dore C, Gallant CJ, Lepage P, Verner A, Van D, V, Hudson TJ, Abel L, Schurr E (2004) Susceptibility to leprosy is associated with PARK2 and PACRG. Nature 427:636-640

213. Mira MT, Alcais A, Van Thuc N, Thai VH, Huong NT, Ba NN, Verner A, Hudson TJ, Abel L, Schurr E (2003) Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. Nat Genet 33:412-415

214. Mitsos LM, Cardon LR, Fortin A, Ryan L, LaCourse R, North RJ, Gros P (2000) Genetic control of susceptibility to infection with Mycobacterium tuberculosis in mice. Genes Immun 1:467-477

215. Mitsos LM, Cardon LR, Ryan L, LaCourse R, North RJ, Gros P (2003) Susceptibility to tuberculosis: a locus on mouse chromosome 19 (Trl-4) regulates Mycobacterium tuberculosis replication in the lungs. Proc Natl Acad Sci U S A 100:6610-6615

216. Mody GM, Hammond MG (1994) Differences in Hla-Dr Association with Rheumatoid-Arthritis Among Migrant Indian Communities in South-Africa. British Journal of Rheumatology 33:425-427

217. Mody GM, Hammond MG, Naidoo PD (1989) HLA associations with rheumatoid arthritis in African blacks. J Rheumatol 16:1326-1328

218. Moller M, Kwiatkowski R, Nebel A, van Helden PD, Hoal EG, Schreiber S (2007) Allelic variation in BTNL2 and susceptibility to tuberculosis in a South African population. Microbes Infect 9:522-528

219. Moller M, Nebel A, Kwiatkowski R, van Helden PD, Hoal EG, Schreiber S (2007) Host susceptibility to tuberculosis: CARD15 polymorphisms in a South African population. Mol Cell Probes 21:148-151

220. Montpetit A, Nelis M, Laflamme P, Magi R, Ke X, Remm M, Cardon L, Hudson TJ, Metspalu A (2006) An evaluation of the performance of tag SNPs derived from HapMap in a Caucasian population. PLoS Genet 2:e27

221. Morrison NA, Qi JC, Tokita A, Kelly PJ, Crofts L, Nguyen TV, Sambrook PN, Eisman JA (1994) Prediction of bone density from vitamin D receptor alleles. Nature 367:284-287

222. Motulsky AG (1960) Metabolic polymorphisms and the role of infectious diseases in human evolution. Human Biology 32:28-62

223. Mountain A (2003) The First People of the Cape. David Philips Publishers, Claremont, South Africa

224. Mueller A, Strange PG (2004) The chemokine receptor, CCR5. International Journal of Biochemistry & Cell Biology 36:35-38

225. Mueller JC (2004) Linkage disequilibrium for different scales and applications. Brief Bioinform 5:355-364

226. Mummidi S, Ahuja SS, McDaniel B, Ahuja SK (1997) Human CC chemokine receptor 5 (CCR5) gene: Multiple transcripts with 5'-end heterogeneity and evidence for alternative promoter usage. Molecular Biology of the Cell 8:1810

227. Mummidi S, Catano G, Lam L, Hoefle A, Telles V, Begum K, Jimenez F, Ahuja SS, Ahuja SK (2001) Extensive repertoire of membrane-bound and soluble dendritic cell-specific ICAM-3-grabbing nonintegrin 1 (DC-SIGN1) and DC-SIGN2 isoforms. Inter-individual variation in expression of DC-SIGN transcripts. J Biol Chem 276:33196-33212

228. Murray CJ, Styblo K, Rouillon A (1990) Tuberculosis in developing countries: burden, intervention and cost. Bull Int Union Tuberc Lung Dis 65:6-24

229. Nagel RL (2005) Epistasis and the genetics of human diseases. C R Biol 328:606-615

230. Nagpal S, Na S, Rathnachalam R (2005) Noncalcemic actions of vitamin D receptor ligands. Endocr Rev 26:662-687

231. Nattermann J, Ahlenstiel G, Berg T, Feldmann G, Nischalke HD, Muller T, Rockstroh J, Woitas R, Sauerbruch T, Spengler U (2006) The tandem-repeat polymorphism of the DC-SIGNR gene in HCV infection. J Viral Hepat 13:42-46

232. Nejentsev S, Godfrey L, Snook H, Rance H, Nutland S, Walker NM, Lam AC, Guja C, Ionescu-Tirgoviste C, Undlien DE, Ronningen KS, Tuomilehto-Wolf E, Tuomilehto J, Newport MJ, Clayton DG, Todd JA (2004) Comparative high-resolution analysis of linkage disequilibrium and tag single nucleotide polymorphisms between populations in the vitamin D receptor gene. Hum Mol Genet 13:1633-1639

233. Newport M, Levin M (1999) Genetic susceptibility to tuberculosis. J Infect 39:117-121

234. Newport MJ, Goetghebuer T, Weiss HA, Whittle H, Siegrist CA, Marchant A (2004) Genetic regulation of immune responses to vaccines in early life. Genes Immun 5:122-129

235. Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res 17:2503-2516

236. Nicewonger J, Suck G, Bloch D, Swaminathan S (2004) Epstein-Barr virus (EBV) SM protein induces and recruits cellular Sp110b to stabilize mRNAs and enhance EBV lytic gene expression. J Virol 78:9412-9422

237. Nickel RG, Casolaro V, Wahn U, Beyer K, Barnes KC, Plunkett BS, Freidhoff LR, Sengler C, Plitt JR, Schleimer RP, Caraballo L, Naidu RP, Levett PN, Beaty TH, Huang SK (2000) Atopic dermatitis is associated with a functional mutation in the promoter of the C-C chemokine RANTES. J Immunol 164:1612-1616

238. Nicodemus KK, Liu W, Chase GA, Tsai YY, Fallin MD (2005) Comparison of type I error for multiple test corrections in large single-nucleotide polymorphism studies using principal components versus haplotype blocking algorithms. BMC Genet 6 Suppl 1:S78

239. Nishijima S, Sugaya K, Naito A, Morozumi M, Hatano T, Ogawa Y (2002) Association of vitamin D receptor gene polymorphism with urolithiasis. J Urol 167:2188-2191

240. Nordborg M, Tavare S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet 18:83-90

241. Novembre J, Galvani AP, Slatkin M (2005) The geographic spread of the CCR5 Delta32 HIV-resistance allele. PLoS Biol 3:e339

242. Nurse GT, Weiner JS, Jenkins T (1985) The peoples of Southern Africa and their affinities. Clarendon Press, Oxford

243. O'Brien SJ (1991) Ghetto legacy. Curr Biol 1:209-211

244. Ogunkolade BW, Boucher BJ, Prahl JM, Bustin SA, Burrin JM, Noonan K, North BV, Mannan N, McDermott MF, Deluca HF, Hitman GA (2002) Vitamin D receptor (VDR) mRNA and VDR protein levels in relation to vitamin D status, insulin secretory capacity, and VDR genotype in Bangladeshi Asians. Diabetes 51:2294-2300

245. Omar MA, Hammond MG, Asmal AC (1984) HLA-A, B, C and DR antigens in young South African blacks with Type 1 (insulin-dependent) diabetes mellitus. Diabetologia 26:20-23

246. Omar MA, Hammond MG, Motala AA, Seedat MA (1988) HLA class I and II antigens in South African Indians with NIDDM. Diabetes 37:796-799

247. Orme IM (2003) The mouse as a useful model of tuberculosis. Tuberculosis (Edinb) 83:112-115

248. Orren A, Taljaard D, de Toit E (1985) HLA-A, B, C and DR antigen associations in insulin dependent diabetes mellitus (IDDM) in South African Negro (black) and Cape coloured people. Tissue Antigens 26:332-339

249. Ortona L, Federico G (1998) Pulmonary and extrapulmonary tuberculosis. Rays 23:64-77

250. Pan H, Yan BS, Rojas M, Shebzukhov YV, Zhou H, Kobzik L, Higgins DE, Daly MJ, Bloom BR, Kramnik I (2005) Ipr1 gene mediates innate immunity to tuberculosis. Nature 434:767-772

251. Pani MA, Knapp M, Donner H, Braun J, Baur MP, Usadel KH, Badenhoop K (2000) Vitamin D receptor allele combinations influence genetic susceptibility to type 1 diabetes in Germans. Diabetes 49:504-507

252. Petersen DC, Kotze MJ, Zeier MD, Grimwood A, Pretorius D, Vardas E, van Rensburg EJ, Hayes VM (2001) Novel mutations identified using a comprehensive CCR5-denaturing gradient gel electrophoresis assay. AIDS 15:171-177

253. Petersen DC, Laten A, Zeier MD, Grimwood A, van Rensburg EJ, Hayes VM (2002) Novel mutations and SNPs identified in CCR2 using a new comprehensive denaturing gradient gel electrophoresis assay. Human Mutation 20:253-259

254. Picard C, Fieschi C, Altare F, Al Jumaah S, Al Hajjar S, Feinberg J, Dupuis S, Soudais C, Al Mohsen IZ, Genin E, Lammas D, Kumararatne DS, Leclerc T, Rafii A, Frayha H, Murugasu B, Wah LB, Sinniah R, Loubser M, Okamoto E, Al Ghonaium A, Tufenkeji H, Abel L, Casanova JL (2002) Inherited interleukin-12 deficiency: IL12B genotype and clinical phenotype of 13 patients from six kindreds. Am J Hum Genet 70:336-348

255. Pirie FJ, Hammond MG, Motala AA, Omar MA (2001) HLA class II antigens in South African Blacks with type I diabetes. Tissue Antigens 57:348-352

256. Plenge R, Rioux JD (2006) Identifying susceptibility genes for immunological disorders: patterns, power, and proof. Immunol Rev 210:40-51

257. Pohlmann S, Soilleux EJ, Baribaud F, Leslie GJ, Morris LS, Trowsdale J, Lee B, Coleman N, Doms RW (2001) DC-SIGNR, a DC-SIGN homologue expressed in endothelial cells, binds to human and simian immunodeficiency viruses and activates infection in trans. Proc Natl Acad Sci U S A 98:2670-2675

258. Pritchard JK, Donnelly P (2001) Case-control studies of association in structured or admixed populations. Theor Popul Biol 60:227-237

259. Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1-14

260. Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65:220-228

261. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170-181

262. Puzyrev VP, Freidin MB, Rudko AA, Strelis AK, Kolokolova OV (2002) [Polymorphisms of the candidate genes for genetic susceptibility to tuberculosis in the Slavic population of Siberia: a pilot study]. Mol Biol (Mosk) 36:788-791

263. Quast TM, Browning RF (2006) Pathogenesis and clinical manifestations of pulmonary tuberculosis. Dis Mon 52:413-419

264. Quesniaux V, Fremond C, Jacobs M, Parida S, Nicolle D, Yeremeev V, Bihl F, Erard F, Botha T, Drennan M, Soler MN, Le Bert M, Schnyder B, Ryffel B (2004) Toll-like receptor pathways in the immune responses to mycobacteria. Microbes Infect 6:946-959

265. Quintans B, Alvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo A (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. Forensic Sci Int 140:251-257

266. R Development Core Team (2006) R: A language and environment for statistical computing. www.r-project.org. R foundation for statistical computing, Vienna, Austria,

267. Rani R, Fernandez-Vina MA, Stastny P (1998) Associations between HLA class II alleles in a North Indian population. Tissue Antigens 52:37-43

268. Raviglione MC (2003) The TB epidemic from 1992 to 2002. Tuberculosis (Edinb) 83:4-14

269. Raviglione MC (2006) The Global Plan to Stop TB, 2006-2015. Int J Tuberc Lung Dis 10:238-239

270. Reed MB, Domenech P, Manca C, Su H, Barczak AK, Kreiswirth BN, Kaplan G, Barry CE, III (2004) A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature 431:84-87

271. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. Nature 411:199-204

272. Reich DE, Goldstein DB (2001) Detecting association in a case-control study while correcting for population stratification. Genet Epidemiol 20:4-16

273. Reich DE, Lander ES (2001) On the allelic spectrum of human disease. Trends Genet 17:502-510

274. Reichman LB (1996) How to ensure the continued resurgence of tuberculosis. Lancet 347:175-177

275. Richardson M, Van Lill SW, van der Spuy GD, Munch Z, Booysen CN, Beyers N, van Helden PD, Warren RM (2002) Historic and recent events contribute to the disease dynamics of Beijing-like Mycobacterium tuberculosis isolates in a high incidence region. Int J Tuberc Lung Dis 6:1001-1011

276. Richeldi L (2006) An update on the diagnosis of tuberculosis infection. Am J Respir Crit Care Med 174:736-742

277. Risch N, Burchard E, Ziv E, Tang H (2002) Categorization of humans in biomedical research: genes, race and disease. Genome Biol 3:comment2007

278. Rockett KA, Brookes R, Udalova I, Vidal V, Hill AV, Kwiatkowski D (1998) 1,25-Dihydroxyvitamin D3 induces nitric oxide synthase and suppresses growth of Mycobacterium tuberculosis in a human macrophage-like cell line. Infect Immun 66:5314-5321

279. Rockhill B, Newman B, Weinberg C (1998) Use and misuse of population attributable fractions. Am J Public Health 88:15-19

280. Rom WN, Garay SM (2004) Tuberculosis. Lippincott Williams & Wilkins, Philadelphia

281. Romero R, Kuivaniemi H, Tromp G, Olson J (2002) The design, execution, and interpretation of genetic association studies to decipher complex diseases. Am J Obstet Gynecol 187:1299-1312

282. Rook GA, Steele J, Fraher L, Barker S, Karmali R, O'Riordan J, Stanford J (1986) Vitamin D3, gamma interferon, and control of proliferation of Mycobacterium tuberculosis by human monocytes. Immunology 57:159-163

283. Roscioli T, Cliffe ST, Bloch DB, Bell CG, Mullan G, Taylor PJ, Sarris M, Wang J, Donald JA, Kirk EP, Ziegler JB, Salzer U, McDonald GB, Wong M, Lindeman R, Buckley MF (2006) Mutations in the gene encoding the PML nuclear body protein Sp110 are associated with immunodeficiency and hepatic veno-occlusive disease. Nat Genet 38:620-622

284. Rossouw M, Nel HJ, Cooke GS, van Helden PD, Hoal EG (2003) Association between tuberculosis and a polymorphic NFkappaB binding site in the interferon gamma gene. Lancet 361:1871-1872

285. Roth DE, Soto G, Arenas F, Bautista CT, Ortiz J, Rodriguez R, Cabrera L, Gilman RH (2004) Association between Vitamin D Receptor Gene Polymorphisms and Response to Treatment of Pulmonary Tuberculosis. J Infect Dis 190:920-927

286. Rouillon A, Perdrizet S, Parrot R (1976) Transmission of tubercle bacilli: The effects of chemotherapy. Tubercle 57:275-299

287. Rousseau J, Williams RC, Creemers P (2003) HLA DRB1, DQA1, DQB1 and DPB1 polymorphisms in Namibian Khoi and San and in the Xhosa and South African mixed-ancestry populations. South African Journal of Science 99:81-89

288. Roy S, Frodsham A, Saha B, Hazra SK, Mascie-Taylor CG, Hill AV (1999) Association of vitamin D receptor genotype with leprosy type. J Infect Dis 179:187-191

289. Roy S, Frodsham A, Saha B, Hazra SK, Mascie-Taylor CGN, Hill AVS (1999) Association of vitamin D receptor genotype with leprosy type. Journal of Infectious Diseases 179:187-191

290. Rudwaleit M, Tikly M, Gibson K, Pile K, Wordsworth P (1995) HLA Class-II Antigens Associated with Systemic Lupus-Erythematosus in black South-Africans. Annals of the Rheumatic Diseases 54:678-680

291. Rybicki BA, Walewski JL, Maliarik MJ, Kian H, Iannuzzi MC (2005) The BTNL2 gene and sarcoidosis susceptibility in African Americans and Whites. Am J Hum Genet 77:491-499

292. Ryu S, Park YK, Bai GH, Kim SJ, Park SN, Kang S (2000) 3'UTR polymorphisms in the NRAMP1 gene are associated with susceptibility to tuberculosis in Koreans. Int J Tuberc Lung Dis 4:577-580

293. Sabeti PC, Walsh E, Schaffner SF, Varilly P, Fry B, Hutcheson HB, Cullen M, Mikkelsen TS, Roy J, Patterson N, Cooper R, Reich D, Altshuler D, O'brien S, Lander ES (2005) The case for selection at CCR5-Delta32. PLoS Biol 3:e378

294. Sadek MI, Sada E, Toossi Z, Schwander SK, Rich EA (1998) Chemokines induced by infection of mononuclear phagocytes with mycobacteria and present in lung alveoli during active pulmonary tuberculosis. Am J Respir Cell Mol Biol 19:513-521

295. Sakuntabhai A, Turbpaiboon C, Casademont I, Chuansumrit A, Lowhnoo T, Kajaste-Rudnitski A, Kalayanarooj SM, Tangnararatchakit K, Tangthawornchaikul N, Vasanawathana S, Chaiyaratana W, Yenchitsomanus PT, Suriyaphol P, Avirutnan P, Chokephaibulkit K, Matsuda F, Yoksan S, Jacob Y, Lathrop GM, Malasit P, Despres P, Julier C (2005) A variant in the CD209 promoter is associated with severity of dengue disease. Nat Genet 37:507-513

296. Salmon JF, Martell R (1994) The role of ethnicity in primary angle-closure glaucoma. S Afr Med J 84:623-626

297. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, Saragosti S, Lapoumeroulie C, Cognaux J, Forceille C, Muyldermans G, Verhofstede C, Burtonboy G, Georges M, Imai T, Rana S, Yi YJ, Smyth RJ, Collman RG, Doms RW, Vassart G, Parmentier M (1996) Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382:722-725

298. Saukkonen JJ, Bazydlo B, Thomas M, Strieter RM, Keane J, Kornfeld H (2002) Beta-chemokines are induced by Mycobacterium tuberculosis and inhibit its growth. Infect Immun 70:1684-1693

299. Schaaf HS, Michaelis IA, Richardson M, Booysen CN, Gie RP, Warren R, van Helden PD, Beyers N (2003) Adult-to-child transmission of tuberculosis: household or community contact? Int J Tuberc Lung Dis 7:426-431

300. Schulze TG, McMahon FJ (2002) Genetic association mapping at the crossroads: which test and why? Overview and practical guidelines. Am J Med Genet 114:1-11

301. Selvaraj P, Narayanan PR, Reetha AM (2000) Association of vitamin D receptor genotypes with the susceptibility to pulmonary tuberculosis in female patients & resistance in female contacts. Indian J Med Res 111:172-179

302. Selwyn PA, Hartel D, Lewis VA, Schoenbaum EE, Vermund SH, Klein RS, Walker AT, Friedland GH (1989) A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. N Engl J Med 320:545-550

303. Sharma OP (2000) Hypercalcemia in granulomatous disorders: a clinical review. Curr Opin Pulm Med 6:442-447

304. Shaw MA, Collins A, Peacock CS, Miller EN, Black GF, Sibthorpe D, Lins-Lainson Z, Shaw JJ, Ramos F, Silveira F, Blackwell JM (1997) Evidence that genetic susceptibility to Mycobacterium tuberculosis in a Brazilian population is under oligogenic control: linkage study of the candidate genes NRAMP1 and TNFA. Tuber Lung Dis 78:35-45

305. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29:308-311

306. Shetty N, Shemko M, Vaz M, D'Souza G (2006) An epidemiological evaluation of risk factors for tuberculosis in South India: a matched case control study. Int J Tuberc Lung Dis 10:80-86

307. Siddiqui MR, Meisner S, Tosh K, Balakrishnan K, Ghei S, Fisher SE, Golding M, Shanker Narayan NP, Sitaraman T, Sengupta U, Pitchappan R, Hill AV (2001) A major susceptibility locus for leprosy in India maps to chromosome 10p13. Nat Genet 27:439-441

308. Simmonds B (2004) Tuberculosis in twins. Pitman Medical Publishing Company,

309. Simon G (1966) Radiology in epidemiological studies and some therapeutic trials. Br Med J 2:491-494

310. Singh KK, Barroga CF, Hughes MD, Chen J, Raskino C, McKinney RE, Spector SA (2003) Genetic influence of CCR5, CCR2, and SDF1 variants on human immunodeficiency virus 1 (HIV-1)-related disease progression and neurological impairment, in children with symptomatic HIV-1 infection. Journal of Infectious Diseases 188:1461-1472

311. Singh KK, Barroga CF, Hughes MD, Chen J, Raskino C, McKinney RE, Spector SA (2004) Prevalence of chemokine and chemokine receptor polymorphisms in seroprevalent children with symptomatic HIV-1 infection in the United States. Jaids-Journal of Acquired Immune Deficiency Syndromes 35:309-313

312. Sinnwell JP, Schaid DJ (2006) Statistical Analysis of Haplotypes with Traits and Covariates when Linkage Phase is Ambigious. R package version 1.2.2.

313. Skwor TA, Sedberry AS, Mackie JT, Russell K, Berghman LR, McMurray DN (2006) BCG vaccination of guinea pigs modulates Mycobacterium tuberculosis-induced CCL5 (RANTES) production in vitro and in vivo. Tuberculosis (Edinb) 86:419-429

314. Slatter ML, Yakumo K, Hoffman M, Neuhausen S (2001) Variants of the VDR gene and risk of colon cancer (United States). Cancer Causes Control 12:359-364

315. Smith MW, O'Brien SJ (2005) Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. Nat Rev Genet 6:623-632

316. Snyder GA, Colonna M, Sun PD (2005) The structure of DC-SIGNR with a portion of its repeat domain lends insights to modeling of the receptor tetramer. J Mol Biol 347:979-989

317. Soborg C, Andersen AB, Madsen HO, Kok-Jensen A, Skinhoj P, Garred P (2002) Natural resistance-associated macrophage protein 1 polymorphisms are associated with microscopy-positive tuberculosis. J Infect Dis 186:517-521

318. Soilleux EJ (2003) DC-SIGN (dendritic cell-specific ICAM-grabbing non-integrin) and DC-SIGN-related (DC-SIGNR): friend or foe? Clin Sci (Lond) 104:437-446

319. Soilleux EJ, Barten R, Trowsdale J (2000) DC-SIGN; a related gene, DC-SIGNR; and CD23 form a cluster on 19p13. J Immunol 165:2937-2942

320. Soilleux EJ, Morris LS, Lee B, Pohlmann S, Trowsdale J, Doms RW, Coleman N (2001) Placental expression of DC-SIGN may mediate intrauterine vertical transmission of HIV. J Pathol 195:586-592

321. Soilleux EJ, Morris LS, Leslie G, Chehimi J, Luo Q, Levroney E, Trowsdale J, Montaner LJ, Doms RW, Weissman D, Coleman N, Lee B (2002) Constitutive and induced expression of DC-SIGN on dendritic cell and macrophage subpopulations in situ and in vitro. J Leukoc Biol 71:445-457

322. Sorensen TI, Nielsen GG, Andersen PK, Teasdale TW (1988) Genetic and environmental influences on premature death in adult adoptees. N Engl J Med 318:727-732

323. Squire SB, Obasi A, Nhlema-Simwaka B (2006) The Global Plan to Stop TB: a unique opportunity to address poverty and the Millennium Development Goals. Lancet 367:955-957

324. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM (1997) Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A 94:9869-9874

325. Stead WW (1992) Genetics and resistance to tuberculosis. Could resistance be enhanced by genetic engineering? Ann Intern Med 116:937-941

326. Stead WW, Senner JW, Reddick WT, Lofgren JP (1990) Racial differences in susceptibility to infection by Mycobacterium tuberculosis. N Engl J Med 322:422-427

327. Stegelmann F, Bastian M, Swoboda K, Bhat R, Kiessler V, Krensky AM, Roellinghoff M, Modlin RL, Stenger S (2005) Coordinate expression of CC chemokine ligand 5, granulysin, and perforin in CD8+ T cells provides a host defense mechanism against Mycobacterium tuberculosis. J Immunol 175:7474-7483

328. Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162-1169

329. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978-989

330. Sterne JA, Rodrigues LC, Guedes IN (1998) Does the efficacy of BCG decline with time since vaccination? Int J Tuberc Lung Dis 2:200-207

331. Stoffels K, Overbergh L, Giulietti A, Verlinden L, Bouillon R, Mathieu C (2006) Immune regulation of 25-hydroxyvitamin-D3-1alpha-hydroxylase in human monocytes. J Bone Miner Res 21:37-47

332. Sun JL, Meng HX, Cao CF, Tachi Y, Shinohara M, Ueda M, Imai H, Ohura K (2002) Relationship between vitamin D receptor gene polymorphism and periodontitis. J Periodontal Res 37:263-267

333. Surdo PL, Bottomley MJ, Sattler M, Scheffzek K (2003) Crystal structure and nuclear magnetic resonance analyses of the SAND domain from glucocorticoid modulatory element binding protein-1 reveals deoxyribonucleic acid and zinc binding regions. Mol Endocrinol 17:1283-1295

334. Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2:930-942

335. Szeszko JS, Healy B, Stevens H, Balabanova Y, Drobniewski F, Todd JA, Nejentsev S (2006) Resequencing and association analysis of the SP110 gene in adult pulmonary tuberculosis. Hum Genet

336. Tailleux L, Maeda N, Nigou J, Gicquel B, Neyrolles O (2003) How is the phagocyte lectin keyboard played? Master class lesson by Mycobacterium tuberculosis. Trends Microbiol 11:259-263

337. Tailleux L, Pham-Thi N, Bergeron-Lafaurie A, Herrmann JL, Charles P, Schwartz O, Scheinmann P, Lagrange PH, de Blic J, Tazi A, Gicquel B, Neyrolles O (2005) DC-SIGN induction in alveolar macrophages defines privileged target host cells for mycobacteria in patients with tuberculosis. PLoS Med 2:e381

338. Tailleux L, Schwartz O, Herrmann JL, Pivert E, Jackson M, Amara A, Legres L, Dreher D, Nicod LP, Gluckman JC, Lagrange PH, Gicquel B, Neyrolles O (2003) DC-SIGN is the major Mycobacterium tuberculosis receptor on human dendritic cells. Journal of Experimental Medicine 197:121-127

339. Talbot EA, Perkins MD, Silva SF, Frothingham R (1997) Disseminated bacille Calmette-Guerin disease after vaccination: case report and review. Clin Infect Dis 24:1139-1146

340. Tassaneetrithep B, Burgess TH, Granelli-Piperno A, Trumpfheller C, Finke J, Sun W, Eller MA, Pattanapanyasat K, Sarasombath S, Birx DL, Steinman RM, Schlesinger S, Marovich MA (2003) DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. J Exp Med 197:823-829

341. Taylor PR, Martinez-Pomares L, Stacey M, Lin HH, Brown GD, Gordon S (2005) Macrophage receptors and immune recognition. Annu Rev Immunol 23:901-944

342. Taype CA, Castro JC, Accinelli RA, Herrera-Velit P, Shaw MA, Espinoza JR (2006) Association between SLC11A1 polymorphisms and susceptibility to different clinical forms of tuberculosis in the Peruvian population. Infect Genet Evol 6:361-367

343. Tekkel M, Rahu M, Loit HM, Baburin A (2002) Risk factors for pulmonary tuberculosis in Estonia. Int J Tuberc Lung Dis 6:887-894

344. Thakkinstian A, D'Este C, Eisman J, Nguyen T, Attia J (2004) Meta-analysis of molecular association studies: Vitamin D receptor gene polymorphisms and BMD as a case study. Journal of Bone and Mineral Research 19:419-428

345. Therneau T, Lumey T (2006) Survival analysis including penalised likelihood. R package version 2.20.

346. Thomas DC, Witte JS (2002) Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiol Biomarkers Prev 11:505-512

347. Thwaites GE, Tran TH (2005) Tuberculous meningitis: many questions, too few answers. Lancet Neurol 4:160-170

348. Thye T, Browne EN, Chinbuah MA, Gyapong J, Osei I, Owusu-Dabo E, Niemann S, Rusch-Gerdes S, Horstmann RD, Meyer CG (2006) No associations of human pulmonary tuberculosis with Sp110 variants. J Med Genet 43:e32

349. Tishkoff SA, Williams SM (2002) Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet 3:611-621

350. Toossi T, Johnson JL, Kanost RA, Wu MD, Luzze H, Peters P, Okwera A, Joloba M, Mugyenyi J, Mugerwa JD, Aung H, Ellner JJ, Hirsch CS (2001) Increased replication of HIV-1 at sites of Mycobacterium tuberculosis infection: Potential mechanisms of viral activation. Journal of Acquired Immune Deficiency Syndromes 28:1-8

351. Tosh K, Campbell SJ, Fielding K, Sillah J, Bah B, Gustafson P, Manneh K, Lisse I, Sirugo G, Bennett S, Aaby P, McAdam KP, Bah-Sow O, Lienhardt C, Kramnik I, Hill AV (2006) Variants in the SP110 gene are associated with genetic susceptibility to tuberculosis in West Africa. Proc Natl Acad Sci U S A 103:10364-10368

352. Tosh K, Meisner S, Siddiqui MR, Balakrishnan K, Ghei S, Golding M, Sengupta U, Pitchappan RM, Hill AV (2002) A region of chromosome 20 is linked to leprosy susceptibility in a South Indian population. J Infect Dis 186:1190-1193

353. Trafzer CE (1998) Coughing blood: tuberculosis deaths and data on the Yakama Indian Reservation, 1911-64. Can Bull Med Hist 15:251-276

354. Tufariello JM, Chan J, Flynn JL (2003) Latent tuberculosis: mechanisms of host and bacillus that contribute to persistent infection. Lancet Infectious Diseases 3:578-590

355. Turner D, Choudhury F, Reynard M, Railton D, Navarrete C (2002) Typing of multiple single nucleotide polymorphisms in cytokine and receptor genes using SNaPshot. Hum Immunol 63:508-513

356. Uitterlinden AG, Fang Y, van Meurs JBJ, Pols HAP, van Leeuwen JPTM (2004) Genetics and biology of vitamin D receptor polymorphisms. Gene 338:143-156

357. Uitterlinden AG, Fang Y, van Meurs JBJ, van Leeuwen H, Pols HAP (2004) Vitamin D receptor gene polymorphisms in relation to vitamin D related disease states. Journal of Steroid Biochemistry and Molecular Biology 89-90:187-193

358. Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, Nagy M, Gaede KI, Franke A, Haesler R, Koch A, Lengauer T, Seegert D, Reiling N, Ehlers S, Schwinger E, Platzer M, Krawczak M, Muller-Quernheim J, Schurmann M, Schreiber S (2005) Sarcoidosis is associated with a truncating splice site mutation in BTNL2. Nat Genet 37:357-364

359. Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques 37:226-231

360. van de Vosse E, Hoeve MA, Ottenhoff TH (2004) Human genetics of intracellular infectious diseases: molecular and cellular immunity against mycobacteria and salmonellae. Lancet Infect Dis 4:739-749

361. Van der Ross RE (1993) 100 Questions about Coloured South Africans. UWC Printing Department, Cape Town

362. van der Spuy GD, Warren RM, Richardson M, Beyers N, Behr MA, van Helden PD (2003) Use of genetic distance as a measure of ongoing transmission of Mycobacterium tuberculosis. J Clin Microbiol 41:5640-5644

363. van Helden PD, Moller M, Babb C, Warren R, Walzl G, Uys P, Hoal E (2006) TB epidemiology and human genetics. Novartis Found Symp 279:17-31

364. van Helden PD, Victor T, Warren RM (2006) The "source" of drug-resistant TB outbreaks. Science 314:419-420

365. van Kooyk Y, Appelmelk B, Geijtenbeek TBH (2003) A fatal attraction: Mycobacterium tuberculosis and HIV-1 target DC-SIGN to escape immune surveillance. Trends in Molecular Medicine 9:153-159

366. van Kooyk Y, Geijtenbeek TBH (2003) DC-sign: Escape mechanism for pathogens. Nature Reviews Immunology 3:697-709

367. van Rie A, Beyers N, Gie RP, Kunneke M, Zietsman L, Donald PR (1999) Childhood tuberculosis in an urban population in South Africa: burden and risk factor. Arch Dis Child 80:433-437

368. van Rie A, Warren R, Richardson M, Victor TC, Gie RP, Enarson DA, Beyers N, van Helden PD (1999) Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. N Engl J Med 341:1174-1179

369. van Wyk CW, Groblerrabie AF, Martell RW, Hammond HG (1994) HLA-Antigens in Oral Submucous Fibrosis. Journal of Oral Pathology & Medicine 23:23-27

370. Verver S, Warren RM, Beyers N, Richardson M, van der Spuy GD, Borgdorff MW, Enarson DA, Behr MA, van Helden PD (2005) Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. Am J Respir Crit Care Med 171:1430-1435

371. Verver S, Warren RM, Munch Z, Richardson M, van der Spuy GD, Borgdorff MW, Behr MA, Beyers N, van Helden PD (2004) Proportion of tuberculosis transmission that takes place in households in a high-incidence area. Lancet 363:212-214

372. Verver S, Warren RM, Munch Z, Vynnycky E, van Helden PD, Richardson M, van der Spuy GD, Enarson DA, Borgdorff MW, Behr MA, Beyers N (2004) Transmission of tuberculosis in a high incidence urban community in South Africa. Int J Epidemiol 33:351-357

373. Victor TC, Streicher EM, Kewley C, Jordaan AM, van der Spuy GD, Bosman M, Louw H, Murray M, Young D, van Helden PD, Warren RM (2007) Spread of an emerging Mycobacterium tuberculosis drug-resistant strain in the western Cape of South Africa. Int J Tuberc Lung Dis 11:195-201

374. Victor TC, Warren R, Beyers N, van Helden PD (1997) Transmission of multidrug-resistant strains of Mycobacterium tuberculosis in a high incidence community. Eur J Clin Microbiol Infect Dis 16:548-549

375. Vidal SM, Malo D, Vogan K, Skamene E, Gros P (1993) Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg. Cell 73:469-485

376. Voight BF, Pritchard JK (2005) Confounding from cryptic relatedness in case-control association studies. PLoS Genet 1:e32

377. Wacholder S, Rothman N, Caporaso N (2002) Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiol Biomarkers Prev 11:513-520

378. Walzl G, Beyers N, van Helden P (2005) TB: a partnership for the benefit of research and community. Transactions of the Royal Society of Tropical Medicine and Hygiene 99:S15-S19

379. Walzl G, Beyers N, van Helden P (2005) TB: a partnership for the benefit of research and community. Transactions of the Royal Society of Tropical Medicine and Hygiene 99:S15-S19

380. Wang WY, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6:109-118

381. Warnes G, Leisch F (2005) genetics: Population Genetics. R package version 1.2.0.

382. Warren R, Hauman J, Beyers N, Richardson M, Schaaf HS, Donald P, van Helden P (1996) Unexpectedly high strain diversity of Mycobacterium tuberculosis in a high-incidence community. S Afr Med J 86:45-49

383. Warren R, Richardson M, van der SG, Victor T, Sampson S, Beyers N, van Helden P (1999) DNA fingerprinting and molecular epidemiology of tuberculosis: use and interpretation in an epidemic setting. Electrophoresis 20:1807-1812

384. Warren RM, Victor TC, Streicher EM, Richardson M, Beyers N, Gey van Pittius NC, van Helden PD (2004) Patients with active tuberculosis often have different strains in the same sputum specimen. Am J Respir Crit Care Med 169:610-614

385. Watashi K, Hijikata M, Tagawa A, Doi T, Marusawa H, Shimotohno K (2003) Modulation of retinoid signaling by a cytoplasmic viral protein via sequestration of Sp110b, a potent transcriptional corepressor of retinoic acid receptor, from the nucleus. Mol Cell Biol 23:7498-7509

386. Weiss KM, Clark AG (2002) Linkage disequilibrium and the mapping of complex human traits. Trends Genet 18:19-24

387. Whitfield GK, Remus LS, Jurutka PW, Zitzer H, Oza AK, Dang HT, Haussler CA, Galligan MA, Thatcher ML, Encinas DC, Haussler MR (2001) Functionally relevant polymorphisms in the human nuclear vitamin D receptor gene. Mol Cell Endocrinol 177:145-159

388. Wickremasinghe MI, Thomas LH, O'Kane CM, Uddin J, Friedland JS (2004) Transcriptional mechanisms regulating alveolar epithelial cell-specific CCL5 secretion in pulmonary tuberculosis. J Biol Chem 279:27199-27210

389. Wiedermann CJ, Kowald E, Reinisch N, Kaehler CM, von L, I, Pattison JM, Huie P, Sibley RK, Nelson PJ, Krensky AM (1993) Monocyte haptotaxis induced by the RANTES chemokine. Curr Biol 3:735-739

390. Wilbur AK, Salter KL, Hurtado AM, Hill KR, Stone AC (2007) Vitamin D receptor gene polymorphisms and susceptibility M. tuberculosis in Native Paraguayans. Tuberculosis (Edinb)

391. Wilkinson RJ, Llewelyn M, Toossi Z, Patel P, Pasvol G, Lalvani A, Wright D, Latif M, Davidson RN (2000) Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London: a case-control study. Lancet 355:618-621

392. Williamson C, Loubser SA, Brice B, Joubert G, Smit T, Thomas R, Visagie M, Cooper M, van der Ryst E (2000) Allelic frequencies of host genetic variants influencing susceptibility to HIV-1 infection and disease in South African populations. AIDS 14:449-451

393. Wilson JF, Goldstein DB (2000) Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. Am J Hum Genet 67:926-935

394. Winkler C, Modi W, Smith MW, Nelson GW, Wu XY, Carrington M, Dean M, Honjo T, Tashiro K, Yabe D, Buchbinder S, Vittinghoff E, Goedert JJ, O'Brien TR, Jacobson LP, Detels R, Donfield S, Willoughby A, Gomperts E, Vlahov D, Phair J, O'Brien SJ (1998) Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. Science 279:389-393

395. Wise J (1998) WHO identifies 16 countries struggling to control tuberculosis. BMJ 316:957

396. Wittke-Thompson JK, Pluzhnikov A, Cox NJ (2005) Rational inferences about departures from Hardy-Weinberg equilibrium. Am J Hum Genet 76:967-86

397. World Health Organization (2005) WHO report 2005: Global tuberculosis control - surveillance, planning, financing. Geneva,

398. World Health Organization (2007) WHO Report 2007: Global tuberculosis control - surveillance, planning, financing. Geneva,

399. Yang B, Houlberg K, Millward A, Demaine A (2004) Polymorphisms of chemokine and chemokine receptor genes in type 1 diabetes mellitus and its complications. Cytokine 26:114-121

400. Yun JS, Lazarus AA, Gilbert R (2006) Latent tuberculous infection. Dis Mon 52:441-445

401. Zhang W, Shao L, Weng X, Hu Z, Jin A, Chen S, Pang M, Chen ZW (2005) Variants of the natural resistance-associated macrophage protein 1 gene (NRAMP1) are associated with severe forms of pulmonary tuberculosis. Clin Infect Dis 40:1232-1236

402. Zmuda JM, Cauley JA, Ferrell RE (2000) Molecular epidemiology of vitamin D receptor gene variants. Epidemiol Rev 22:203-217