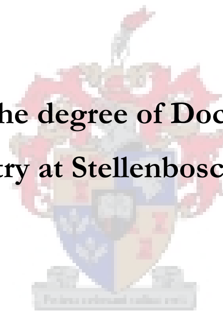
ANALYSIS OF HOST DETERMINING FACTORS IN SUSCEPTIBILITY TO TUBERCULOSIS IN THE SOUTH AFRICAN COLOURED POPULATION

Erika de Wit



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

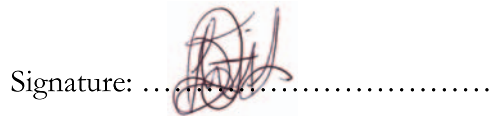
Promoter: Prof. Eileen Hoal

Co-promoter: Prof. Paul van Helden

December 2009

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: …08/12/2009……

Erika de Wit

US number: 13797719

Copyright © 2009 Stellenbosch University
All rights reserved

SUMMARY

The infectious disease tuberculosis (TB) still represents a global threat due to its devastating effect on health and the subsequent high mortality rate. Previous studies have indicated that host genetic factors are implicated in host susceptibility to TB. Since TB is a complex disease, it can be assumed that susceptibility to *M. tuberculosis* has multiple genetic causative factors (as well as environmental causes).

The current study focussed on a number of South African Coloured (SAC) individuals, some of whom were TB cases and others controls. Population substructure was tested in the admixed SAC population as it can be a strong confounding factor for association studies. Our results using the programme STRUCTURE indicated no population substructure in the SAC population. We further investigated the population structure of the SAC group using Affymetrix 500k SNP chip data which showed that the SAC population group has 4 major ancestral components: the Khoesan, European, African and Asian (Indian).

A number of candidate polymorphisms in eight genes, previously indicated to play an important role in TB susceptibility, were tested in case-control associations studies. We found statistically significant associations between *IFNGR1, IL-8, IL-1Ra* and *NRAMP1* polymorphisms and TB susceptibility in the SAC population.

It has become increasingly evident that gene-gene interactions play a far more important part in an individual's susceptibility to a complex disease than single polymorphisms would on their own. The importance of epistasis was clearly identifiable in this study with only four associations found between the individual variants and TB susceptibility, but eight instances of statistically significant gene-gene interactions. A combined data set consisting of 106 variants constructed from our database and also used for gene-gene interaction analysis yielded numerous statistically significant interactions.

The interaction between the genotype of the human host and the bacterial strain genotype was also investigated and yielded interesting results.

Owing to various polymorphisms in several cytokine genes, the protein levels of the main modulators of the immune system, cytokines and chemokines, are changed in several diseases such as infectious diseases and may affect susceptibility or resistance to TB. The functional polymorphisms or haplotype patterns in some of these cytokine genes might be vital for protective immune responses and may serve as biomarkers of protection or susceptibility to TB. The present study investigated 18 cytokines including pro-inflammatory, anti-inflammatory and chemokine factors in healthy (mantoux positive or negative) children using the Linco-plex immunoassay, and investigated potential interactions.

The basic research will one day contribute to personalised genetics which may benefit infectious diseases such as TB. If individuals can be identified as potentially more vulnerable, they may require different vaccination strategies, a higher index of suspicion if exposed to TB, and prophylactic treatment.

OPSOMMING

Die infektiewe siekte tuberkulose (TB) is steeds ‘n gevaar wat die hele wêreld bedreig weens die groot impak op gesondheid en die gevolglike hoë mortaliteit. Vorige studies het bevind dat die gasheer se genetiese faktore wel betrokke mag wees by die gasheer se vatbaarheid vir TB. Aangesien TB ‘n komplekse siekte is, kan dit aanvaar word dat vatbaarheid tot *M. tuberculosis* veelvuldige genetiese oorsaaklike faktore (sowel as omgewingsoorsake) het.

Hierdie studie het gefokus op ‘n aantal Suid-Afrikaanse Kleurling (SAC) individue, waarvan sommige TB pasiënte en ander kontroles was. Die gemengde SAC populasie is getoets vir populasie-stratifikasie, aangesien stratifikasie ‘n sterk verwarrende invloed op pasiënt-kontrole studies kan hê. Ons resultate is verkry met behulp van die program STRUCTURE en het aangedui dat daar geen populasie sub-struktuur tussen die pasiënte en kontroles was nie. Ons het ook die populasiesamestelling van die SAC groep ondersoek met data verkrygbaar van die Affymetrix 500k enkel nukleotied polimorfisme mikroskyfie. Hierdie data het getoon dat die SAC populasie uit 4 hoof voorouerlike komponente bestaan naamlik die Khoesan, Europeërs, Afrikane en Asiate (Indiërs).

‘n Aantal kandidaat polimorfismes in agt gene, wat volgens vorige studies ‘n belangrike rol in TB vatbaarheid te speel, was in hierdie pasiënt-kontrole assosiasie studie bestudeer. Ons het statistiese beduidende verwantskappe tussen *IFNGR1*, *IL-8*, *IL-1Ra* en *NRAMP1* polimorfismes en TB vatbaarheid in die SAC populasie gevind.

Dit het al hoe meer duidelik geword dat geen-geen interaksies ‘n baie belangriker rol in ‘n individu se vatbaarheid vir ‘n komplekse siekte speel as enkel polimorfismes op hul eie. Die belang van epistase kon duidelik in hierdie studie geïdentifiseer word met slegs vier assosiasies wat tussen die individuele variante en TB vatbaarheid gevind is, in vergelyking met agt statisties beduidende geen-geen interaksies. ‘n Gekombineerde datastel wat uit ons databasis saamgestel is en wat 106 variante bevat is ook in ‘n aparte geen-geen interaksie analise gebruik, wat verskeie statisties beduidende interaksies getoon het.

Die interaksie tussen die menslike gasheer genotipe en die bakteriese stam genotipe is ook in hierdie studie ondersoek en het interessante resultate opgelewer.

Veranderde proteïen uitdrukking van die hoofmoduleerders van die immuunsisteem, sitokine en chemokine, kom voor in verskeie siektes soos infektiewe siektes weens verskillende polimorfismes in verskeie sitokien-gene. Sulke polimorfismes kan ook vatbaarheid vir of weerstandigheid teen TB beïnvloed. Die funksionele polimorfismes of haplotipe patrone in sommige van hierdie sitokien-gene mag noodsaaklik wees vir beskermende immuunresponse en mag ook as biomerkers vir beskerming teen of vatbaarheid vir TB dien. Hierdie studie het 18 sitokiene (insluitend pro-inflammatoriese-, anti-inflammatoriese- en chemokiene faktore), sowel as potensiële interaksies in gesonde (mantoux positiewe of negatiewe) kinders, ondersoek met behulp van die Linco-plex immuno-analise.

Hierdie basiese navorsing sal eendag in die toekoms bydrae tot persoonlike genetiese analises wat tot voordeel kan wees vir infektiewe siektes soos TB. Indien individue as potensieël meer vatbaar vir TB geïdentifiseer kan word, kan sulke persone ander vaksineringstrategieë sowel as voorkomende behandeling vereis.

This thesis is dedicated to my husband, Renier de Wit, my parents, Fred and Sophia Truter and Leon and Rina de Wit

ACKNOWLEDGEMENTS

I would like to thank numerous people and institutions for their contribution and assistance to this thesis:

Firstly my promoter Prof. Eileen Hoal and co-promoter Prof. Paul van Helden. Thank you for accepting me, not just as your PhD student but also as part of your little Tygerberg family. Thank you for always being there with advice or any help that I might have needed. Whenever I had a problem I always knew I could come to you and always walk out there feeling much better. Thank you for accepting me for who I am with my "not so bright moments" and the fact that I can never sit still. The lessons I have learned from you are of immeasurable value. I am extremely grateful to both of you.

To everyone in the Division of Molecular Biology and Human Genetics, in particular Lindsey Adams, Kim Stanley, Marna Esterhuyse, Marisa Tait, Muneeb Salie, Chandre Wagman and Annelle Muller. Thank you for your friendship, support, interesting discussions and all our essential tea-time trips to the cafeteria. Thank you to Rene Veikondis for the microsatellite analysis, Prof. Rob Warren for the bacterial strain data and Dr. Gian van der Spuy for all the computer and database assistance.

Dr. Marlo Möller, thank you for not being just a colleague, but a very special friend. Thank you for never being too busy, to help, give advice or simply listen. "*Friends are like stars, they come and go, but the ones that stay are the ones that glow" - Roxy Quiksilver-*

Special thanks to André Loxton, Nonhlanhla Nene and Novel Chegou my PhD partners-in-crime. Thank you for all the emotional support and discussions. You are three very special people and I wish you all the best.

The National Research foundations (NRF), the South African Medical Research Council (MRC), the Division of Molecular Biology and Human Genetics and Stellenbosch University for financial support during this project.

The individuals that participated in this study, without them this research would not have been possible.

My uncle Pierre Smith and sister Lizana Truter. Uncle baie dankie vir al jou ondersteuning deur die jare ek waardeer dit ongelooflik baie! Dankie vir al jou belangstelling, ondersteuning en die impak wat jy tot dusver op my lewe gehad het. Sus, baie dankie vir elke sms om sterkte te sê en vir jou invloed op my lewe. Was dit nie vir die feit dat ek my man moes staan teen jou nie het ek nie die deursettingsvermoë gehad om 'n PhD te voltooi nie.

My parents Leon and Rina de Wit. Baie dankie vir al julle liefde en ondersteuning. Dankie vir elke oproep, e-pos of sms. Dankie dat julle my laat voel soos julle eie dogter en dat julle so lief is vir my. Ek is baie lief vir Pappa en Mamma en waardeer julle ongelooflik baie!.

My parents Fred and Sophia Truter. Baie dankie vir alles wat Pappa en Mamma vir my gedoen het. Dankie vir die opvoeding, die geleenthede en vir wie ek vandag is. Dankie vir al julle oneindige liefde, ondersteuning, begrip, sms'e en e-posse. Ek is bevoorreg om ouers soos julle te hê. Ek is baie lief vir Pappa en Mamma!

To my husband Renier de Wit. Engel ek weet nie eers waar om te begin nie want jy is net ongelooflik. Dankie vir al jou ondersteuning en aanmoediging van die 1ste jaar af. Dankie vir al jou liefde en vir die wonderlike man wat jy vir my is. Ek is baie lief vir jou! "*Twee vaar beter as een. Hulle inspanning kom tot iets. As die een val, kan die ander hom ophelp. Maar as een val wat alleen is, is daar niemand om hom op te help nie. As twee langs mekaar slaap, word hulle warm, maar hoe sal een wat alleen is, warm word? Een alleen kan oorweldig word, twee saam kan weerstand bied. 'n Driedubbele tou breek nie maklik nie". (Prediker 4:9-12).*

Last but not least to God Almighty for giving me the strength and the ability to finish this degree. You were with me every step of the way and every second of every day as You are my whole life. I love You and thank you for everything.

See the end of chapters for additional acknowledgements.

TABLE OF CONTENTS

Declaration……………………………………………………………………………...........ii

Summary…………………………………………………………………………….........iii

Opsomming………………………………………………………………………….....v

Acknowledgments……………………………………………………………………....viii

Table of contents……………………………………………………………………………x

List of abbreviations and symbols…………………………………………….........xiv

List of figures…………………………………………………………………….....xx

List of tables…………………………………………………………….........xxvii

List of manuscripts prepared for submission...xxxi

CHAPTER ONE

INTRODUCTION

1.1 Tuberculosis………………………………………………………………...............2

1.1.1 The origin and history of tuberculosis…………………………………….........3

1.1.2 The history of TB in South Africa…………………………………….........4

1.2 Mycobacterium tuberculosis………………………………………………………….5

1.2.1 The *M. tuberculosis* genome…………………………………………….........6

1.2.2 TB strain families……………………………………………………….......6

1.3 Pathogenesis of TB……………………………………………………………........8

1.3.1 TB and the immune system……………………………………………........9

1.3.2 Recurrent TB……………………………………………………………........11

1.4 Vaccination……………………………………………………………………......11

1.5 Diagnosis…………………………………………………………………………....12

1.6 Treatment……………………………………………………………………………...13

1.7 Drug-resistant TB……………………………………………………………………...14

1.8 TB and the human immunodeficiency virus (HIV)………………………………….15

1.9 Host genetic susceptibility to TB……………………………………………..........17

1.9.1 Candidate gene approach…………………………………………………..........17

1.9.2 Linkage analysis…………………………………………………….................18

1.9.3 Case-control association studies……………………………………………........20

1.9.4 Population stratification……………………………………………….............21

1.9.5 TB and previous genetic susceptibility studies……………………………...........21

1.9.6 Genes investigated in this study……………………………………………..........23

CHAPTER TWO

GENETIC STRUCTURE OF THE SOUTH AFRICAN COLOURED POPULATION

2.1 Introduction…………………………………………………………………........39

2.1.1 Heterogeneity in association studies………………………………………........40

2.1.2 Confounding…………………………………………………………………...41

2.1.3 Genomic adjustment…………………………………………………………...44

2.1.3.1 Non-model based approach……………………………………………...44

2.1.3.2 Model based approach to population structure…………………………..44

2.1.4 Markers……………………………………………………………………........45

2.2 Structure……………………………………………………………………….........46

2.2.1 Problems with the interpretation of K…………………………………………47

2.3 African populations……………………………………………………………47

2.3.1 African origin of modern humans……………………………………………...48

2.3.2 South African population……………………………………………………....48

2.4 Materials and Methods………………………………………………….............53

2.4.1 Study population………………………………………………………………53

2.4.2 Population stratification………………………………………………………..54

2.4.3 Population structure……………………………………………………………55

2.5 Results………………………………………………………………………….....57

2.5.1 Population stratification………………………………………………………..57

2.5.2 Population structure……………………………………………………………59

2.6 Discussion…………………………………………………………......................68

2.7 Acknowledgements...71

CHAPTER THREE

ASSOCIATION STUDIES IN EIGHT CANDIDATE GENES

3.1 Introduction…………………………………………………………………….....73

3.1.1 Surfactant protein D (SP-D)…………………………………………………74

3.1.2 Mannose binding lectin (MBL)………………………………….....................75

3.1.3 Interferon-gamma (IFN-γ) and Interferon-gamma receptor 1 (IFNGR1)……......76

3.1.4 Interleukin 8 (IL-8)…………………………………………………….............78

3.1.5 Regulated upon Activation, Normal T-cell Expressed and Secreted chemokine (RANTES), also known as CCL5……………………………………………........79
3.1.6 Natural resistance-associated macrophage protein 1 (NRAMP1) also known as Solute carrier family 11a member 1 (SLC11A1)……………………………........81
3.1.7 Interleukin-1 receptor antagonist (IL-1Ra)……………………………………....83
3.2 Materials and Methods……………………………………………………..............84
3.2.1 SP-D…………………………………………………………………………...84
3.2.2 MBL…………………………………………………………………………....85
3.2.3 IFN-γ and IFNGR1……………………………………………………………87
3.2.4 IL-8…………………………………………………………………………….88
3.2.5 RANTES………………………………………………………………………89
3.2.6 NRAMP1………………………………………………………………………89
3.2.7 IL-1Ra…………………………………………………………………………90
3.3 Results……………………………………………………………………………....90
3.3.1 SP-D……………………………………………………………………........90
3.3.2 MBL…………………………………………………………………….........90
3.3.3 IFN-γ and IFNGR1…………………………………………………………....93
3.3.4 IL-8…………………………………………………………………………….94
3.3.5 RANTES………………………………………………………………………95
3.3.6 NRAMP1………………………………………………………………………96
3.3.7 IL-1Ra…………………………………………………………………………97
3.9 Discussion…………………………………………………………………..........101

CHAPTER FOUR

GENE-GENE AND GENE- *(M. TUBERCULOSIS)* STRAIN INTERACTIONS IN TUBERCULOSIS SUSCEPTIBILITY

4.1 Introduction…………………………………………………………………….105
4.2 Materials and Methods…………………………………………………………….108
4.2.1 Study population……………………………………………………………...108
4.2.2 Genotyping…………………………………………………………………...109
4.2.3 Bacterial strains……………………………………………………………….109
4.2.4 Statistical analysis……………………………………………………………...110
4.3 Results……………………………………………………………………………..111

4.3.1 Association studies……111
4.3.2 Optimal model……113
4.3.3 Gene-gene interactions……114
4.3.4 Host genotype and bacterial genotype analysis……118
4.4 Discussion……123
4.5 Acknowledgements……126

CHAPTER FIVE

MEASUREMENT OF CYTOKINES BY MULTIPLEX ANALYSIS AMONG CHILDREN FROM A HIGH TB INCIDENCE AREA

5.1 Introduction……128
5.2 Materials and Methods……130
5.2.1 Study cohort……130
5.2.2 Sample collection and antigen stimulation……131
5.2.3 Measurement of cytokines using the Linco-plex immunoassay……131
5.2.4 Statistical analysis……132
5.3 Results……132
5.3.1 Mantoux results……132
5.3.2 Age……138
5.3.3 Correlation between BCG and ESAT-6 for all the cytokines……140
5.3.4 Comparison of cytokine production levels……142
5.3.5 Factor analysis……144
5.3.6 Comparison between ELISA and Lincoplex data……147
5.4 Discussion……150
5.5 Acknowledgements……156

CHAPTER SIX

CONCLUDING REMARKS

6.1 Potential limitations of the study……161
6.2 Future directions……162
6.3 Conclusion……166

REFERENCES……167

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviations are listed in alphabetical order.

A

A	Adenine nucleotide
AIC	Akaike's criterion
AIDS	Acquired Immune Deficiency Syndrome
AIMs	Ancestry informative markers
AML	Acute myeloid leukemia
Arg	Arginine
ARMS	Amplification refractory mutation system
Asp	Aspartic acid

B

BALF	Bronchoalveolar lavage fluid
BCG	Bacillus Calmette-Guérin
bp	Base pair
BRLMM	Bayesian Robust Linear Model with Mahalanobis distance classifier

C

C	Cytosine nucleotide
CD	Cluster designation (cluster of differentiation)
CDCV	Common disease common variant
CFP-10	Culture filtrate protein-10
CNVs	Copy number variations
CRD	Carbohydrate recognition domain
Cys	Cysteine
CYP3A4	Cytochrome P450 3A4

D

DC	Dentritic cell
DC-SIGN	Dentritic cell-specific ICAM-3 grabbing nonintegrin
DM	Dynamic model
DNA	Deoxyribonucleic acid

dNTPs	Deoxynucleotide triphosphates
DOTS	Directly observed treatment short course
DRD2	Dopamine D2 receptor

E

ELISA	Enzyme-linked immunosorbant assay
ENA-78	Epithelial cell-derived neutrophil activating protein 78
ESAT-6	6-kDa early secretory antigenic target

F

Fe	Iron
For	Forward

G

G	Guanine nucleotide
GAG	Glyco-amino glycans
GC	Genomic control
Glu	Glutamic acid
Gly	Glycine
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GROα	Growth-related oncogene α
GWAS	Genome-wide association studies

H

HGDP	Human Genome Diversity Project
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HWE	Hardy-Weinberg equilibrium

I

IFN-γ	Interferon-gamma
IFNGR1	Interferon-gamma receptor 1
IL	Interleukin

IL-1Ra	Interleukin-1 receptor antagonist
IL-1α	Interleukin-1 alpha
IL-1β	Interleukin-1 beta
IP-10	Interferon-inducible protein -10

K

K	Populations clusters
Kb	kilo base
kDA	Kilodalton

L

LAM	Latin American and Mediterranean
LCA	Latent-class approaches
LCC	Low copy clade
LD	Linkage disequilibrium
LG	Löwenstein-Jensen medium

M

MBL	Mannose-binding lectin
M. bovis	*Mycobacterium bovis*
MBP	Mannose-binding protein
MCP-1	Macrophage chemotactic protein-1
MDR-TB	Multidrug-resistant tuberculosis
Met	Methionine
$MgCl_2$	Magnesium Chloride
MHC	Major histocompatibility complex
MIP-1α	Macrophage inhibitory protein -1 alpha
MIRU-VNTR	Mycobacterial interspersed repetitive unit-variable number tandem repeat
ml	Milliliter
mM	Millimolar
mtDNA	Mitochondrial DNA
M. tuberculosis	*Mycobacterium tuberculosis*

MΦ	Macrophage

N

n	Number of samples
ng	Nanogram
NH_2	Amino group
NK	Natural killer
NOS2A	Nitric oxide synthase 2A gene
NRAMP1	Natural resistance-associated macrophage protein 1
nsSNP	Nonsynonymous single nucleotide polymorphism
NUI	National University of Ireland

O

OR	Odds ratio

P

P	Probability value
PAMP	Pathogen associated molecular pattern
PCA	Principal component analysis
PCR	Polymerase chain reaction
PGG	Principal genetic groups
$(P[K\|X])^a$	Posterior probability
PPD	Purified protein derivative (of *M. tuberculosis*)
PTB	Pulmonary tuberculosis
PTST	Persistently negative tuberculin skin test

R

r	Correlation coefficient
RANTES	Regulated upon activation, normal T-cell expressed and secreted
RAO	Recent African origin
Rev	Reverse
RFLP	Restriction fragment length polymorphism

S

SA	Structure association
SAC	South African Coloured(s)
SB	Sodium boride
SCD40L	Soluble CD40 ligand
Ser	Serine
SLC11A1	Solute carrier family 11a member 1
SNP	Single nucleotide polymorphisms
SP-D	surfactant protein D
SP110	Nuclear body protein / speckled 110 kDa protein
sSNP	Synonymous single nucleotide polymorphism
STRs	Short tandem repeats

T

T	Thymine nucleotide
TB	Tuberculosis
TBE	Tris–boric acid-ethylene diamine tetra-acetic acid buffer
TBM	Tuberculosis meningitis
TDI-FP	Template-directed dye-terminator incorporation with fluorescence polarization detection
TDT	Transmission disequilibrium test
Thr	Threonine
Th1	T helper 1
Th2	T helper 2
TLR	Toll-like receptor
T_m	Melting temperature
TNF	Tumor necrosis factor
TST	Tuberculin skin test

U

U	Units
UCT	University of Cape Town
US	University of Stellenbosch

UTR	Untranslated region
UV	Ultra violet
UWC	University of Western Cape
V	
V	Volt
VDR	Vitamine D receptor gene
VNTR	Variable number of tandem repeats
W	
WBA	Whole blood assay
WHO	World Health Organization
w/v	Weight per volume
X	
XDR-TB	Extensively drug-resistant tuberculosis
3’	3 prime end
5’	5 prime end
%	Percent
^{o}C	Degrees Celsius
µl	microlitre
µM	micro molar
µg	microgram
χ^2	Chi square

LIST OF FIGURES

Fig. 1.1: The worldwide incidence of tuberculosis for 2007 from the WHO Report....... 2

Fig. 1.2: The three PGG assigned to the eight different clusters based on the sSNP 8

Fig. 1.3: The complex interactions between the host, environment and Mycobacteria (adapted from (Casanova & Abel,2002)... 9

Fig. 1.4: Diagram explaining the two different types of immune responses (Th-1 and Th-2) (Ducati *et al.,* 2006; Meya & McAdam, 2007). (Note: this is somewhat oversimplified, but the Th1/Th2 balance is a widely accepted idea)................. 10

Fig. 1.5: The five fundamental elements of the DOTS strategy... 13

Fig. 1.6: The worldwide estimated HIV prevalence in new TB cases for 2007 from..... 16

Fig. 1.7: Strategies for identifying susceptibility genes, adapted from (Alcaïs & Abel, 2004).. 18

Fig. 1.8: Genetic involvement in the TB disease process. A simplified representation of different outcomes after *M. tuberculosis* infection and some of the genes that may be involved at various stages. The bacteria enter the respiratory system of the host via inhaled droplets and are engulfed by macrophages (MΦ) and dendritic cells. There are three potential outcomes after inhalation of *M. tuberculosis*: i) Infection develops into active TB. ii) *M. tuberculosis* is immediately killed by the pulmonary immune system. iii) Infection does not progress to active disease, because the bacteria are contained in granulomas. During active TB, *M. tuberculosis* can disseminate from granulomas. (Adapted from (Kaufmann & McMichael,2005).. 22

Fig. 2.1: Classic confounding and population stratification. In confounding, an exposure of interest correlated with the true risk factor but not truly causal of

disease, can be misleading seen to be associated with risk of disease. In population structure, the genotype of interest correlated with the true risk factor because both are correlated with ethnicity it can be incorrectly seen to be associated with risk of disease (Wacholder *et al.*, 2002) 42

Fig. 2.2: The recent African origin model, multiregional model and assimilation model to explain the evolution of modern humans 48

Fig. 2.3: The traditional distribution of the different population groups in South Africa (Adapted from Lombard *et al.*, 2006). 49

Fig. 2.4: Bar plots for estimated membership from structure for *K* = 1- 6. Each vertical line represents an individual on the X-axis while the Y-axis indicated the individual's estimated membership fraction in the different clusters (*K*=1-6). 58

Fig. 2.5: Estimates of the number of ancestral populations (K) for the SAC and combined HapMap and HGDP samples under an admixture model using the STRUCTURE software. The estimated probability of the data given the model is plotted against increasing K for each of the subsets of SNP data used (see Methods) 61

Fig. 2.6: Proportion of each individual's ancestry for the number of ancestral populations from k = 2 to the estimated number of ancestral populations with greatest probability (Fig. 2.5D). Plots shown are for unlinked Ancestry Informative Markers. Population labels are as in Table 2.1.............................. 62

Fig. 2.7: Proportion of each individual's ancestry for the number of ancestral populations from k = 2 to the estimated number of ancestral populations with greatest probability (Fig. 2.5B). Plots shown are for unlinked random SNPs. 62

Fig. 2.8: Proportion of each individual's ancestry for the number of ancestral populations from k = 2 to the estimated number of ancestral populations with greatest probability (Fig. 2.5A). Plots shown are for random linked SNPs..... 63

Fig. 2.9: Proportion of each individual's ancestry for the number of ancestral populations from k = 2 to the estimated number of ancestral populations with greatest probability (Fig. 2.5C). Plots shown are for linked Ancestry Informative Markers. 64

Fig. 2.10: Proportion of each individual's ancestry (K = 4) sorted (in ascending order from left to right) by the proportion of ancestry for each of the major contributions to the SAC. 65

Fig. 2.11: Proportion of each individual's ancestry derived using the linkage model in STRUCTURE for the optimal number of ancestral populations (K=4).......... 66

Fig. 2.12: Plot of the first four eigen vectors in the PCA analysis of SAC, HapMap 3 and HGDP populations selected as putative ancestral populations for the South African Coloured population.. 67

Fig. 3.1: *SP-D* and the polymorphism investigated in this study....................................... 74

Fig. 3.2: The structure of MBL. Adapted from (Turner & Hamvas, 2000).................... 75

Fig. 3.3: *MBL* and the polymorphisms investigated in this study. 76

Fig. 3.4: [A] *IFN-γ* and [B] *IFNGR1* and the polymorphisms investigated in this study. 77

Fig. 3.5: *IL-8* and the polymorphism (+100 C/T) investigated in this study. 79

Fig. 3.6: *RANTES* and the polymorphism investigated in this study. 79

Fig. 3.7: RANTES attracts inflammatory cells from the bloodstream to the site of infection. Leukocytes roll along the endothelium via interactions with selectins. Upon inflammation, macrophages in the bloodstream release interleukin-1 (IL-1) and tumour necrosis factor (TNF) which in turn promotes

the release of chemokines (including RANTES). The various cell types follow the chemotactic gradient to the infection site. In addition RANTES binds to glycoaminoglycans (GAG) on the endothelial cell surface. Once the T lymphocytes reach the injury site they produce large amount of RANTES. Adapted from (Krensky & Ahn, 2007) 80

Fig. 3.8: The NRAMP1 protein is a transmembrane cation iron transporter. It is recruited to the phagosome of the macrophage after both bacteria and iron are ingested. NRAMP1 transports iron and protons such as H^+ out of the phagosome into the macrophage cytoplasm. Adjusted from (Gruenheid et al., 1997). 81

Fig. 3.9: *NRAMP1* and the polymorphisms investigated in this study. 82

Fig. 3.10: The *IL-1* gene cluster: Interleukin -1 receptor 1 (*IL-1R1*), interleukin-1 alpha and beta (*IL-1a* & *IL-1β*) and interleukin-1 receptor antagonist (*IL-1Ra*). *IL-1Ra* is expanded to show the polymorphism investigated in this study. Adapted from (Walley et al., 2004). 83

Fig. 3.11: An example of how an ARMS-PCR works (Strachan & Read, 1999) 87

Fig. 3.12: Restriction enzyme analysis of the Met11Thr SNP in *SP-D*............................. 91

Fig. 3.13: Restriction enzyme analysis of the 3 SNPs located in exon 1 of *MBL*, [A] PCR product with MBLexon1 primer set, [B] PCR product with MBL D allele primer set, [C] Digestion with *Ban*I (B allele), [D] Digestion with *Mbo*II (C allele), [E] Digestion with *Mlu*I (D allele). 92

Fig. 3.14: The *IFN-γ* +874 A/T polymorphism detected using the ARMS method........ 93

Fig. 3.15: Detection of the polymorphism +1050 G/T in *IFNGR1*: [A] PCR product followed by, [B] restriction enzyme analysis.. 93

Fig. 3.16: Detection of all three genotypes of the +100 C/T *IL-8* polymorphism........... 95

Fig. 3.17: The detection of all three genotypes of the *RANTES* - 403 A/G polymorphism. 96

Fig. 3.18: Examples of an electropherogram sample run for microsatellite analysis of the 5'$(GT)_n$ and 3'UTR variants in the *NRAMP1* gene. [A] The 5'$(GT)_n$ variant: homozygote 118/118 heterozygote 118/120, homozygote 120/120 and heterozygote 120/122. [B] The 3'UTR variant: homozygote N/N, heterozygote N/D and homozygote D/D. (N= no deletion, D = deletion). . 98

Fig. 3.19: All alleles obtained with the microsatellite analysis of the 86bp VNTR in the *IL-1Ra* gene. The grey areas are the bins set up for the different alleles. 99

Fig. 4.1: A simplified illustration, explaining the process from sputum collection to the *M. tuberculosis* strain information. 109

Fig. 4.2: Graphical representation of modelled interactions 117

Fig. 4.3: An illustration of all pair-wise interaction analysis for the combined dataset adjusted for age and gender. ....Appendix A

Fig. 4.4: Graphical representations of significant trait-genotype associations 122

Fig. 5.1: A simplified version of some downstream effects after infection with *M. tuberculosis*... 129

Fig. 5.2: The distribution of the Mantoux readings for all the study participants 133

Fig. 5.3: Correlation between Mantoux readings > 0 mm and IFN-γ production [pg/ml] for BCG stimulated samples, ESAT-6 stimulated samples and unstimulated samples. r = Spearman correlation. 135

Fig. 5.4 A: IFN-γ production [pg/ml] in individuals with Mantoux reading 0 mm vs Mantoux reading > 0 mm for BCG stimulated samples, ESAT-6 stimulated samples and unstimulated samples. The vertical bars denote 0.95 confidence intervals. 136

Fig. 5.4 B: IL-8 production [pg/ml] in individuals with Mantoux reading 0 mm vs Mantoux reading > 0 mm for BCG stimulated samples, ESAT-6 stimulated samples and unstimulated samples. The vertical bars denote 0.95 confidence intervals.. 136

Fig. 5.5: IL-4 production [pg/ml] with Mantoux reading 0 mm vs Mantoux reading > 0 mm for BCG stimulated samples, ESAT-6 stimulated samples and unstimulated samples. The vertical bars denote 0.95 confidence intervals. .. 137

Fig. 5.6 A: Correlation between IFN-γ production [pg/ml] and age for BCG stimulated samples, ESAT-6 stimulated samples and unstimulated samples. 139

Fig. 5.6 B: Correlation between IL-4 production [pg/ml] and age for BCG stimulated samples, ESAT-6 stimulated samples and unstimulated samples. 139

Fig. 5.7: Correlation between BCG and ESAT-6 stimulated responses for [A] IFN-γ, [B] IL-8, [C] IL-1Ra and [D] IL-4 141

Fig. 5.8: A summary of the Spearman correlation matrix, portraying the r values of major correlations found. 142

Fig. 5.9: A matrix of the Spearman correlation coefficients (r values) between antigen-specific cytokine production levels of all 18 cytokines. The upper quadrant portrays the result of stimulation by the ESAT-6 antigen (blue) and the lower quadrant by BCG antigen (green). The values in red text indicate statistically significant correlations with values ≥ 0.7 in yellow blocks and values ≥ 0.8 are shown in red blocks. 149

Fig. 5.10: Two scree plots of [A] BCG stimulated and [B] ESAT-6 stimulated samples of all the multiple variables. The blue line represents the principal component (PC) of the actual data; the red line represents the PC of the stimulated data and the green line represents the PC of the resampled data 145

Fig. 5.11: Histograms of [A] BCG stimulated and [B] ESAT-6 stimulated samples for IFN-γ levels. The blue bars represent the Lincoplex data and the green bars ELISA data. 149

Fig. 5.12: Correlation between [A] BCG stimulated and [B] ESAT-6 stimulated samples of IFN-γ. 149

Fig. 5.13: The role of IL-17A, IL-17E and IL-17F in the expression of airway inflammation. Dentritic cell (DC), macrophage (MΦ), Growth-related oncogene α (GROα), Epithelial cell-derived neutrophil activating protein 78 (ENA-78) (Kawaguchi *et al.*, 2004)….. 153

LIST OF TABLES

Table 1.1: Factors responsible for the resumption of TB 14

Table 1.2: Evidence suggesting a strong association between TB and HIV in Africa 16

Table 1.3: A comparison between linkage-type studies and association studies (Lander & Schork, 1994) 18

Table 1.4: Chromosomal regions identified by genome-wide linkage studies of TB 20

Table 1.5: Articles reporting significant linkage or association between susceptibility genes and TB (Burgner et al., 2006). 26

Table 1.6: Articles reporting no significant linkage or association between possible 33

Table 2.1: Explanations, other than population stratification, for false-positive reports in studies with unrelated controls (Wacholder *et al.*, 2002). 43

Table 2.2: Comparison between the model-free genomic control approach (GC) and the model-based structured association (SA) approach. 46

Table 2.3: Putative ancestral populations included in population structure analysis of South African Coloureds (SAC). 56

Table 2.4: Alternate SNP marker selection strategies employed in the STRUCTURE analysis. 56

Table 2.5: The estimated posterior probabilities for *K* in the South African Coloured samples. 59

Table 2.6: Mean and standard error on proportion of ancestry for each of four populations contributing to South African Coloureds (SAC), for admixture

and linkage models. Mean and standard errors are from multiple STRUCTURE analysis (n=5) with the number of ancestral populations (K) equal to 4. 60

Table 3.1: Primers used in the PCR amplification of variants investigated in this study 86

Table 3.2: Genotypic and allelic distribution of the Met11Thr SNP in South African Coloured TB patients compared to controls. A single *P*-value is given for genotype and allele association, adjusted for age and gender........................... 91

Table 3.3: Genotypic and allelic distribution of the *MBL* exon 1 variant. 91

Table 3.4: Genotypic and allelic distribution of the *IFN-γ* variant in TB cases *vs* controls and TBM cases *vs* controls, adjusted for age and gender. 94

Table 3.5: Genotypic and allelic distribution of the *IFNGR1* variant in TB cases *vs* controls and TBM cases *vs* controls, adjusted for age en gender..................... 94

Table 3.6: Genotypic and allelic distribution of the *IL-8* variant in TB cases *vs* controls and TBM cases *vs* controls, adjusted for age and gender. 95

Table 3.7: Genotypic and allelic distribution of the -403 A/G *RANTES* variant in TB cases *vs* controls and TBM cases *vs* controls, adjusted for age and gender. ... 96

Table 3.8: Genotypic and allelic distribution of the *NRAMP1* 5' $(GT)_n$ repeat in TB cases *vs* controls and TBM cases *vs* controls. 97

Table 3.9: Genotypic and allelic distribution of the *NRAMP1* 3'UTR deletion in TB cases *vs* controls and TBM cases *vs* controls, adjusted for age and gender. ... 97

Table 3.10: *IL-1Ra* allele frequencies among TB cases and controls in this South African population 100

Table 3.11: *IL-1Ra* 86bp VNTR genotype and allele frequencies among TB and TBM cases and controls. 100

Table 4.1: Genes associated with TB 106

Table 4.2: Case-control analysis of polymorphisms and TB in the South African Coloured population. 112

Table 4.3: *IL-1Ra* 86bp VNTR genotype and allele frequencies among TB cases and controls. 113

Table 4.4: Summary of the optimal model for logit(TB), selected from the model...... 114

Table 4.5: Pairs of variants showing significant interactions in models adjusted for age and gender. For variants with more than 2 alleles, the specific allele in the model is given. 114

Table 4.6: Pairs of variants showing highly significant interactions ($P < 0.01$) in models adjusted for age and gender. The gene name is given followed by the rs number for the SNP and the specific allele for variants with more than 2 alleles. 120

Table 4.7: *P*-values for dichotomous trait-genotype (allele count) association tests, adjusted for age. 121

Table 5.1: Cytokines included in the 18-plex Lincoplex assay 132

Table 5.2: Two line plots [A] BCG stimulated and [B] ESAT-6 stimulated samples of all the multiple variables 145

Table 5.3: Factor loadings for the BCG stimulated samples of which all were Varimax normalized. Cytokine loadings ≥ 0.7 are marked in red, > 0.5 in yellow and those present in two factors are marked in orange. 146

Table 5.4: Factor loadings for the ESAT-6 stimulated samples of which all were Varimax normalized. Cytokine loadings ≥ 0.7 are highlighted in red, > 0.5 in yellow and those present in two factors are highlighted in orange. 147

Table 6.1: Summary of al the polymorphisms investigated in the case-control study in the South African Coloured population. 159

LIST OF MANUSCRIPTS ACCEPTED, SUBMITTED OR PREPARED FOR SUBMISSION

Erika de Wit, Wayne Delport, Emile Chimusa Rugamika, Ayton Meintjes, Marlo Möller, Paul D. van Helden, Cathal Seoighe, & Eileen G. Hoal: GENOME-WIDE ANALYSIS OF THE STRUCTURE OF THE SOUTH AFRICAN COLOURED POPULATION IN THE WESTERN CAPE.(*Submitted to Human Genetics*)

Erika de Wit, Lize van der Merwe, Paul D. van Helden & Eileen G. Hoal: GENE-GENE INTERACTIONS IN TUBERCULOSIS SUSCEPTIBILITY IN A SOUTH AFRICAN POPULATION. (*Submitted to Human Molecular Genetics*)

Erika de Wit, Marlo Möller & Eileen G. Hoal: GENETIC PERSPECTIVES OF TUBERCULOSIS IN SOUTHERN AFRICA. Genomic variation and genetic disorders of developing countries. In print, Oxford University Press, 2009 (*Submitted*)

Marlo Möller, **Erika de Wit** & Eileen G. Hoal (2009) PAST, PRESENT AND FUTURE DIRECTIONS IN HUMAN GENETIC SUSCEPTIBILITY TO TUBERCULOSIS. (*FEMS Immunol Med Microbiol Accepted)*

CHAPTER 1

INTRODUCTION

1.1 TUBERCULOSIS

Tuberculosis (TB) is an important disease in the recent history of mankind due to its devastating effect on health and the subsequent high mortality rate, currently throughout the world. This infectious disease remains a scourge especially in low- and middle - income countries and is still responsible for a large number of deaths each year (Styblo, 1988). In 1993 TB was declared a global public health emergency by the World Health Organization (WHO), the only disease thus far to be singled out. One third of the world's population is thought to be infected with *Mycobacterium tuberculosis* (*M. tuberculosis*) and at risk for progression to active TB (Ducati *et al.,* 2006; Dye *et al.*, 1999). The worldwide estimates of TB for 2007 were 9.27 million new cases and between 2 to 3 million deaths (World Health Organization, 2009). TB is responsible for 100 000 deaths in children worldwide and it is estimated that by the year 2020, TB will still be one of the 10 leading causes of the global disease burden resulting in 1 billion more people being infected (Murray & Salomon, 1998; Pasqualoto & Ferreira, 2001). The estimated incidence of TB can be seen in Fig. 1.1 (World Health Organization, 2009). It is estimated that pulmonary TB killed more people in the nineteenth century in Britain than the combined total of other diseases (Cartwright, 1977) and currently this disease still represents a global threat (Ducati *et al.*, 2006).

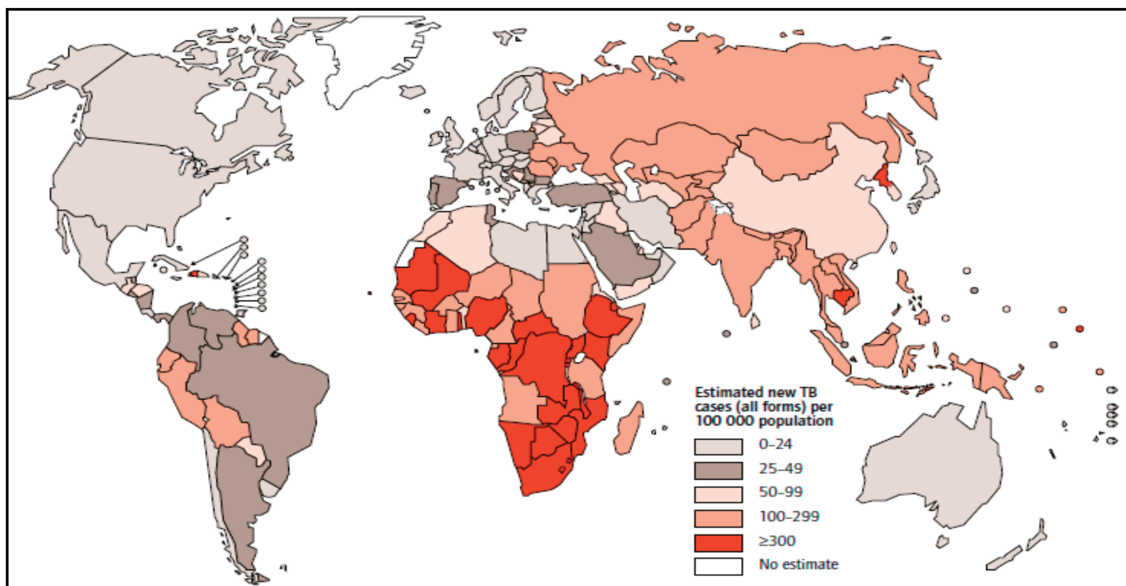


Fig. 1.1 The worldwide incidence of tuberculosis for 2007 from the WHO Report 2009 Global Tuberculosis Control (World Health Organization, 2009).

The history of TB is fundamentally linked to changes in demographic, social and environmental conditions. Therefore events such as war, urbanization, immigration and colonization play a major role in the course of TB in different population groups (Metcalf, 1991).

1.1.1 The origin and history of tuberculosis

When and where TB disease occurred for the first time is still unknown, but evidence shows that it has been prevalent for several thousand years. The *M. tuberculosis* organism was first discovered by Koch in 1882 (Koch, 1882). *M. tuberculosis* and *M. bovis,* which cause TB mainly in humans or cattle and other animals respectively, are genetically very similar, therefore it has been speculated that tuberculosis infection in humans evolved from a mutant form of *M. bovis* (Coovadia & Benatar, 1991). A more recent theory is that domesticated animals contracted the common mycobacterial ancestor from humans (de la Rua-Domenech, 2006). The earliest evidence of tuberculosis in animals comes from elephants in India prior to 2000 BC (Manchester, 1984) and the earliest written evidence of human pulmonary TB comes from Asia (Morse, 1967). The first person to describe the contagious nature of TB was Aristotle (384-322 BC) (Leff *et al.*, 1979). However, no definite theory has been established regarding the origin and spread of TB since the information is still very fragmented due to absence of data, unreliable statistics, under-reporting and over-diagnosis.

For the duration of the industrial revolution in the 18th and 19th century, towns began to grow rapidly, leading to overcrowding. Since TB was not thought of as an infectious disease, no safety measures were taken to avoid transmission, giving rise to the endemic nature of the disease. European colonialism brought about unfavorable changes in the population demographics, way of life and surroundings of many indigenous people in certain parts of the world and in Africa making them more susceptible to TB (Roelsgaard *et al.*, 1964). Even in Europeans, where TB has been endemic for centuries, lifestyle changes affect TB incidence. For example during the first and second world war the mortality rate of TB increased above the trend due to changes in household conditions but mostly due to poor nutrition (Coovadia & Benatar, 1991).

The earliest information regarding TB in indigenous people from Australia and Africa comes from European travellers and missionaries but its reliability is questionable. The absence of data is a huge problem in determining the history of TB particularly in developing countries where the figures are unreliable or unavailable (Metcalf, 1991). Due to the unreliable records there were, prior to the twentieth century, no reliable estimates of TB incidence available for any country worldwide. Therefore the first recorded date for TB in the Cape and Natal colonies in South Africa was in 1904 (Bryder, 1988). TB deaths were extensively under-reported due to the stigma surrounding them, and were often wrongly ascribed to pneumonia or bronchitis (Fishberg, 1922; Swan, 1985). Over-diagnoses started in the twentieth century because of the inability to differentiate between diseases with similar clinical features, such as pneumonia, heart conditions and lung cancer.

Much controversy exists regarding susceptibility to TB in specific ethnic groups. A high incidence of TB in a specific group does not necessarily imply that the group as such is more susceptible to TB, since ethnic groups are in many instances associated with specific socio-economic status which is an important determinant of disease risk. Several theories existed for the ethnic differences, one of these being the theory of "virgin soil/populations" that started in the late nineteenth to early twentieth century. The term "virgin soil" refers to populations that have never been exposed to infection and therefore have had no chance to acquire resistance against the causative organism, making them more susceptible (Metcalf, 1991). After investigating this theory it was found that it can only be applied to infectious diseases such as smallpox, where individuals obtain permanent immunity after infection, however this theory does not hold in the case of TB since environmental factors can reverse apparent immunity. Therefore this theory does not explain most susceptibility to TB (Metcalf, 1991).

1.1.2 The history of TB in South Africa

Socio-economic, political and demographic factors played a major role in determining the incidence of TB in South Africa. Information prior to the 1940s and 1950s, regarding the prevalence of TB and extent of the disease, is incomplete due to poor data collection. As previously mentioned, the reporting of TB started in 1904 in the Cape and Natal (Union of South Africa, 1914) but in 1943 the following was stated about the notification figures:

"…cannot be relied upon to find an accurate estimate of the incidence of the disease in South Africa [as] notification of tuberculosis … is very poorly observed [and] is often erroneous, being based on purely clinical and not radiological or bacteriological grounds." (Dormer *et al.*, 1943)

It is thought that the initial spread of TB amongst indigenous South Africans is closely related to the amount of contact with the Europeans (McVicar, 1932). TB was first described as a severe problem amongst the Khoikhoi people who lived in the south-western parts of South Africa in the early 19th century, whereas black people were infected with TB only later in the century (Metcalf, 1991). The San people were particularly devastated by TB after European contact (McVicar, 1932). Industrial development in South Africa led to rapid urbanization of black and coloured (to be defined later, see chapter 2) people in the late nineteenth century, resulting in the simultaneous increase in TB mortality rates in Johannesburg, Durban and Cape Town (Packard, 1989).

It has been suggested that the mining conditions in South Africa played a major role in the spread of TB. Very high mortality rates were reported amongst black mine workers in the Witwatersrand area in 1912. Conditions in the mines favoured the spread of TB infection, such as poor ventilation and high humidity which was exacerbated by miners working in close proximity to one another (Packard, 1989). After 1912, tuberculin tests were conducted in the mines and mine workers who tested positive for TB were sent home. The miners had inadequate diets and scurvy was prevalent, increasing their chances of contracting TB as vitamin C deficiency can lower resistance to TB (Metcalf, 1991). In 1924 Allen reported that poor living conditions contributed significantly to the high TB mortality rate (Allen, 1924).

1.2 MYCOBACTERIUM TUBERCULOSIS

TB is a chronic granulomatous disease in humans and other mammals. Four closely related mycobacterial species: *M. tuberculosis* (mostly human tubercle bacillus), *M. bovis* (mostly bovine and other animals), *M. microti* (mostly small mammals) and *M. africanum* (mostly human and bovine) together also known as the *M. tuberculosis* complex, can cause TB (Greenwood *et al.*, 1997). *M. tuberculosis* is responsible for most human TB cases and is characterized as a weakly gram-positive bacterium that is non-motile, non-sporulating, non-

capsulating, has straight or slightly curved rods and may occur in clumps or individually (Casanova & Abel, 2002; Greenwood *et al.*, 1997; Murray *et al.*, 1999).

1.2.1 The M. tuberculosis genome

The complete genome of *M. tuberculosis* laboratory strain H37Rv was published in 1998 and this led to substantial advances in the field of bacterial genomics (Cole *et al.,* 1998; Fleischmann *et al.*, 2002). The *M. tuberculosis* genome is highly conserved and consists of 4000 genes rich in repetitive DNA, coding for enzymes involved in lipolysis and lipogenesis (Ducati *et al.*, 2006). Since then a number of other *M. tuberculosis* and related mycobacterial strains have been sequenced. Comparative genomic analysis of *M. tuberculosis* revealed DNA conservation between chromosomes. Therefore nonsynonymous single nucleotide polymorphisms (nsSNP) that create an amino acid change and synonymous single nucleotide polymorphisms (sSNP) provide genetic information that allow for different *M. tuberculosis* strains to be distinguished from each other. Both types of SNPs have a specific function regarding TB. The nsSNPs correlate with the resistance to anti-tuberculosis agents whereas the sSNPs can provide the basis for genetic drift and indicate evolutionary relationship between mycobacterial strains (Mathema *et al.*, 2006). Some of the most common typing techniques used to study the epidemiology of TB are IS*6110* RFLP analysis, spoligotyping, MIRU-VNTR and SNP analysis (Mathema *et al.*, 2006).

1.2.2 TB strain families

A hypothesis suggesting that the modern *M. tuberculosis* strains underwent an evolutionary bottleneck ± 20 000 years ago has been formulated based on the high level of conserved DNA in the *M. tuberculosis* housekeeping genes and on the analysis of rare synonymous SNPs (sSNPs) and deletions (Brosch *et al.,* 2002; Kapur *et al.*, 1994; Sreevatsan *et al.*, 1997). One of the ways by which *M. tuberculosis* has evolved is by deletion or duplication of genomic segments. Some of these deletions can be used as phylogenetic markers since they occurred only once (Hirsh *et al.*, 2004).

Several strains such as LAM (Latin American and Mediterranean), Haarlem, C strain and X strain exist all over the world, and the W-Beijing family of strains has been studied most extensively (Malik & Godfrey-Faussett, 2005). An investigation into 875 strains from 80 different countries led to the hypothesis that *M. tuberculosis* originated and drifted together

with humans "out of " Africa (Gagneux *et al.*, 2006). Other evidence that differs from that of Gagneux *et al.*, suggests that six lineages of *M.tuberculosis* have adapted to specific populations (Maartens & Wilkinson, 2007) and Hanekom *et al.* showed that strains from a defined sublineage may have been selected by a human population in a defined geographical setting (Hanekom *et al.*, 2007). This argument is also supported by the fact that the HLA allele frequencies, of which the HLA genotype have been associated with susceptibility to *M. tuberculosis*, varies widely between human populations with different historical backgrounds, with some alleles completely absent in some populations (Lombard *et al.*, 2006)

The Beijing family of strains, most easily recognised by its distinctive spoligotyping pattern (Bifani *et al.*, 2002), was first reported in 1995 (van Soolingen *et al.*, 1995). This is the most dominant strain worldwide and has been reported in many countries, such as Vietnam, Thailand, Hong Kong, Russia and South Africa (Anh *et al.,* 2000; Chan *et al.*, 2001; Drobniewski *et al.*, 2002; Glynn *et al.*, 2006; Prodinger *et al.*, 2001). It has been suggested that Beijing strains are hypermutable and have greater virulence compared to other strains (Rad *et al.*, 2003) and have also been associated with a higher frequency of drug resistance (Dale *et al.,* 2005; Rad *et al.*, 2003; van der Spuy *et al.*, 2009).

From the Mycobacterial complex, *M.tuberculosis* has evolved into three principle genetic groups (PGG 1, 2 and 3). Based on the combinations of two polymorphisms at the *kat*G codon 463 and *gyr*A codon 95 site, isolates of *M.tuberculosis* were assigned to one of these three groups (Sreevatsan *et al.*, 1997). PGG 1, known as the ancestral group, is genetically more diverse than PGG 2 and 3 since it is older and therefore has accumulated variation.

Based on the sSNP analysis, strains were also assigned to eight different clusters (I-VIII) of genetically related organisms on the basis of phylogenetic analysis of 230 sSNPs. The PGG 1 was assigned to clusters I and II, PGG 2 was assigned to clusters III-VI and PGG 3 which also contain H37 was assigned to clusters VII and VIII (Fig. 1.2) (Gutacker *et al.*, 2006).

It has been suggested that strain fitness that refers to the heritable variation among members of a given species or phylogenetic lineage can play a major role in transmission

(Cohen & Murray, 2004). However, this is still poorly understood. Certain factors that might enhance the fitness in a population may include the ability of the strains to: (1) endure and reproduce within the hostile environment of the host, (2) modulate the host response and (3) up or down regulate expression of specific genes to adjust the effectiveness of antimycobacterial agents (Mathema *et al.*, 2006).

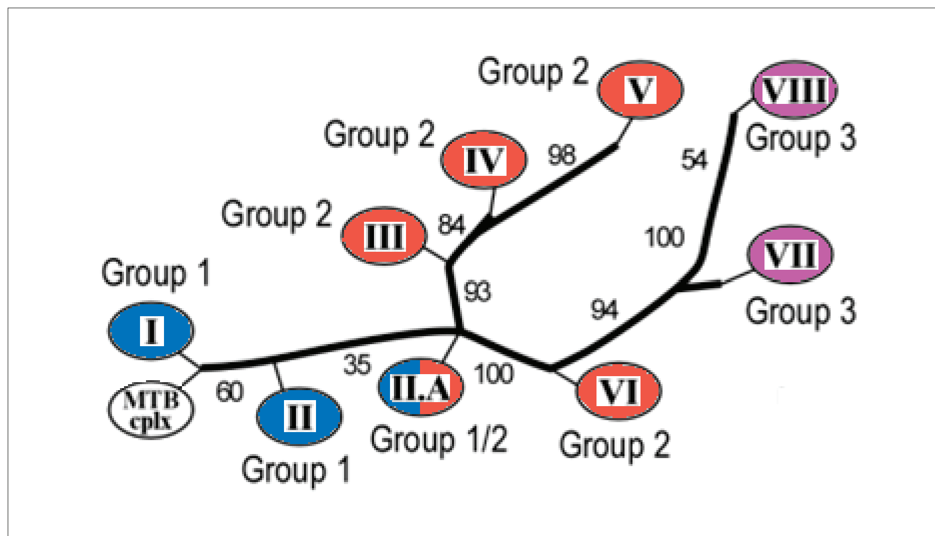


Fig. 1.2 The three PGG assigned to the eight different clusters based on the sSNP analysis. The *M.tuberculosis* complex includes *M. bovis*, *M. africanum*, *M. microtti*, and *M. canettii* strains. Bootstrapping was used to assess the genetic distances (adapted from (Gutacker *et al.*, 2006).

1.3 PATHOGENESIS OF TB

M. tuberculosis is a pathogenic bacterium that is usually transmitted through the coughing of a patient with pulmonary TB releasing a droplet of nuclei containing *M. tuberculosis* particles of about 1-5µm in diameter and the inhaling of this droplet by another person. After inhalation, the bacillus usually establishes itself in the lungs, being the initial site of infection, where it is phagocytosed by the alveolar macrophages after which a vital host cellular immune response involving T helper cells, cytokines and a large number of chemokines can either lead to protective immunity or progression to disease (Henderson *et al.,* 1997; Roach *et al.*, 2002). It is also known that a combination of environmental and host genetic factors play a major role in the outcome of exposure and infection (Fig. 1.3) (Casanova & Abel, 2002). It is clear that susceptibility to TB is polygenetic therefore it is unlikely that dysfunction in only one gene will allow progression to TB after infection (Meya & McAdam, 2007). Currently, it is thought that in 10% of immunocompetent

individual, the infection is not contained and they will develop active TB whereas the other 90% will never develop disease (Murray *et al.*, 1990).

The tubercle bacillus is a intracellular pathogen that survives in the macrophage (Greenwood *et al.*, 1997). The T lymphocytes attracted to the site release lymphokines which activate the macrophages to form a granuloma (cluster). The macrophages in the granuloma are active and use oxygen which results in anoxia, which together with acidosis in the centre of the lesion, kills most of the tubercle bacilli. Granuloma formation is in most cases sufficient to limit primary infection, however not all bacilli are destroyed. The rest can presumably remain dormant as long as the individual's immune system prevails. The dormant bacilli can be reactivated either spontaneously or due to a impairment in the host's immune system giving rise to post-primary disease (Greenwood *et al.*, 1997).

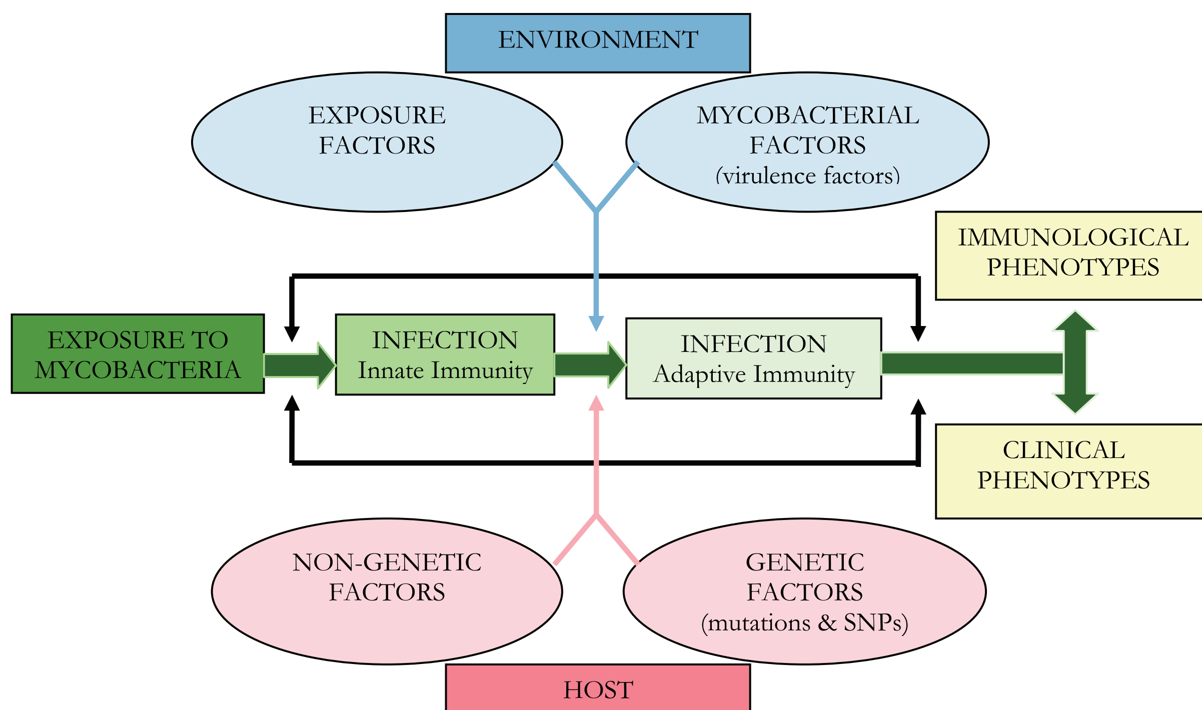


Fig. 1.3 The complex interactions between the host, environment and Mycobacteria (adapted from (Casanova & Abel, 2002).

1.3.1 TB and the immune system

The human immune system acts by mounting either a cellular or humoral response. Both of these responses involve T helper lymphocytes (Th) that are produced by Th-1 and Th-2 maturation pathways (Fig. 1.4). The T lymphocytes are grouped into CD4 or CD8

according to cluster differentiation. The Th-1 type immune response is responsible for granuloma formation, whereas the Th-2 type immune response can cause the secretion of cytokines that are responsible for the inactivation of macrophage proliferation (Labidi *et al.*, 2001). Macrophage-derived pro-inflammatory cytokines such as interleukin-6 (IL-6), IL-12, tumor necrosis factor (TNF) and interferon gamma (INF-γ) derived from natural killer cells are responsible for sustaining or inhibiting bacterial growth (Ducati *et al.*, 2006).

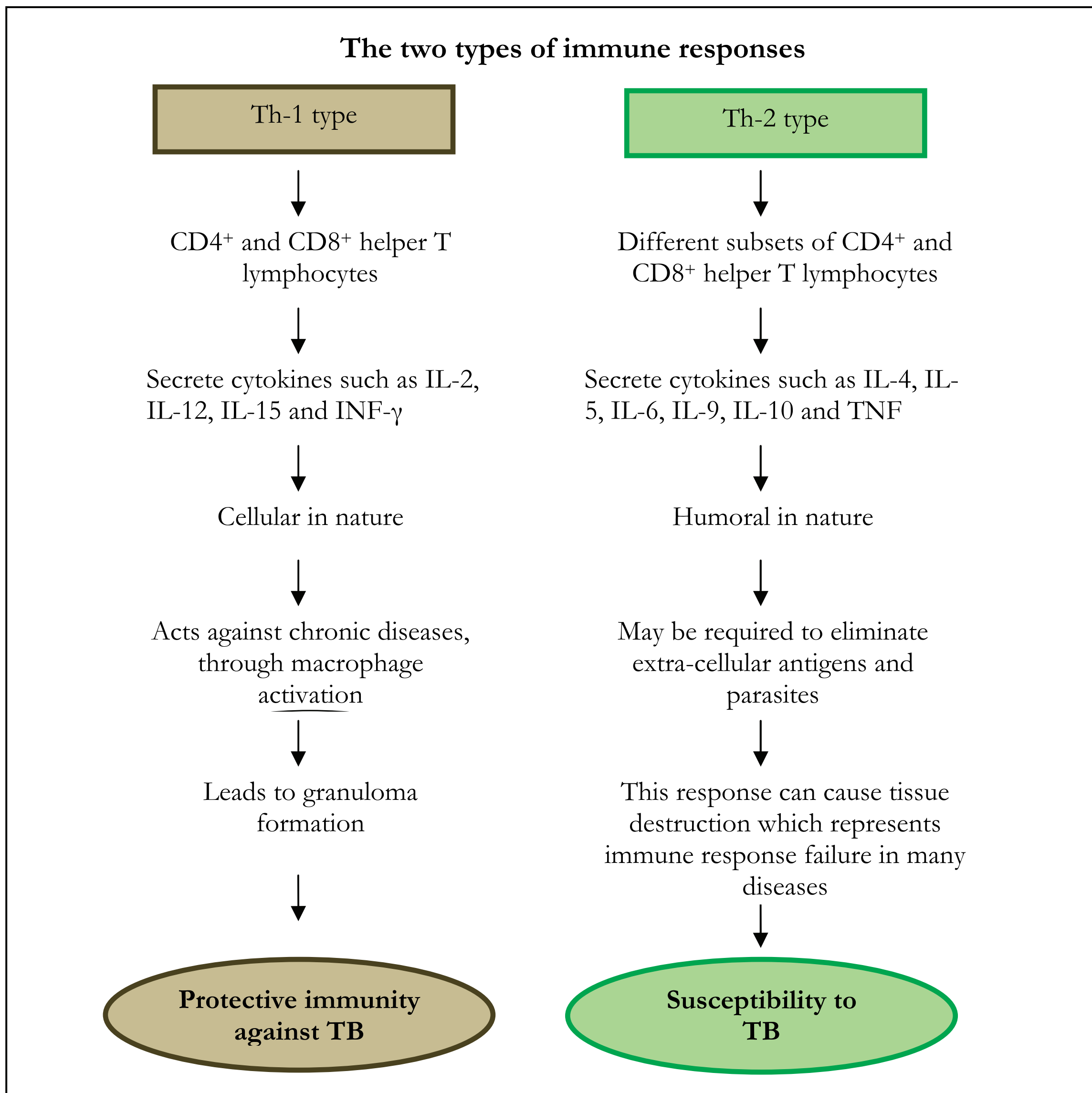


Fig. 1.4 Diagram explaining the two different types of immune responses (Th-1 and Th-2) (Ducati *et al.,* 2006; Meya & McAdam, 2007). (Note: this is somewhat oversimplified, but the Th1/Th2 balance is a widely accepted idea).

human immunization, the first vaccines that were produced focused on the protection against acute infections, where the Th-2 type immune response can be efficient (Grange *et al.*, 1995). However more recent attempts focused on developing single antigens into vaccines (Labidi *et al.*, 2001). For a continuing protective immunity, new vaccines should be developed that combine selected antigens with potent adjuvants and stimulate the suitable immunological pathway. Therefore the ideal mycobacterial adjuvant would be one that induces an increase in Th-1 type response and a parallel inhibition of Th-2 response (Grange *et al.*, 1995) (See section 1.4).

1.3.2 Recurrent TB

Recently it was reported that the age–adjusted incidence rate of TB due to reinfection among patients with successful prior treatment in Cape Town was four times higher than the rate of new TB cases (Verver *et al.*, 2005). Individuals, especially in high incidence communities, that had TB previously have a greater chance of developing disease when reinfected. Some confounding factors such as HIV and socioeconomic status might have biased the estimates, however this still indicates that a group of individuals may exist with a predisposition to TB infection or that TB increases susceptibility to recurrence (Bellamy, 2006; Casanova *et al.*, 2002; Yew & Leung, 2005).

1.4 VACCINATION

Globally, the vaccine used in fighting the battle against TB is the intradermal BCG (bacilli Calmette-Guerin) vaccine. This vaccine is easy to administer, generally without any serious complications and has been administered to more than 2 – 3 billion people since 1948 (Labidi *et al.*, 2001). BCG is a living attenuated vaccine derived from a strain of *M. bovis* and selected for its limited virulence in humans. The protective effect of this vaccination varies between populations. In some cases, BCG had a 78% protection rate while in other cases no protection was detected (Greenwood *et al.*, 1997). This difference might be ascribed to the pathogenic population-wide exposure to mycobacteria or the ineffectiveness of the BCG vaccine as demonstrated in a meta-analysis of controlled BCG trials (Colditz *et al.*, 1994). This vaccine in general has been effective in preventing TB meningitis (TBM) in children but does not give protection to pulmonary TB (PTB) in adults (Colditz *et al.,* 1994; Orme, 2001; Sterne *et al.*, 1998). Currently there are two major

problems with the BCG vaccine: 1) it has been shown that sub-strains are formed as a result of genetic modifications therefore complete avirulence acquisition is not unsuspected for the vaccine making immunization impractical and 2) after many years of BCG vaccination there can be a steady loss of T memory cell population (Orme, 2001). This, together with the increasing number of TB cases worldwide, highlight the necessity for developing a better vaccine (Young & Dye, 2006). It was previously suggested that two vaccines are needed; one to prevent infection of the pathogen (that applies to two thirds of the world's population) and another one to eradicate infection already established (that applies to one third of the world's population) (Hess & Kaufmann, 1999).

1.5 DIAGNOSIS

The diagnosis of TB is not simple, and is usually algorithm-based, which is a combination of several factors, such as clinical symptoms (coughing, fever, night sweats, weight loss), a positive tuberculin test (TST), chest x-rays and by microscopy or culture. Diagnosing TB is still based on the long tedious process of detecting the acid-fast bacilli under a microscope or by culturing the bacterium. Therefore a new diagnostic tool with high sensitivity, which is rapid and affordable is sorely needed for diagnosing TB. The deletion that classifies BCG includes two antigenic proteins, namely 6-kDa early secretory antigenic target (ESAT-6) and culture filtrate protein-10 (CFP-10) (Mahairas *et al.*, 1996). Both these antigens are essentially limited to *M. tuberculosis* therefore many assays that test for the presence of TB infection are based on IFN-γ release by the T cells in response to ESAT-6 and CFP-10 (Connell *et al.*, 2006).

The TST is used worldwide as a diagnostic test to detect infection, irrespective of whether the infection is old or recent (Bloom & Murray, 1992). The TST has some limitations regarding the antigens ESAT-6 and CFP-10, active and latent TB and BCG vaccination. Cross-reactions from exposure to other mycobacteria may induce low levels of tuberculin reactivity (Greenwood *et al.*, 1997). False positives can be produced as protein derivative contains antigens present in BCG and non-pathogenic mycobacteria, or false negatives can occur in immunocompromised patients or in early primary TB. The TST is therefore not the best test to use since it cannot differentiate between active or latent TB, but on the other hand it can indicate the presence of *M. tuberculosis* infection (Wilkinson *et al.*, 2005).

With molecular techniques, such as PCR, TB can now be diagnosed in 1 day instead of 2 months.

Diagnosing drug resistant TB can take up to 6-8 weeks. Resistance to Rifampicin usually acts as a marker for multi-drug resistance and this can be detected by rapid molecular methods such as the polymerase chain reaction (PCR) (Morgan *et al.*, 2005). In patients infected with both TB and HIV the sputum-smear test for TB is often negative, therefore the WHO issued diagnostic guidelines for areas where the HIV incidence is very high (World Health Organization, 2006).

1.6 TREATMENT

The first treatment that TB patients, including HIV-infected TB patients, would receive according to guidelines set by the WHO, consists of the combined administration of rifampicin, isoniazid, pyrazinamide and streptomycin/ethambutol for the first 2 months followed by a combination of rifampicin and isoniazid for at least another 4 months (Jindani *et al.,* 2004; World Health Organization, 2006). The directly observed treatment short course (DOTS) is the internationally recommended strategy to control TB and consists of five fundamental elements (Fig. 1.5).

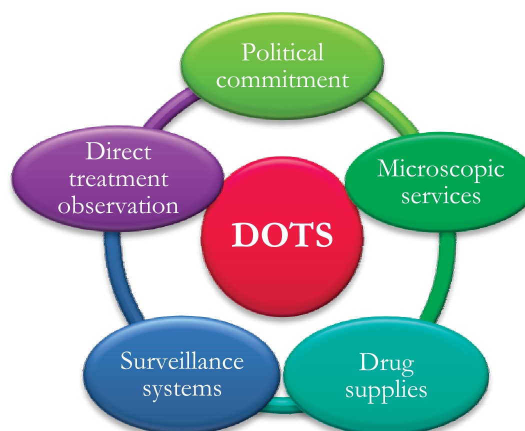


Fig. 1.5 The five fundamental elements of the DOTS strategy

The DOTS strategy was responsible for a decrease in the incidence of TB in many areas of the world, but is far from complete or ideal for multidrug-resistant TB. Therefore new diagnostic tools, better vaccines and new drugs need to be developed. People with latent TB can also be treated with only isoniazid for six months, however this usually happens in

first world countries only (Ducati *et al.*, 2006). Several factors are responsible for the resumption of TB (Table 1.1) of which the most important is social inequality as it leads to malnutrition, poor living conditions and low education levels (Ducati *et al.*, 2006).

Table 1.1 Factors responsible for the resumption of TB
Increase in drug resistance
HIV/AIDS pandemic
Increase of injectable drug users
Social structure
Poor nutrition
Increase amount of immigrants
Aging of the world population
Transmission amongst environments (prisons, hospitals and shelters)
Social inequality
Degradation of health care systems

1.7 DRUG-RESISTANT TB

Drug-resistant TB can arise from two major sources: (1) acquired resistance due to inadequate treatment that leads to the selection of resistant strains in patients and (2) primary resistance that develops when an individual is infected or re-infected with a resistant strain (Victor *et al.*, 2007). Previous studies have stated the existence of an association between the occurrence of spontaneous mutations in *M. tuberculosis* and drug resistance (David & Newman, 1971). Since acquired resistance due to antibiotic pressure can lead to accumulation of these drug-resistant mutants the use of combination therapy (using several antibiotics) to reduce the appearance of drug resistance during therapy became prevalent (David, 1971; East African/British Medical Research Council, 1972). The long duration of treatment combined with the side effects of TB treatment, irregular drug supply, poor adherence and malabsorption are contributing factors to drug resistance (Davies, 1998; Singh *et al.*, 2007). In the early 1990s, multidrug-resistant tuberculosis (MDR-TB) emerged as a growing phenomenon globally, whereafter extensively drug-resistant tuberculosis (XDR-TB) followed (Mathema *et al.*, 2006). South Africa has been identified as one of the high-burden countries for drug-resistant TB with a staggering number of more than 10 000 number of cases reported for 2007 (World Health Organization, 2009).

MDR-TB is by definition the resistance to the two important frontline antibiotics, rifampicin and isoniazid (Telenti & Iseman, 2000). It was identified in the Western Cape, South Africa for the first time in 1985 (Weyer *et al.*, 1995). It is estimated that more than 4% of TB patients worldwide are MDR and more than 40% of these have previously been treated for TB (Zignol *et al.*, 2006). The WHO estimated that 300 000 – 600 000 new cases of MDR-TB may emerge every year (World Health Organization, 2006). The mortality rate for MDR-TB is estimated at 40% – 60% which is equivalent to patients untreated for TB (Bloom & Murray, 1992). It has been suggested that three main factors: 1) easy accessibility to anti-TB medication, 2) patient noncompliance towards treatment and 3) unsuitable administration of drugs by clinicians, contributed to the development of MDR/XDR (Riley, 1993). MDR represents a global problem.

XDR-TB can be defined as resistance to the important frontline antibiotics (rifampicin and izoniazid), second line drugs such as ofloxacin and at least one injectable (kanamycin, amikacin, capreomycin) (Maartens & Wilkinson, 2007). An outbreak of XDR has recently been reported in South Africa where 52 of 53 HIV positive patients died within 16 days. The *M. tuberculosis* strain in these XDR patients had the same genetic history indicating recent transmission or possibly nosocomial transmission (Gandhi *et al.*, 2006). If MDR-TB is not detected when chemotherapy is started, inappropriate antibiotic use will inevitably result in the development of XDR-TB as the bacteria may be exposed to fewer active drugs than required (van Helden *et al.*, 2006).

1.8 TB AND THE HUMAN IMMUNODEFICIENCY VIRUS (HIV)

The first documented proof of a link between TB and HIV was in New York in 1992 (Bloom & Murray, 1992). It was estimated that individuals infected with HIV and latent TB are predisposed to develop active TB at a rate of 7 - 10% per year compared to 8 - 10% per lifetime in HIV-negative individuals (Bloom & Murray, 1992; Selwyn *et al.*, 1989; Selwyn *et al.*, 1992). It has been reported that the risk of TB is two fold higher one year after HIV infection (Sonnenberg *et al.*, 2005). The association between TB and HIV in sub-Saharan Africa is alarming and distressing (Corbett *et al.*, 2006). Severe suppression of the immune system in individuals increases their risk of active disease and a TB incidence rate of as high as 30% has been reported in South African patients with advanced HIV

(Wood *et al.*, 2000). It is also suggested that HIV infection is responsible for the transmission of most of the TB between adults in sub-Saharan Africa (Glynn *et al.*, 2005). The estimated HIV prevalence in new TB cases is the highest in Africa with more than 50% of new TB cases in South Africa co-infected with both TB and HIV (Fig. 1.6) (World Health Organization, 2009).

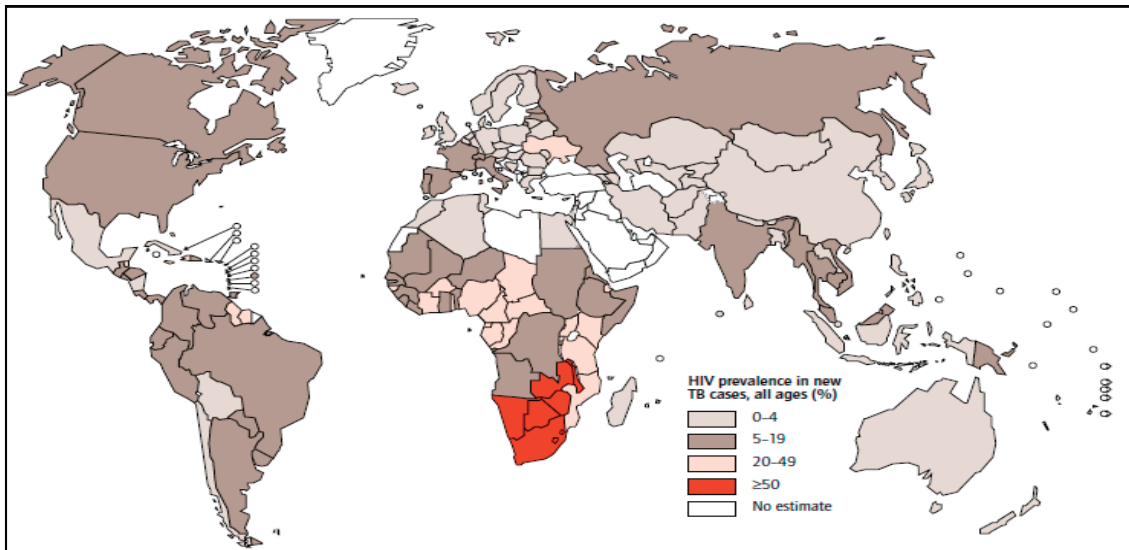


Fig. 1.6 The worldwide estimated HIV prevalence in new TB cases for 2007 from the WHO Report 2009 Global Tuberculosis Control (World Health Organization, 2009).

TB induces AIDS progression in HIV-positive patients by the production of cytokines and a decrease in CD4 cells therefore co-infection of HIV and TB represents a major problem (Ducati *et al.*, 2006). Evidence suggesting that a strong association exists between TB and HIV was supported in a study done in Africa (Table 1.2) (Ducati *et al.*, 2006). MDR-TB has been called "the most malignant opportunistic infection yet associated with HIV infection" (Nolan, 1997).

Table 1.2 Evidence suggesting a strong association between TB and HIV in Africa
TB is the most important disease associated to HIV/AIDS
TB is the major cause of death amongst HIV patients
The incidence of TB increased with the appearance of AIDS
HIV is more prevelant in people with TB
Active TB is more prevelant in HIV-positive people

1.9 HOST GENETIC SUSCEPTIBILITY TO TB

Since it is estimated that only 10% of one third of the world's population infected with *M. tuberculosis* will develop active TB, the question arises: what makes them different from the other 90%? In search of a genetic based answer to this question, many approaches have been used, including animal models, case-control association and genome-wide linkage studies (Bellamy, 2006). Adoption studies (Sorensen *et al.*, 1988), twin studies (Comstock, 1978) and familial clustering (Fine, 1981) have indicated that host genetic factors play an important role in host susceptibility to TB.

The genetic susceptibility component of TB seems to be scattered across many genes as well as different genes being implicated in different populations, which corresponds with the fact that *M. tuberculosis* might be population specific (Maartens & Wilkinson, 2007). Genes may not play a role in the risk of infection upon exposure but they might influence the risk of disease development and the course of disease (Daniel, 1997). Many studies indicated that *M. tuberculosis* applies a powerful selective pressure on the human genome (Bloom & Small, 1998). The genetic component of susceptibility to TB has been studied extensively in the past few years resulting in a rapid increase in available information.

Since TB is a complex disease, it can be assumed that susceptibility to *M. tuberculosis* has multiple genetic causative factors (as well as environmental causes), therefore a single approach would not be able to identify all the genes involved (Bellamy, 2006). Different approaches will be discussed below.

1.9.1 Candidate gene approach

The selection of a candidate gene for a genetic study can be based on two approaches, one being by experiment, using data from genome scans, linkage analysis or animal studies to identify candidate genes, and secondly by hypothesis, where the known or presumed function of the gene provides " probable cause" for association with the disease (Fig. 1.7). Linkage analysis and case-control association studies are based on two completely different approaches (Table 1.3).

Table 1.3 A comparison between linkage-type studies and association studies (Lander & Schork, 1994)

Linkage studies	Association studies
Test whether they show correlated transmission within a pedigree	Test whether a disease and an allele show correlated occurrence in a population
Focus on concordant inheritance	Focus on population frequencies
Can detect linkage without association	Can detect association without linkage

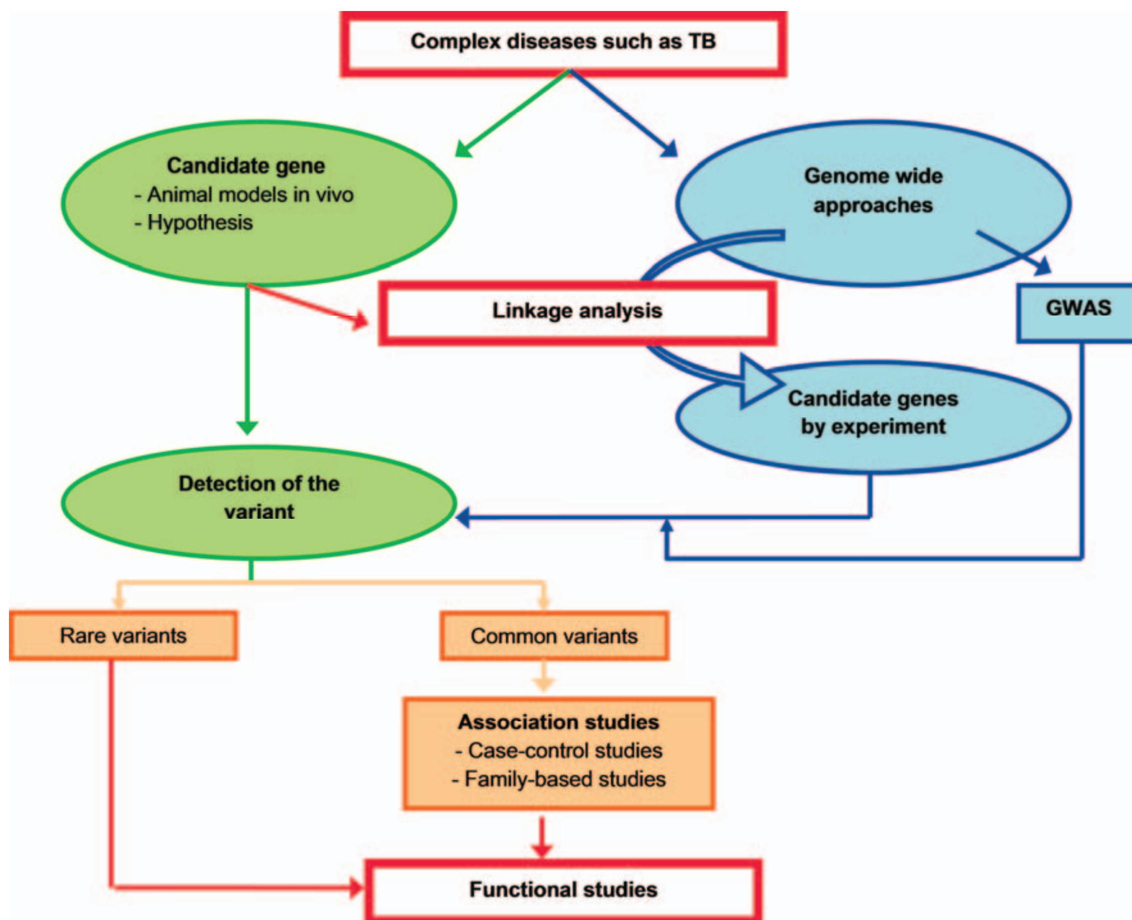


Fig. 1.7 Strategies for identifying susceptibility genes, adapted from (Alcaïs & Abel, 2004).

1.9.2 Linkage analysis

Linkage studies are used to trace chromosomal regions containing putative susceptibility genes, either by a genome-wide scan, which ensures that all major genomic regions involved in disease susceptibility are identified, or by concentrating on a candidate region.

This provides the opportunity to find new genes and pathways that might not previously have been suspected to contribute to the disease studied. The inheritance pattern of phenotypes and genotypes observed in a certain pedigree can be explained by linkage analysis which proposes a model. Linkage analysis has successfully mapped hundreds of monogenic traits (Lander & Schork, 1994). This type of analysis is very successful for simple Mendelian traits but to find a precise model that adequately explains the inheritance pattern of a complex trait such as TB can be difficult or impossible.

Model-based or parametric linkage analysis by the logarithm of odds (lod) score method needs a defined model that specifies the relationship between the phenotype and factors that may influence its expression. The model is provided by segregation analysis. Model-free or non-parametric linkage studies are used when little is known about the relationship between the phenotype and the gene, which is generally the case with complex diseases. Once evidence for linkage has been found, fine genetic and physical maps are constructed to narrow down the interval on the chromosome and allow gene identification by candidate gene selection (when the function of the gene is known) or by positional cloning (when the function is unknown) (Abel & Dessein, 1997).

The chromosome 17q11-q21 candidate region, syntenic to mouse chromosome 11 which was previously identified as a susceptibility region for another intracellular pathogen (Roberts *et al.*, 1993), was chosen for a TB linkage study (Jamieson *et al.*, 2004). This indicated that the chromosome 17q11-q21 region contained TB susceptibility genes, and further evaluation of this region suggested that four genes contributed separately to susceptibility (Jamieson *et al.*, 2004). A model-based linkage study performed in a Canadian pedigree using microsatellite markers flanking the natural resistance associated macrophage protein 1 gene (*NRAMP1*, renamed solute carrier family 11A member 1 (*SLC11A1*)), revealed linkage between TB and the 2q35 loci (Greenwood *et al.*, 2000).

The first model-free genome-wide linkage study for tuberculosis was conducted using affected sibling pairs from West and South African families where two genomic regions on 15q and Xq27 showed evidence of linkage with TB (Bellamy *et al.*, 2000). To date six genome-wide linkage scans have been done in TB genetics (Table 1.4). The minimal overlap between susceptibility regions observed in these studies is probably because linkage

of a genotype for a genetic marker with TB may be unique to a specific family or population and thus impossible to identify in other studies. These studies are also exquisitely sensitive to the phenotype definition.

However, as discussed previously, it is very difficult to find a precise model that adequately explains the inheritance pattern of a complex trait such as TB. Association studies have much greater power for detecting genes of small effect than linkage analysis (Risch & Merikangas, 1996), provided that the sample size is adequate. Therefore association studies are the most popular study design for investigating host susceptibility to TB.

Table 1.4 Chromosomal regions identified by genome-wide linkage studies of TB

Population	Chromosomal region	TB phenotype	Reference
South African	15q11-13	TB	(Bellamy *et al.*, 2000)
The Gambia	Xq	TB	
Brazilian	10q26.13	TB	(Miller *et al.*, 2004)
	11q12.3	TB	
	20p12.1	TB	
Moroccans	8q12-q13	TB	(Baghdadi *et al.*, 2006)
South African	6p21-q23	TB	(Cooke *et al.*, 2008)
Malawian	20q13.31-33	TB	
Ugandans	2q21-2q24	PTST-	(Stein *et al.*, 2008)
	5p13-5q22	PTST-	
	7p22-7p21	TB	
	20q13	TB	
Thais	5q23.2-31.3	TB	(Mahasirimongkol *et al.*, 2009)
	17p13.3-13.1	TB, CA	
	20p13-12.3	TB, CA	

TB, current or previous microbiologically confirmed tuberculosis
PTST-, persistently negative tuberculin skin test
CA, ordered subset analysis with minimum age at onset of disease as covariate

1.9.3 Case-control association studies

Population based case-control association studies are the classic association studies where the allele frequency of a specific marker is compared between unrelated cases (affected individuals) and controls (unaffected individuals) when Hardy-Weinberg equilibrium (HWE) holds (Lander & Schork, 1994). Case-control association studies can detect weak effects and several of these studies replicated or validated associations found in previous studies or detected novel pathways that might be involved in pathogenesis (Maartens & Wilkinson, 2007). However, an association can arise for one of three reasons: 1) When the

specific allele is the actual cause of the disease, 2) when the allele does not cause the trait but is in linkage disequilibrium with the actual cause of the disease, 3) as an artifact of population admixture (Lander & Schork, 1994). It is very important therefore, to address population stratification and to replicate association studies to confirm the results.

1.9.4 Population stratification

Since population stratification can arise when the genetic background of cases and controls is different, it is important that both these groups should be sampled from the same area and that cases and controls should be carefully matched on their ethnic, and by implication, genetic backgrounds (Reich & Goldstein 2001; Cardon & Palmer 2003). Case-control association studies are valid tests for association, provided that the issue of population stratification is addressed (Pritchard & Rosenberg 1999). The role population stratification plays in genetic studies as a whole and in this particular study will be discussed in more detail in chapter two.

1.9.5 TB and previous genetic susceptibility studies

The genetic component of TB susceptibility seems to be scattered across many genes and it has been studied extensively in the past few years resulting in a vast amount of information (Fig. 1.8). Several candidate genes, amongst others *HLA*, *VDR*, *DC-SIGN*, *NRAMP1, IFN-γ* and *MBL* have been associated with TB (Table 1.5) whereas others were not associated (Table 1.6). The genes *NRAMP1, IFN-γ* and *MBL* will be discussed in section 1.9.6 as they formed part of this study.

The human leukocyte antigen (HLA) region consists of approximately 200 genes, many of which are involved in antigen presentation. Genes that are involved in protective immunity show greater variance than other genes (Murphy, 1993). This is also the case for the HLA region, which varies between populations and is highly polymorphic. It has been postulated that this is the effect of different selection pressures, such as infectious disease (Lombard *et al.*, 2006). The *HLA* class I and class II genes are involved in antigen presentation to T cells and each protein binds a different range of peptides (Janeway *et al.*, 2001). There are three class I α-chain genes in humans, namely HLA-A, -B and –C. There are four subclasses of genes in the HLA class II region, named HLA-DR, -DP, -DM and –DQ respectively. *HLA* genes have been examined in several TB susceptibility studies and

were some of the first genes to be investigated and associated with the disease. The HLA-DR2 is the most consistently associated with TB in several populations such as India (Meyer *et al.,* 1998; Ravikumar *et al.*, 1999), Thai (Vejbaesya *et al.*, 2002), Indonesia (Selvaraj *et al.*, 1998) and Russia (Khomenko *et al.*, 1990). It was found that the HLA-DQB1*0503 influence the progression of TB in the Cambodian population (Goldfeld *et al.*, 1998) whereas the DQB1*0601 was associated with TB in the Thai and South Indian population (Ravikumar *et al.,* 1999; Vejbaesya *et al.*, 2002). DRB1*1302 was associated with TB susceptibility in a Venda population (Lombard *et al.*, 2006).

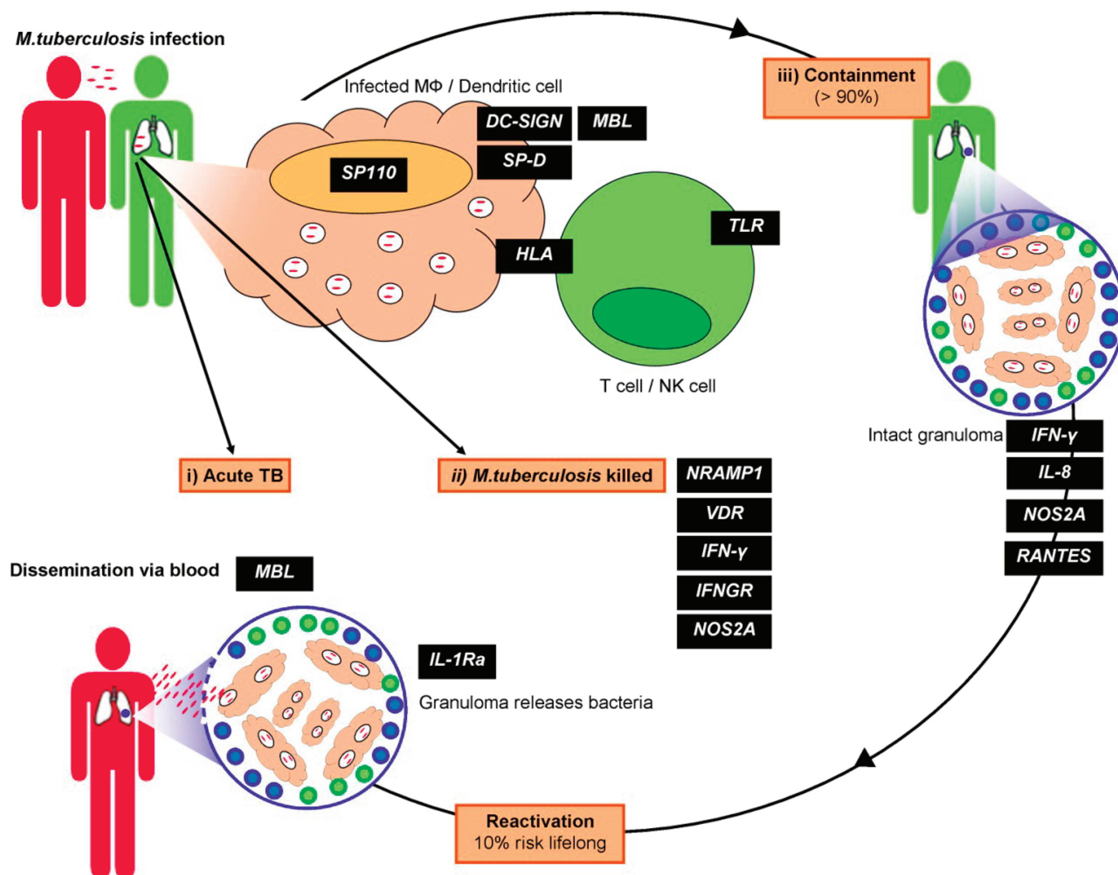


Fig. 1.8 Genetic involvement in the TB disease process.
A simplified representation of different outcomes after *M. tuberculosis* infection and some of the genes that may be involved at various stages. The bacteria enter the respiratory system of the host via inhaled droplets and are engulfed by macrophages (MΦ) and dendritic cells. There are three potential outcomes after inhalation of *M. tuberculosis*: i) Infection develops into active TB. ii) *M. tuberculosis* is immediately killed by the pulmonary immune system. iii) Infection does not progress to active disease, because the bacteria are contained in granulomas. During active TB, *M. tuberculosis* can disseminate from granulomas. (Adapted from (Kaufmann & McMichael, 2005).

The vitamin D receptor (*VDR*) gene mediates the effects of the active metabolite of vitamin D, 1,25-dihydroxyvitamin D3, which suppresses the growth of *M. tuberculosis* in vitro (Denis, 1991; Rockett *et al.*, 1998) by stimulating cell-mediated immunity and activating monocytes (Rook *et al.*, 1986). Conflicting results have been found in association studies of VDR in TB (Bellamy *et al.,* 1999; Bornman *et al.*, 2004; Selvaraj *et al.*, 2000; Wilkinson *et al.*, 2000). A meta-analysis of studies indicated that results were inconclusive and that the studies were underpowered (Lewis *et al.*, 2005). A recent study in South Africans determined that the ApaI "AA" genotype and "T"- containing TaqI genotypes predicted a faster response to TB treatment, but did not detect an association with TB in a case-control analysis (Babb *et al.*, 2007).

1.9.6 Genes investigated in this study

Eight genes, previously indicated to play an important role in TB susceptibility, were chosen as candidate genes for this study and will be discussed in detail in the chapters to follow (See Fig 1.8).

The gene encoding the collectin surfactant protein D (SP-D) is located on chromosome 10 and has previously been associated with TB in a Mexican population (Floros *et al.*, 2000). SP-D plays a crucial role in the pulmonary innate immune response by interacting with the macrophage cell surface molecules and therefore modulating phagocytosis of *M. tuberculosis* (Ferguson *et al.*, 2002). SP-D has been used as a biomarker for pulmonary disease states as it is present on the extrapulmonary epithelial surfaces and in serum (Sørensen *et al.*, 2006).

The gene encoding the collectin mannose-binding protein (MBP) also known as the mannose-binding lectin (MBL) is mapped to chromosome 10. MBL is a component of the innate immune system and plays an important role in the modulation of inflammation (Turner, 2003). It has become more evident that MBL plays a multifaceted role in many diseases and that a deficiency of the protein, due to three mutations in exon 1, has been shown to be associated with increased susceptibility to infectious disease (Summerfield *et al.*, 1997) but equivocal results have been found in the case of TB (Bellamy *et al.,* 1998; Selvaraj *et al.*, 1999), where promoting the uptake of bacteria into macrophages may be advantageous to the bacterium, and the variant alleles have been associated with protection

against TB and particularly tuberculous meningitis in South Africa (Hoal-van Helden *et al.*, 1999), Denmark (Soborg *et al.*, 2003) and Turkey (Cosar *et al.*, 2008).

An effective response by the host during bacterial infections in the lung (*M. tuberculosis* infection) is the recruitment and activation of inflammatory agents to the lung through chemokine activity. Interleukin 8 (IL-8) is a tissue derived peptide and a major chemokine that is secreted by several cell types in response to inflammatory stimuli (Ma *et al.*, 2003). In tuberculosis patients increased levels of IL-8 have been reported in the bronchoalveolar lavage fluid (BALF), pleural exudates or cerebrospinal fluid, compared to the controls (Dlugovitzky *et al.,* 1997; Mastroianni *et al.*, 1994; Sadek *et al.*, 1998). Polymorphisms in the *IL-8* gene, located on chromosome 4, have been associated with TB in a case-control study as well as a TDT study (Ma *et al.*, 2003).

The interferon-gamma (IFN-γ) pathway is one of the most well-known pathways involved in TB since it is a flagship Th1 cytokine and plays a vital role in the protective immune response against *M. tuberculosis* infection (Vidyarani *et al.*, 2006). Several polymorphisms have been indentified in *IFN-γ* and in the α and β chains of the interferon-gamma receptor (*IFNGR*) gene that were mapped to chromosome 12 and 6 respectively (Papanicolaou *et al.,* 1997; Zimonjic *et al.*, 1995). IFN-γ and IFNGR form a vital complex in determining the outcome of the biological effects of IFN-γ as the functional IFNGR is extremely important for containing *M. tuberculosis* (Dorman *et al.*, 2004). A previous study showed that defects in either of these two genes influence the availability of IFN-γ and therefore individuals were more prone to mycobacterial infections (Ottenhoff *et al.*, 1998). One of the most studied polymorphisms in *IFN-γ* is located in the first intron (+874 T/A) and have been associated with TB susceptibility in several populations such as South Africa where it was found in both the case-control and TDT study (Rossouw *et al.*, 2003), Spanish (Lopez-Maderuelo *et al.*, 2003) and Chinese (Tso *et al.*, 2005). A study that summarizes all the different studies is the meta-analysis performed on this particular SNP which indicated a significant protection, with the T/T genotype, to tuberculosis in different ethnic population groups throughout the world. Therefore it was shown that the +874 T/A SNP can be a major genetic marker for TB resistance. (Pacheco *et al.*, 2008).

The gene encoding the chemokine Regulated upon Activation, Normal T-cell Expressed and Secreted (RANTES), also known as CCL5, is located on chromosome 17. Its ligand is CCR5. RANTES recruits several types of inflammatory cells including monycytes and T lymphocytes and is induced after *M. tuberculosis* infection. In tuberculosis patients, an increased level of RANTES has been found in bronchoalveolar lavage fluid and in lung alveoli during active pulmonary tuberculosis (Sadek *et al.*, 1998). It has been hypothesized that these polymorphisms have an effect on TB in that RANTES recruits *M. tuberculosis*-infected macrophages and suppresses intracellular growth of the bacterium.

The gene encoding the natural resistance-associated macrophage protein (*NRAMP1*), also known as the solute carrier family 11A member 1 (*SLC11A1)*, is located on chromosome 2 and plays a role in macrophage activation. NRAMP1 regulates cytoplasmic cation levels by specifically regulating the iron metabolism in the macrophages (Wyllie *et al.*, 2002) and variants in the *NRAMP1* gene may alter the function of this protein. Genetic variants of *NRAMP1* have been associated with TB in several studies leading to the meta-analysis of *NRAMP1* polymorphisms which confirmed its association with tuberculosis susceptibility (Li *et al.*, 2006). The 5' $(GT)_n$ variant has been shown to be of functional importance as it influenced NRAMP1 expression in a luciferase reporter assay (Searle & Blackwell, 1999). The 118 allele ($(GT)_9$) has been associated with higher expression and resistance to TB whereas the 120 allele ($(GT)_{10}$) has been associated with lower promoter activity and susceptibility to TB (Awomoyi *et al.,* 2002; Hoal *et al.*, 2004). The 3'UTR variant has been linked to a significantly increased risk of pulmonary tuberculosis in West Africa (Bellamy *et al.*, 1998), Asia (Ryu *et al.*, 2000) and South Africa (Hoal *et al.*, 2004).

The interleukin-1 receptor antagonist (*IL-1Ra*) gene is located on chromosome 2 in close proximity to the *IL-1a* and *IL-1β* genes. Both these gene products bind to the receptor and set off the recruitment and activation of macrophages leading to potent pro-inflammatory immune responses (Dinarello, 1988). IL-1Ra also binds to the receptor but acts as a competitive inhibitor of IL-1 biological activity, therefore the interaction between IL-1 and IL-1Ra determines the initiation, persistence or termination of the pro-inflammatory response (McIntyre *et al.*, 1991). Polymorphisms in *IL-1Ra* have previously been associated with TB susceptibility (Gomez *et al.,* 2006; Wilkinson *et al.*, 1999).

Table 1.5 Articles reporting significant linkage or association between susceptibility genes and TB (Burgner *et al*., 2006).

Candidate gene	Population	Phenotype	Sample Size	Reported Results	Reference
MHC Class (Region I, II, III)					
A1 Supertype	Indian	PTB & Miliary TB	TB cases = 235; Controls = 289	OR = 0.43; pc = 0.001	(Balamurugan *et al*., 2004)
A10; B8	Indian (South)	Smear neg. PTB	TB cases = 152; Controls = 404	$p < 0.01$	(Brahmajothi *et al*., 1991)
A2; B5	Egyptian	PTB	TB cases = 42; Controls = 156	N/A	(Hafez *et al*., 1985)
A26; B17; B27; DR14	Iranian	PTB	TB cases = 44; Controls = 108	$p < 0.05$ ($pc > 0.05$)	(Mahmoudzadeh-Niknam *et al*., 2003)
B12	Thai	PTB	TB cases = 35; Controls = 35?	Decreased Frequency	(Chandanayingyong *et al*., 1988)
B14	Italian	Cavitary TB	TB cases = 54; Controls = 1089	RR = 3.9; p = 0.001	(Ruggiero *et al*., 2004)
Bw15	African American	PTB with Cavitation	TB cases = 60; Controls = 100	Increased Frequency	(Al Arif *et al*., 1979)
Bw46	Thai	PTB	TB cases = 35; Controls = 35?	Increased Frequency	(Chandanayingyong *et al*., 1988)
Cw Allotype 1	Indian	PTB & Miliary TB	TB cases = 235; Controls = 289	OR = 1.69; p = 0.005	(Balamurugan *et al*., 2004)
Cw Allotype 2	Indian	PTB & Miliary TB	TB cases = 235; Controls = 289	OR = 2.31; p = 0.000004	(Balamurugan *et al*., 2004)
DQ B57 (Asp/Asp)	Cambodian	PTB	TB cases = 436; Controls = 107	OR = 3.05; p = 0.001	(Delgado *et al*., 2006)
DQA1*0101; DQB1*0501; DRB1*1501	Mexican	PTB	TB cases = 65; Controls = 95	OR = 6.16 - 7.92	(Teran-Escandon *et al*., 1999)
DQA1*0301; DQA1*0501	Iranian	PTB	TB cases = 40; Controls = 100	OR = 0.25; OR = 0.53	(Amirzargar *et al*., 2004)
DQA1*0601, DQB1*0301	Thai	PTB	TB cases = 82; Controls = 160	OR = 0.4, $p < 0.02$	(Vejbaesya *et al*., 2002)
DQB1*02	Polish	PTB	TB cases = 38; Controls = 58	OR = 0.39; p = 0.01	(Dubaniewicz *et al*., 2003)
DQB1*0402; DR4; DR8	Mexican	PTB	TB cases = 65; Controls = 95	Decreased Frequency	(Teran-Escandon *et al*., 1999)

Candidate gene	Population	Phenotype	Sample Size	Reported Results	Reference
DQB1*05	Polish	PTB	TB cases = 38; Controls = 58	OR = 2.84; pc = 0.002	(Dubaniewicz *et al.*, 2003)
DQB1*0502	Thai	PTB	TB cases = 82; Controls = 160	OR = 2.06, p = 0.01 (pc = 0.13)	(Vejbaesya *et al.*, 2002)
DQB1*0503	Cambodian	PTB	TB cases = 126; Controls = 88	p = 0.005	(Goldfeld *et al.*, 1998)
DR2	Indian	PTB	25 Families	p = 0.001	(Singh *et al.*, 1983)
DR2	Russian	PTB	N/A	Increased Frequency	(Khomenko *et al.*, 1990)
DR2	Indian (South)	Smear pos. PTB	TB cases = 204; Controls = 404	Attributable risk = 0.29; p = 0.01	(Brahmajothi *et al.*, 1991)
DR2	Indian (North)	PTB	TB cases = 153; Controls = 289	RR = 1.8; pc = 0.029	(Rajalingam *et al.*, 1996)
DR2; DQ1	Indian	PTB	TB cases = 209; Controls = 122	RR = 2.3; RR = 2.8	(Selvaraj *et al.*, 1998)
DR2; DQw1	Indonesian	PTB	TB cases = 101; Controls = 65	Attributable risk = 36% & 39%	(Bothamley *et al.*, 1989)
DR2; DRw53	Tuvinian Russian	PTB	N/A	N/A	(Pospelov *et al.*, 1996)
DR3	Mexican	PTB	TB cases = 51; Controls = 54	Decreased Frequency	(Cox *et al.*, 1988)
DR3	Russian	PTB	N/A	Decreased Frequency	(Khomenko *et al.*, 1990)
DR4	Thai	PTB	TB cases = 35; Controls = 35?	Increased Frequency	(Chandanayingyong *et al.*, 1988)
DR4	Italian	Cavitary TB	TB cases = 54; Controls = 1089	RR = 2.7; p = 0.001	(Ruggiero *et al.*, 2004)
DRB1 *13 & DRB1*14	Tuvinian Russian	PTB	14 Pedigrees	Transmitted more frequently	(Pospelova *et al.*, 2005)
DRB1*07; DQA1*0101	Iranian	PTB	TB cases = 40; Controls = 100	OR = 2.7; OR = 2.66	(Amirzargar *et al.*, 2004)
DRB1*0803	Korean	Drug Resistant PTB	TB cases = 81; Controls = 200	OR = 2.63; pc = 0.047	(Kim *et al.*, 2005a)
DRB1*11	Chinese	PTB	TB cases = 74; Controls = 90	RR = 0.12; $p < 0.05$	(Wang *et al.*, 2001)
DRB1*13	Polish	PTB	TB cases = 31; Controls = 58	RR = 0.04; $p < 0.001$	(Dubaniewicz *et al.*, 2000)

Candidate gene	Population	Phenotype	Sample Size	Reported Results	Reference
DRB1*15	Chinese	PTB	TB cases = 74; Controls = 90	RR = 2.91; p < 0.05	(Wang *et al.*, 2001)
DRB1*1501	Indian	PTB	TB cases = 22; Controls = 36	p < 0.05	(Sriram *et al.*, 2001)
DRB1*1501; DQB1*0601	Southern Indian	Sputum pos. PTB	TB cases = 126; Controls = 87	OR = 2.68; OR = 2.32	(Ravikumar *et al.*, 1999)
DRB1*16	Polish	PTB	TB cases = 31; Controls = 58	RR = 9.7; p < 0.01	(Dubaniewicz *et al.*, 2000)
HSP70-1A	Indian (North)	DR15 neg. PTB	N/A	RR = 12.6; p = 0.02	(Rajalingam *et al.*, 2000)
TAP-A/F	Indian (North)	PTB	TB cases = 57; Controls - 40	RR = 4.3; pc = 0.01	(Rajalingam *et al.*, 1997)
TNF	Sicilian	PTB	TB cases = 45; Controls = 100	p = 0.05	(Scola *et al.*, 2003)
TNF (-308G; -238A)	Colombian	PTB	TB cases = 135; Controls = 430	OR = 1.8; OR = 2.2	(Correa *et al.*, 2005)
TNF -238	Gambian	PTB	TB cases = 206; Controls = 229	OR = 2.54; p = 0.00001	(Hill *et al.*, 1996)
Other candidate genes					
BCHE (cholinesterase)	Russian	PTB	N/A	RR = 6.92	(Gadzhiev *et al.*, 1987)
CCL18 (rs2015086; rs14304)	Brazilian	PTB	92 Pedigrees (627 Ind)	RR = 0.4; RR = 0.38	(Jamieson *et al.*, 2004)
CCL2 (-2518G)	Mexican	PTB	TB cases = 445; Controls = 334	OR = 2.43; p = 0.0003	(Flores-Villanueva *et al.*, 2005)
CCL2 (-2518G)	Korean	PTB	TB cases = 129; Controls = 162	OR = 2.63; p = 0.0001	(Flores-Villanueva *et al.*, 2005)
CCL4 (rs1719144)	Brazilian	PTB	92 Pedigrees (627 Ind)	RR = 0.35; p = 0.002	(Jamieson *et al.*, 2004)
CR1 (Q1022H)	Malawian	HIVneg. PTB	TB cases = 196; Controls = 670	OR = 3.12; p = 0.03	(Fitness *et al.*, 2004)
ESD (Esterase)	Tuvinian Russian	PTB	TB cases = 73; Controls = 251	N/A	(Matrakshin *et al.*, 1993)
GT(n) & D543N	Japanese	PTB	TB cases = 202; Controls = 267	OR = 2.07; p = 0.0003	(Gao *et al.*, 2000)

Candidate gene	Population	Phenotype	Sample Size	Reported Results	Reference
Haptoglobin (2:2)	Russian	PTB	TB cases = 223; Controls = 567	Significantly Increased	(Kharakter *et al.*, 1990)
IFNG	Caucasian	Smear pos. PTB	TB cases = 113; Controls = 207	OR = 3.75; 0.0017	(Lopez-Maderuelo *et al.*, 2003)
IFNG	South African	PTB	TB cases = 313; Controls = 235	OR = 1.64; p = 0.0055	(Rossouw *et al.*, 2003)
IFNG (874 TT)	Sicilian	PTB	TB cases = 45; Controls = 97	p = 0.02	(Lio *et al.*, 2002)
IFNG (874T/A)	Hong Kong Chinese	PTB	TB cases = 385; Controls = 451	OR = 2.24; p < 0.001	(Tso *et al.*, 2005)
IFNG (874T/A)	Colombian	TB	TB cases = 190; Controls = 135	p = 0.01	(Henao *et al.*, 2006)
IFNG (T/T 874)	Croatian	Microscopy pos. TB	TB cases = 54; Controls = 175	OR = 3.12; p = 0.012	(Etokebe *et al.*, 2006)
IFNGR1	Croatian	PTB	TB cases = 120; Controls = 87	p = 0.02	(Fraser *et al.*, 2003)
IFNGR1 (-611A, -56C)	Caucasian	DNTM & PTB	TB cases = 55; Controls= 86	Increased in Cases; p = 0.004 & 0.03	(Rosenzweig *et al.*, 2004)
IFNGR1 (FA1 (CA)n)	Croatian	PTB	TB cases = 244; Controls = 521	OR = 0.24; p = 0.0023	(Bulat-Kardum *et al.*, 2006)
IL10 (-1082)	Cambodian	PTB	TB cases = 358; Controls = 106	OR = 1.84; p = 0.01	(Delgado *et al.*, 2002)
IL10 (-1082)	Malawian	HIVpos. PTB	TB cases = 155; Controls = 541	OR = 0.37; p = 0.007	(Fitness *et al.*, 2004)
IL10 (-1082A)	Sicilian	PTB	TB cases = 45; Controls = 100	p = 0.05	(Scola *et al.*, 2003)
IL1B (+3953 TT/TC)	Columbian	PTB	TB cases = 122; Controls = 166	OR = 0.3; p = 0.001	(Gomez *et al.*, 2006)
IL1B (-511C)	Gambian	PTB	TB cases = 335; Controls = 298	OR = 0.58; p = 0.015	(Awomoyi *et al.*, 2005)
IL1RA	Gambian	PTB	TB cases = 404; Controls = 417	OR = 0.46; p = 0.032	(Bellamy *et al.*, 1998)
IL8 (-251A)	Caucasian	PTB	TB cases = 106; Controls = 107	OR = 3.41; p < 0.006	(Ma *et al.*, 2003)

Candidate gene	Population	Phenotype	Sample Size	Reported Results	Reference
IL8 (-251A)	African American	PTB	TB cases = 180; Controls = 167	OR = 3.46; p < 0.01	(Ma *et al.*, 2003)
MBP (B)	South African	TB Meningitis	TB cases = 91; Controls = 79	p = 0.017	(Hoal-van Helden *et al.*, 1999)
MBP (B)	African American	PTB	TB cases = 176; Controls = 71	OR = 0.34; p < 0.01	(El Sahly *et al.*, 2004)
MBP (B; C)	West African	TB Incidence	626 Individuals	r = 0.565; t = 2.273	(Mombo *et al.*, 2003)
MBP (B; C; D & X)	Danish	PTB	TB cases = 59; Controls = 250	p = 0.03	(Soborg *et al.*, 2003)
MBP (B; C; D)	Indian	PTB	TB cases = 202; Controls = 109	OR = 6.5; p = 0.008	(Selvaraj *et al.*, 1999)
MBP (C)	Gambian	PTB	TB cases = 397; Controls = 422	OR = 0.79; p = 0.037	(Bellamy *et al.*, 1998)
MBP (C)	Malawian	HIVpos. PTB	TB cases = 154; Controls = 546	OR = 1.69; p = 0.034	(Fitness *et al.*, 2004)
NOS2A (-1026)	Brazilian	PTB	92 Pedigrees (627 Ind)	RR = 3.25; p = 0.021	(Jamieson *et al.*, 2004)
P2RX7	Gambian	PTB	TB cases = 646; Controls = 694	OR = 0.70; p = 0.003	(Li *et al.*, 2002)
PGM1 (*2+ allele)	South Indian	PTB	TB cases = 204; Controls = ?	N/A	(Papiha *et al.*, 1987)
PGM1 (Phosphoglucomutase)	Indian	PTB	N/A	Significant Difference	(Papiha *et al.*, 1983)
SLC11A1 (3' UTR)	Korean	PTB	TB cases = 192; Controls = 192	OR = 1.85; p = 0.02	(Ryu *et al.*, 2000)
SLC11A1 (CAAA+/del)	Malawian	HIV neg./pos. PTB	TB cases = 239/259; Controls = 762	OR = 0.65; OR = 0.70	(Fitness *et al.*, 2004)
SLC11A1 (D2S424 (distal to SLC11A1))	Canadian Indian	PTB	1 Family (81 Ind)	LOD = 3.81; p = 0.00001	(Greenwood *et al.*, 2000)
SLC11A1 (D543N & TGTG+/del)	Cambodian	PTB	TB cases = 358; Controls = 106	OR = 0.59; p = 0.02	(Delgado *et al.*, 2002)
SLC11A1 (D543N & TGTG+/del)	Chinese Han	PTB	TB cases = 110; Controls = 180	OR = 1.93; OR = 2.22	(Liu *et al.*, 2003)

Candidate gene	Population	Phenotype	Sample Size	Reported Results	Reference
SLC11A1 (D543N)	Japanese	Cavitary Lesion in TB	TB cases = 95; Controls = 90	OR = 5.16	(Abe *et al.*, 2003)
SLC11A1 (D543N; TGTG+/del)	Chinese Han	PTB	TB cases = 120; Controls = 240	OR = 2.59; OR = 1.89	(Liu *et al.*, 2004)
SLC11A1 (GT(9) & TGTG+/del)	South African	PTB	TB cases = 265; Controls = 224	p = 0.002 & p = 0.013	(Hoal *et al.*, 2004)
SLC11A1 (GT(n) & 274C/T)	Houston (USA)	Pediatric TB	184 Nuclear Families	p = 0.04 & OR = 1.75; p = 0.01	(Malik *et al.*, 2005)
SLC11A1 (GT(n))	Gambian	PTB	TB cases = 329; Controls = 324	OR = 1.40; p = 0.024	(Awomoyi *et al.*, 2002)
SLC11A1 (GT(n))	Caucasian US	PTB	TB cases = 135; Controls = 108	OR = 2.02	(Ma *et al.*, 2002)
SLC11A1 (GT(n); INT4; D543N & TGTG+/d)	Gambian	Smear pos. PTB	TB cases = 410; Controls = 417	OR = 4.07; $p < 0.001$	(Bellamy *et al.*, 1998)
SLC11A1 (INT4 & D543N)	Chinese	Severe TB	TB cases = 127; Controls = 91	OR = 2.29; OR = 2.27	(Zhang *et al.*, 2005)
SLC11A1 (INT4 + D543N)	Korean	NTM Lung Disease	TB cases = 41; Controls = 50	OR = 10.88; p = 0.04	(Koh *et al.*, 2005)
SLC11A1 (INT4 C/C, D543N G/A)	Peruvian	PTB	TB cases = 507; Controls = 513	OR = 1.72 & 1.40; $p < 0.05$	(Taype *et al.*, 2006)
SLC11A1 (INT4)	Guinea-Conakry	PTB	44 Families (160 Ind)	$\chi^2 = 4.14$; p = 0.036	(Cervino *et al.*, 2000)
SLC11A1 (INT4)	Danish	Microscopy pos. TB	TB cases = 104; Controls = 176	RR = 1.9; p = 0.013	(Soborg *et al.*, 2002)
SLC11A1 (region)	Brazilian	PTB	37 Pedigrees (287 Ind)	LOD = 0.51; p = 0.025	(Shaw *et al.*, 1997)
SLC11A1 (TGTG+/del)	Chinese Han	PTB	TB cases = 147; Controls = 145	$\chi^2 = 7.79$; $p < 0.01$	(Duan *et al.*, 2003)
SP110	Republic of Guinea	PTB	99 Families	$p < 0.015$	(Tosh *et al.*, 2006)
SP110	Guinea-Bissau	PTB	102 Families	p = 0.002	(Tosh *et al.*, 2006)
SP110	Gambian	PTB	219 Families	$p < 0.02$	(Tosh *et al.*, 2006)
SP-A1	Mexican	PTB	TB cases = 107; Controls = 101	OR = 4.51, p = 0.008	(Floros *et al.*, 2000)

Candidate gene	Population	Phenotype	Sample Size	Reported Results	Reference
SP-A1 (307A, 776T)	Ethiopian	PTB	181 Pedigrees (226 TB cases)	$p < 0.019$	(Malik *et al.*, 2006)
SP-A2	Mexican	PTB	TB cases = 107; Controls = 102	OR = 9.57, $p = 0.38$	(Floros *et al.*, 2000)
SP-A2 (355C, 751C)	Ethiopian	PTB	181 Pedigrees (226 TB cases)	$p < 0.042$	(Malik *et al.*, 2006)
SP-A2 (A1660G; G1649C)	Indian	PTB	TB cases = 17; Controls = 19	OR = 16.3; $p< 0.001$	(Madan *et al.*, 2002)
SP-B (B1012_A, AAGG_1)	Mexican	PTB	TB cases = 107; Controls = 103	OR = 2.36; OR = 0.12	(Floros *et al.*, 2000)
STAT5B (rs2230097)	Brazilian	PTB	92 Pedigrees (627 Ind)	RR = 0.36; $p = 0.038$	(Jamieson *et al.*, 2004)
TLR2 (Arg753Gln)	Turkish	PTB	TB cases = 151; Controls = 116	1.60 - 6.04 Fold Increase	(Ogus *et al.*, 2004)
TLR2 (C2029T)	Tunisian	PTB	TB cases = 33; Controls = 33	$p < 0.0001$	(Ben Ali *et al.*, 2004)
UBE3A	African	PTB	180 Pedigrees	$\chi^2 = 4.17$; $p = 0.03$	(Cervino *et al.*, 2002)
VDR (BsmI; FokI)	Indian	Spinal TB	TB cases = 64; Controls = 103	OR = 2.2; OR = 2.4	(Selvaraj *et al.*, 2004)
VDR (ff)	Gujarati Asian	Extra-pulmonary TB	TB cases = 52; Controls = 116	OR = 2.8	(Wilkinson *et al.*, 2000)
VDR (ff)	Chinese Han	PTB	TB cases = 76; Controls = 171	OR = 3.67	(Liu *et al.*, 2003)
VDR (-ff)	Chinese Han	PTB	TB cases = 120; Controls = 240	OR = 2.35; $p = 0.03$	(Liu *et al.*, 2004)
VDR (FokI; BsmI; ApaI; TaqI)	West African	PTB	382 Trios	$\chi^2 = 22.11$; $p = 0.009$	(Bornman *et al.*, 2004)
VDR (TaqI; FokI)	Peruvian	TB Treatment	TB cases = 103; Controls = 206	RR = 5.6; RR = 9.6	(Roth *et al.*, 2004)
VDR (tt)	Gambian	PTB	TB cases = 408; Controls = 414	OR = 0.53; $p = 0.01$	(Bellamy *et al.*, 1999)

Table 1.6 Articles reporting no significant linkage or association between possible susceptibility genes and TB (Burgner *et al*., 2006).

Candidate Gene	Population	Phenotype	Sample Size	Reported Results	Reference
MHC Region (Class I, II & III)					
Aw30; Aw33; B7; B15; B17	Mexican American	PTB	TB cases = 100; Controls = 100	ns	(Cox *et al*., 1982)
HLA-A, -B, -C	Northern Indian	PTB	TB cases = 124; Controls = 109	ns	(Singh *et al*., 1983)
DR2	Northern Indian	PTB	TB cases = 124; Controls = 109	ns (after correction)	(Singh *et al*., 1983)
HLA	Indian	PTB	21 Families	No Linkage	(Singh *et al*., 1984)
HLA-A & -B	Hong Kong Chinese	PTB	TB cases = 256; Controls = 100	ns	(Hawkins *et al*., 1988)
DRB; DQA, DQB	South Indian	PTB	TB cases = 38; Controls = 36; 12 families	No association; No linkage	(Sanjeevi *et al*., 1992)
TNF	Brazilian	PTB	37 pedigrees (287 Ind)	LOD = 0.01	(Shaw *et al*., 1997)
TNF	Cambodian	PTB	TB cases = 126; Controls = 88	ns	(Goldfeld *et al*., 1998)
HLA	Brazilian	PTB	98 Pedigrees (704 Ind)	ns	(Blackwell, 2001)
TNF -238; -308	Indian	PTB	TB cases = 210; Controls = 120	ns	(Selvaraj *et al*., 2001)
LT	Indian	PTB	TB cases = 210; Controls = 120	ns	(Selvaraj *et al*., 2001)
TNF -1030; -862; -856; -307	Cambodian	PTB	TB cases = 358; Controls = 106	$p > 0.05$	(Delgado *et al*., 2002)
HLA -A; -B; -C	Italian	Current TB	TB cases = 68; Controls = 1089	ns	(Ruggiero *et al*., 2004)
TAP1	Korean	PTB	TB cases = 219; Controls = 210	ns	(Roh *et al*., 2004)
DQB1*0501	Iranian	PTB	TB cases = 40; Controls = 100	ns	(Amirzargar *et al*., 2004)
TNF (-238; -308; -376; -893)	Malawian	HIV neg./ pos. PTB	TB cases ~ 181/144; Controls = 417	ns	(Fitness *et al*., 2004)
LTA (5' UTR m/sat)	Malawian	HIV neg./ pos. PTB	TB cases = 198/237; Controls = 707	ns	(Fitness *et al*., 2004)

Candidate Gene	Population	Phenotype	Sample Size	Reported Results	Reference
TNF (-308)	Colombian	TB	TB cases = 190; Controls = 135	ns	(Henao *et al.*, 2006)
TAP1*0101; TAP2*0101	Columbian	PTB	TB cases = 122; Controls = 166	ns	(Gomez *et al.*, 2006)
Other Candidates:					
CCL3 (-906 promoter m/sat)	Malawian	HIV neg./ pos. PTB	TB cases = 147/206; Controls = 580	ns	(Fitness *et al.*, 2004)
CCL3 (-459)	Mexican	PTB	TB cases = 445; Controls = 518	ns	(Flores-Villanueva *et al.*, 2005)
CCL5 (-471)	Mexican	PTB	TB cases = 445; Controls = 518	ns	(Flores-Villanueva *et al.*, 2005)
CD14 (-159CT)	Columbian	PTB	TB cases = 267; Controls = 112	ns	(Pacheco *et al.*, 2004)
CD14 (-159)	Polish	PTB	Not Given Abstract Only	ns	(Druszczynska *et al.*, 2006)
CXCR1	Caucasian US	PTB	TB cases = 106; Controls = 107	ns	(Ma *et al.*, 2003)
CXCR1	African American	PTB	TB cases = 180; Controls = 167	ns	(Ma *et al.*, 2003)
CXCR2	Caucasian US	PTB	TB cases = 106; Controls = 107	ns	(Ma *et al.*, 2003)
CXCR2	African American	PTB	TB cases = 180; Controls = 167	ns	(Ma *et al.*, 2003)
Haptoglobin (HP)	Zimbabwean	PTB	TB cases = 98; Controls = 98	$p = 0.50$	(Kasvosve *et al.*, 2000)
ICAM1 (179A/T)	Malawian	HIV neg./ pos. PTB	TB cases = 209/217; Controls = 596	ns	(Fitness *et al.*, 2004)Fit4
IFNA17 (551T/G)	Japanese	PTB	TB cases = 114, Controls = 110	$p = 0.155$	(Akahoshi *et al.*, 2004)
IFNB (153C/T)	Japanese	PTB	TB cases = 114, Controls = 110	$p = 0.137$	(Akahoshi *et al.*, 2004)
IFNG (1348T/A)	Japanese	PTB	TB cases = 114, Controls = 110	$p = 0.55$	(Akahoshi *et al.*, 2004)
IFNG (874T/A)	Malawian	HIV neg./ pos. PTB	TB cases = 213/238; Controls = 703	ns	(Fitness *et al.*, 2004)
IFNG (T874A, G2109A)	Croatian	PTB	TB cases = 253; Controls = 519	ns	(Etokebe *et al.*, 2006)

Candidate Gene	Population	Phenotype	Sample Size	Reported Results	Reference
IFNGR1	Gambian	PTB	TB cases = 297; Controls = 285	ns	(Awomoyi *et al.*, 2004)
IFNGR1 (167T/C)	Japanese	PTB	TB cases = 114, Controls = 110	p = 0.213	(Akahoshi *et al.*, 2004)
IFNGR1 (-611; -56)	African American	PTB	TB cases = 76; Controls = 114	ns	(Rosenzweig *et al.*, 2004)
IFNGR1 (-611; -56)	Caucasian	PTB	TB cases = 70; Controls = 128	ns	(Rosenzweig *et al.*, 2004)
IFNGR1 (395)	Iranian	PTB	TB cases = 50; Controls = 54	ns	(Mirsaeidi *et al.*, 2006)
IFNGR2 (839G/A)	Japanese	PTB	TB cases = 114, Controls = 110	p = 0.498	(Akahoshi *et al.*, 2004)
IkBL (T738C)	Columbian	PTB	TB cases = 122; Controls = 166	ns	(Gomez *et al.*, 2006)
IL10	Gambian	PTB	TB cases = 792; Controls = 816	ns	(Bellamy *et al.*, 1998)
IL10	Caucasian	PTB	TB cases = 113; Controls = 207	ns	(Lopez-Maderuelo *et al.*, 2003)
IL12 - 3' UTR	Russian	PTB	TB cases = 58; Controls = 127	ns	(Puzyrev *et al.*, 2002)
IL12p40	Caucasian US	PTB	TB cases = 106; Controls = 107	ns	(Ma *et al.*, 2003)
IL12p40	African American	PTB	TB cases = 180; Controls = 167	ns	(Ma *et al.*, 2003)
IL12RB1 (641A/G)	Japanese	PTB	TB cases = 114, Controls = 110	p = 0.610	(Akahoshi *et al.*, 2003)
IL12RB1 (+705, +1158, +1196, +1637, +1664)	Koreans	PTB	TB cases = 115; Controls = 151	ns	(Lee *et al.*, 2005)
IL12RB2 (365C/T)	Japanese	PTB	TB cases = 114, Controls = 110	p = 0.59	(Akahoshi *et al.*, 2003)
IL1RA	Gujarati Asian	PTB	TB cases = 89; Controls = 114	ns	(Wilkinson *et al.*, 1999)
IL1RA	Indian	PTB	TB cases = 202; Controls = 109	ns	(Selvaraj *et al.*, 2000)

Candidate Gene	Population	Phenotype	Sample Size	Reported Results	Reference
IL1RA	Cambodian	PTB	TB cases = 358; Controls = 106	$X^2 = 3.27$; p = 0.19	(Delgado *et al.*, 2002)
IL1beta	Gujarati Asian	PTB	TB cases = 89; Controls = 114	ns	(Wilkinson *et al.*, 1999)
IL1beta -511; +3953	Cambodian	PTB	TB cases = 358; Controls = 106	p = 0.32 & 0.78	(Delgado *et al.*, 2002)
IL1RN	Gambian	PTB	TB cases = 35; Controls = 298	ns	(Awomoyi *et al.*, 2005)
IL6 (-174G/C)	Colombian	TB	TB cases = 190; Controls = 135	ns	(Henao *et al.*, 2006)
IL8 -251 & +781	Gambian	PTB	TB cases = 284; Controls = 245	p = 0.50; p = 0.42	(Cooke *et al.*, 2004)
MBP (B; C; D)	Hispanic	PTB	TB cases = 198; Controls = 46	ns	(El Sahly *et al.*, 2004)
MBP (B; C; D)	Caucasian	PTB	TB cases = 113; Controls = 69	ns	(El Sahly *et al.*, 2004)
MBP (C; D)	African American	PTB	TB cases = 176; Controls = 71	ns	(El Sahly *et al.*, 2004)
MBP (codons 52, 54 & 57)	Polish	PTB	Not Given Abstract Only	ns	(Druszczynska *et al.*, 2006)
MMP-1 1G/2G	Japanese	PTB	TB cases = 105; Controls = 106	ns	(Ninomiya *et al.*, 2004)
NLI-IF	Caucasian US	PTB	TB cases = 94; Controls = 145	$p > 0.05$	(Ma *et al.*, 2002)
NOD2	Gambian	PTB	TB cases = 320; Controls = 320	ns	(Stockton *et al.*, 2004)
NOS2A (-954)	Mexican	PTB	TB cases = 445; Controls = 518	ns	(Flores-Villanueva *et al.*, 2005)
SLC11A1	Russian (Slavic)	PTB	TB cases = 58; Controls = 127	ns	(Puzyrev *et al.*, 2002)
SLC11A1	Taiwanese	PTB	TB cases = 49; Controls = 48	ns	(Liaw *et al.*, 2002)
SLC11A1	Moroccan	PTB	116 Pedigrees	ns	(El Baghdadi *et al.*, 2003)

Candidate Gene	Population	Phenotype	Sample Size	Reported Results	Reference
SLC11A1 (1703G/A)	Japanese	PTB	TB cases = 114, Controls = 110	$p = 0.144$	(Akahoshi *et al.*, 2004)
SLC11A1	Polish	PTB	TB cases = 85; Controls = 93	ns	(Dubaniewicz *et al.*, 2005)
SLC11A1 (ATA(n))	Gambian	PTB	TB cases = 318; Controls = 146	ns	(Awomoyi *et al.*, 2006)
SLC11A1 (INT4)	Polish	PTB	Not Given Abstract Only	ns	(Druszczynska *et al.*, 2006)
SLC11A2	South African	PTB	TB cases = 265; Controls = 224	ns	(Hoal *et al.*, 2004)
SPP1 (2514C/T)	Japanese	PTB	TB cases = 114, Controls = 110	$p = 0.643$	(Akahoshi *et al.*, 2004)
TGFbeta1 (T869C)	Japanese	PTB	TB cases = 101; Controls = 110	ns	(Niimi *et al.*, 2002)
TGFbeta1	Colombian	TB	TB cases = 190; Controls = 135	ns	(Henao *et al.*, 2006)
TLR2 (Int 2 m/sat)	Malawian	HIV neg./ pos. PTB	TB cases = 215/249; Controls = 742	ns	(Fitness *et al.*, 2004)
TLR2 (Arg677Trp; Arg753Gln)	Korean	NTM Lung Disease	TB cases = 80; Controls = 84	ns	(Ryu *et al.*, 2006)
TLR4 (896A/G)	Malawian	HIV neg./ pos. PTB	TB cases = 162/120; Controls = 427	ns	(Fitness *et al.*, 2004)
TLR4	Gambian	PTB	TB cases = 307; Controls = 298	$X^2 = 0.19$; $p = 1.00$	(Newport *et al.*, 2004)
TNFSF5 (CD40)	West African	PTB	121 Trios	ns	(Campbell *et al.*, 2003)
VDR	Cambodian	PTB	TB cases = 358; Controls = 106	$X^2= 0.99$; $p = 0.60$	(Delgado *et al.*, 2002)
VDR (TaqI; ApaI; BsmI)	Malawian	HIV neg./ pos. PTB	TB cases ~212/225; Controls = 672	ns	(Fitness *et al.*, 2004)

DNTM = Disseminated non-tuberculous Mycobacterial
Ind = Individuals
LOD = Logarithm of the odds
N/A = Not available
ns = Not significant
NTM = Non-Tuberculous Mycobacterial
OR = Odds Ratio
pc = Corrected p-value
PTB = Pulmonary Tuberculosis
RR = Relative risk

CHAPTER 2

GENETIC STRUCTURE OF THE SOUTH AFRICAN COLOURED POPULATION

2.1 INTRODUCTION

Sampling for human genetic studies begins by collecting samples from a predefined population. Usually populations are subjectively defined based on their geographic location, cultural or other characteristic features and perceived racial information even though it might not reflect their genetic makeup (Foster & Sharp, 2002). Race and ethnicity are two very controversial concepts among the general public as well as scientists. Historically, geneticists have had one of two perspectives on these two concepts, one being the eugenics movement that believed in making use of scientific analysis to classify individuals biologically (Huxley, 1951) and the other being the physical anthropology social movement that believed mainly in historical contributions and very little in biological contributions (Boas, 1942). At present the differences between geneticists about these two contrasting perspectives still needs to be resolved.

The word "race" derived from the Spanish word "raza", meaning breed or stock, was a product of the European era of exploration (Smedley, 1999). There are two arguments against racial categorization, one: that race has no biological basis (Editorial, 2001; Schwartz, 2001) and two: that there are racial differences but they are merely cosmetic and do not reflect any additional genetic distinctiveness (Wilson *et al.*, 2001). However, according to Risch (Risch *et al.*, 2002) race is a key factor in genetic studies as it explains a portion of our genetic variation. These patterns of variation provide a biological justification of the use of traditional racial categories (Race ethnicity and genetics working group, 2005). On the other hand Goldstein (Goldstein & Hirschhorn, 2004) argues that race is meaningless and that the focus must shift from geographic ancestry to identifying causal variants, as they would be likely to be informative across the world. Individuals such as Schwartz and Goldstein believe there is only one race – the human race (Schwartz, 2001).

Social labels for different groups such as Chinese, African Americans and Yoruba can be very misleading for genetic studies. The Yoruba population is widely spread over West Africa and includes other smaller groups, therefore labels such as these can mask significant intragroup variation (Foster & Sharp, 2002). Of the genetic variation between individuals, 90 to 95% occurs within human population groups whereas the remaining 5 to

10% occurs between population groups (Jorde *et al.,* 2000; Lewontin, 1972; Rosenberg *et al.*, 2002). This variation has an influence on the average differences in physical characteristics, disease susceptibility and treatment outcome among different populations (Bamshad *et al.*, 2003).

2.1.1 Heterogeneity in association studies

The most common approach used in genetic association studies is a case-control design. This approach needs to be carried out very carefully, since false-positive or negative associations might occur due to the absence of genetic homogeneity amongst the cases and controls (Pritchard & Rosenberg, 1999; Wacholder *et al.*, 2002). Irrespective of the disease status, allele frequencies vary within and between populations, due to their unique genetic and historic background shaped by, amongst others, migration, non-random mating, genetic drift and stochastic variation that can produce a variety of results (Cardon & Palmer, 2003; Satten *et al.*, 2001).

Genetic heterogeneity may arise due to population substructure, which may be cryptic, or due to recent admixture (Ziv & Burchard, 2003). If samples from diseased and healthy groups have different proportions of each ethnic group or, in the case of admixture, different percentages of ancestry from each ancestral subpopulation, they would be classified as stratified, whereas cases and controls that contain equal amounts of each ethnic group would not be stratified (Pritchard & Rosenberg, 1999). Therefore population stratification, by definition, does not occur in homogenous populations (Wacholder *et al.*, 2002).

One of the main reasons why case-control association studies are often not replicated, or produce inconsistent results, can be ascribed to population stratification or population substructure that can induce false positives (Cardon & Palmer, 2003). Usually it is assumed that in a study where a certain mutation is more frequent in the cases (affected individuals) than controls (unaffected individuals), it is associated with disease susceptibility. If a marker locus close to the causative mutation is in linkage disequilibrium with the mutation, differences in allele frequencies between cases and controls will also be detectable (Pritchard & Donnelly, 2001). However in the presence of population stratification this approach would not be valid since high rates of "association" would be detected with

markers that are unlinked to the disease loci (Knowler *et al.,* 1988; Lander & Schork, 1994; Pritchard & Donnelly, 2001).

The contribution of different ancestries to individuals may vary tremendously in an admixed population. For example, African Americans have different proportions of African ancestry, but averaging 20% (Pfaff *et al.*, 2001), Puerto Ricans have ± 30% African ancestry and Mexicans have ± 9% African ancestry (Hanis *et al.*, 1991). Recent population admixture can mask or reverse the true genetic effects underlying a complex disease, therefore it is important that population admixture should be tested in an a association study to ensure that the results are not an artifact (Deng, 2001).

"*Cryptic population structure*" can be defined as structure at a genetic level that is not well described according to ethnic or national labels (Pritchard *et al.*, 2000). For population stratification to affect genetic association studies, two circumstances must be met:

1) Differences in disease prevalence between cases (affected) and controls (unaffected) must be present
2) Variations in allele frequency must exist between groups (Wacholder *et al.*, 2000).

The presence of just one of these criteria is not sufficient to cause population stratification (Cardon & Palmer, 2003).

Since population stratification can arise when the genetic background of cases and controls is different, it is important that both these groups should be sampled from the same place and that cases and controls should be carefully matched on their ethnic and genetic backgrounds (Cardon & Palmer, 2003; Reich & Goldstein, 2001). Case-control association studies are valid tests for association, provided that the issue of population stratification is addressed (Pritchard & Rosenberg, 1999).

2.1.2 Confounding

Population structure may confound genetic association studies (Lander & Schork, 1994; Spielman *et al.*, 1993). The results of genetic association studies with unrelated individuals (cases and controls) may be confounded due to substructure or recent admixture present in

a population (Rosenberg *et al.*, 2002). This might lead to an excess of both false negative (non significant) or false positive (significant), also known as spurious association, results (Chakraborty & Smouse, 1988; Deng, 2001; Ziv & Burchard, 2003). A false positive association can be defined as an association between a marker locus and a phenotype when the marker is unlinked to any causal loci (Pritchard & Rosenberg, 1999). False positives will be a more serious problem than false negatives, since setting the conventional significant *P* value at 0.05 would mean that out of a set of 1000 candidate alleles of which only 10 are truly causal, one would expect 50 false positives (1000 x 5%). This number is considerably more than the actual 5 true positives (10 x 50%) in a study with 50% power. However, both are detrimental to research (Thomas & Witte, 2002; Ziv & Burchard, 2003). The main confounding factor in population stratification is ethnicity, which includes underlying factors such as environment and cultural influences (Fig. 2.1). However there might be explanations, other than population stratification, for false-positive reports in studies with unrelated controls (Table 2.1) (Wacholder *et al.*, 2002).

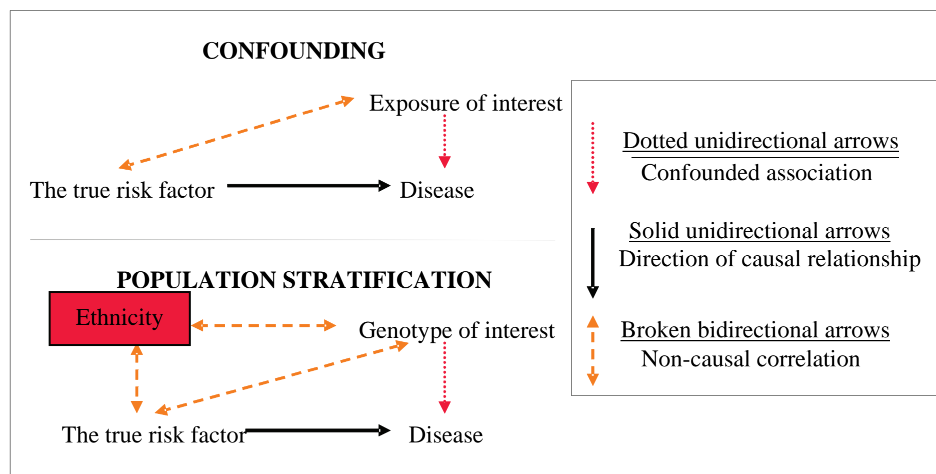


Fig. 2.1. Classic confounding and population stratification. In confounding, an exposure of interest correlated with the true risk factor but not truly causal of disease, can be misleadingly seen to be associated with risk of disease. In population structure, as the genotype of interest is correlated with the true risk factor because both are correlated with ethnicity, it can be incorrectly regarded to be associated with risk of disease (Wacholder *et al.*, 2002).

Mendelian randomization can be defined as the arbitrary transmission of genes from parent to offspring that occurs during gamete formation and conception, therefore the way the phenotype and alleles co-segregate during transmission from parent to offspring (Davey Smith & Ebrahim, 2003). Mendelian randomization relies on getting good estimates from

genetic association studies. Population stratification is of great importance as it could lead to confounded results when the gene-risk factor association and the gene-outcome association is studied in different populations (Lawlor *et al.*, 2008).

"The basic principle of Mendelian randomization is that genetic variants which mirror the biological effects of a modifiable environmental exposure and alter disease risk should be associated with disease risk to the extent predicted by their influence on exposure to the risk factor" (Davey Smith & Ebrahim, 2003; Davey Smith, 2006).

The term Mendelian randomization was first used in assessing the efficiency of sibling bone marrow transplantations in the treatment of acute myeloid leukemia (AML). This was done by comparing the outcome in patients with and without HLA-compatible siblings (Gray & Wheatley, 1991).

Table 2.1. Explanations, other than population stratification, for false-positive reports in studies with unrelated controls (Wacholder *et al.*, 2002).

Alternative reasons for false-positives
The use of clinic patients as cases and medical students as healthy controls
Low power
Common polymorphism with low penetrance and low attributable risk
Multiple comparisons *i.e.* subgroup analysis
Unmeasured environmental factors

The concern about population stratification started with the Pima Indian example (Knowler *et al.*, 1988) where the genetic frequency of a $Gm^{3,5,13,14}$ haplotype of the human immunoglobulin G, was 75% and disease rate (type 2 diabetes) 15% in individuals of tribal descent and 1% versus 60% respectively in those without any tribal ancestry. A spurious association was reported for type II diabetes in Pima Indians, because those who had more Caucasian ancestry also had lower diabetes susceptibility, therefore any markers that were over-represented in the Pima Indians compared to the Caucasians appeared to be associated with the disease, in this case type II diabetes (Pritchard & Rosenberg, 1999; Thomas & Witte, 2002). Other examples where population stratification gave false positives were with *DRD2* and alcoholism (Blum *et al.*, 1990), and *CYP3A4* and prostate cancer among African-Americans (Kittles *et al.*, 2002).

2.1.3 Genomic adjustment

Where a group of unlinked markers is used, population stratification can be addressed by one of two model based methods or a non model-based (model-free) approach:

1) Model-free genomic control (GC) approach uses an over-dispersion model to determine an empirical distribution (Bacanu *et al.,* 2000; Devlin & Roeder, 1999; Reich & Goldstein, 2001).
2) Model-based structure association approaches (SA) evaluate if stratification does exist and then proceed to an association study (Pritchard & Rosenberg, 1999).
3) Model-based latent-class approaches (LCA) distinguish between homogenous subpopulations (Kim *et al.*, 2001).

2.1.3.1 Non-model based approach

<u>i) Genomic control (GC) approach</u>

The genomic control (GC) approach makes use of the over-dispersion model. In the presence of population stratification the null distribution of the χ^2 test for association will tend to shift toward elevated values leading to more false positives. The null distribution could be determined by using enough unlinked markers that are independent of the disease: this is called GC (Bacanu *et al.,* 2000; Devlin & Roeder, 1999; Reich & Goldstein, 2001). However a major drawback of this method is that large numbers of unlinked markers would be needed, which might be too expensive even with the new SNP technology, in order to obtain accurate results (Devlin *et al.*, 2001). This method is restricted to SNPs and it was estimated that 50 or more SNPs would be needed (Bacanu *et al.,* 2000; Devlin & Roeder, 1999). The GC approach also assumes that the effect of population structure is constant over all loci (Devlin *et al.*, 2001). The genomic adjustment approach is simple calculation-wise and can deal with population stratification (Devlin & Roeder, 1999).

2.1.3.2 Model based approach to population structure

<u>i) Structure association (SA)</u>

Another approach is a two-step process where the first step involves a panel of a few dozen markers to test for population stratification and the second step involves evaluating the candidate gene association only if homogeneity is not rejected while correcting for individual admixture (Pritchard & Rosenberg, 1999). The drawback

of this method is that there are no guidelines on what to do if homogeneity is rejected.

Structure association (SA) was developed by Pritchard *et al.* (2000) with the principle that markers unlinked to the candidate genes must be used to infer subpopulation membership (Pritchard *et al.*, 2000). Pritchard and his group proposed that the latent class approach would be able to control for population substructure. The main assumption of the latent class model is that there is no confounding due to population substructure. This Bayesian approach estimates the number of subpopulations *K* (clusters) and the proportion of the genome of each individual that belongs to each *K* (Pritchard *et al.*, 2000) and accommodates admixed populations (see section 2.2). A comparison between the GC and the SA approach can be seen in Table 2.2.

ii) Latent class approach (LCA)

A novel latent-class model was introduced by Satten *et al.* (2001) to draw conclusions about the association between a candidate gene and a disease in a population where population stratification might be present. Satten *et al.* (2001) believed that this model-based approach outperformed that of Pritchard *et al.* for the following reasons:

1) This is not a two-step approach to estimate structure
2) The model is not conditional on the number of subpopulations
3) A straightforward likelihood approach is used in this model and it does not require a Gibbs sampler which is responsible for changing the number or parameter in the model
4) This model accounts for subpopulation structure between cases and controls unlike Pritchard *et al.* (2000) that infer substructure without accounting for cases and controls

2.1.4 Markers

Unlinked randomly selected markers such as restriction fragment length polymorphisms (RFLPs), single nucleotide polymorphisms (SNPs) or microsatellites, can be used to detect these confounding factors or correct for structure under the assumption that Hardy-

Weinberg equilibrium exists within populations (Pritchard *et al.,* 2000; Ziv & Burchard, 2003). It was suggested that microsatellites have more power than biallelic markers and at first it was proposed that 15-20 unlinked microsatellites or ± 30 SNPs would be sufficient to detect population substructure (Pritchard & Rosenberg, 1999).

Table 2.2. Comparison between the model-free genomic control approach (GC) and a model-based approach using structured association (SA) as an example.

Genomic control (GC)	Structured association (SA)
• More general approach, it corrects for confounding due to population substructure and admixture (Devlin *et al.,* 2001)	• Attempts to infer genetic ancestry for each individual sampled (Pritchard & Donnelly, 2001)
• Models effect of confounding in a over - dispersion model (Devlin & Roeder, 1999)	• Models population substructure via latent variables (Pritchard *et al.,* 2001)
• When the effect is ± constant across sub – populations then GC is more powerful than SA (Pritchard & Donnelly, 2001)	• When the effect varies greatly across subpopulations then SA is more powerful than GC (Pritchard & Donnelly, 2001).
• Attempts to eliminate spurious associations	• Attempts to eliminate spurious associations
Limitations	
• Assumes that effect of population structure is constant over all loci (Devlin *et al.,* 2001; Satten *et al.,* 2001)	• Primary source of confounding must be due to population substructure
• Limited to only binary markers	• Marker data must be powerful enough to infer number of subpopulations within the population

In order to determine the presence of substructure in the SAC, the well described and published method (SA), using the program STRUCTURE (Pritchard, Stephens & Donnelly 2000), was chosen.

2.2 STRUCTURE

The program STRUCTURE is based on a Bayesian model clustering algorithm that identifies subgroups with distinctive allele frequencies (Pritchard & Donnelly, 2001). It makes use of unlinked randomly selected markers across the genome to infer population structure (Pritchard *et al.*, 2000). This algorithm, as applied in the program STRUCTURE,

available at (http://pritch.bsd.uchicago.edu) (Pritchard, Stephens & Donnelly 2000), puts individuals into different population clusters (*K*) each characterized by a different set of allele frequencies. *K* is chosen in advance and the posterior probability changes between independent runs with the same *K* and is inherent to the MCMC calculation (Kim *et al.*, 2005b). The admixture model can be exemplified by an individual that has a fraction of his genome from ancestors in a certain population (*K*). All these fractions or membership of the *K* clusters must sum up to a membership coefficient of one across all the *K* clusters, whereafter the posterior average estimates of these proportions will be calculated (Kim *et al.*, 2005b). This program was successfully used in human studies where it was found to produce accurate clustering for both simulated and real data (Bamshad *et al.,* 2003; Pritchard *et al.*, 2000; Romualdi *et al.*, 2002; Rosenberg *et al.*, 2002; Wilson *et al.*, 2001). It has also been used in dog studies (Parker *et al.*, 2004).

2.2.1 Problems with the interpretation of K

When some individuals are assigned to one population to a greater extent than the rest of the group, then population structure is present (Pritchard & Donnelly, 2001). The true value of *K* might often not be known but one should aim for the smallest value of *K* that captures major structure in the dataset. The interpretation of the inferred value of *K* should be treated with caution due to the following reasons:

1) It is difficult to obtain reliable estimates of Pr (XlK), which is the probability of the genotypes given *K*, therefore a fairly *ad hoc* procedure should be chosen that will yield sensible results
2) The posterior distribution of *K* is quite dependent on the priors and modeling assumptions in the Bayesian model-based clustering
3) Biological interpretation can be difficult e.g. the clusters may not correspond with real populations and markers can be affected by gene flow (Pritchard *et al.*, 2000).

2.3 AFRICAN POPULATIONS

More than 2000 distinct ethnic groups exist on the African continent and these are largely responsible for the genetic diversity among the different African populations (Tishkoff & Williams, 2002).

Several studies done on mitochondrial DNA (mtDNA) and nuclear DNA confirmed that the African continent is home to the worlds most genetically diverse group and that there is no single representative African population (Calafell *et al.,* 1998; Hammer *et al.*, 2001; Jorde *et al.*, 2000; Relethford & Jorde, 1999; Stoneking *et al.*, 1997; Tishkoff *et al.*, 2000). It is therefore suggested that Africa is the ancestral homeland of all modern humans.

2.3.1 African origin of modern humans

Information regarding the evolutionary history of the human race will enable us to understand patterns of variation and utilize these patterns in identifying alleles and genotypes that predispose individuals to a certain disease. Archaeologically and genetically, the data supports a recent African origin (RAO). This topic remains highly debatable, since three different models can explain the evolution of modern humans (Fig. 2.2) (Tishkoff & Williams, 2002). The population history of Africa is a very complex system of population diversification due to population expansions and gene flow (Tishkoff & Williams, 2002).

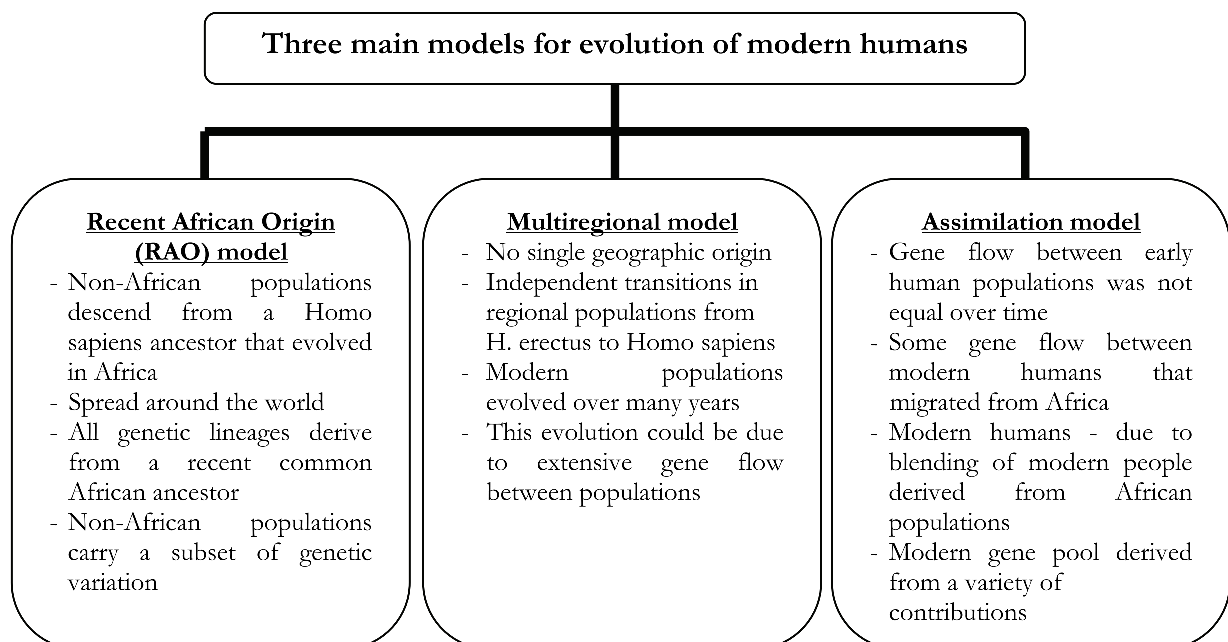


Fig. 2.2 The recent African origin model, multiregional model and assimilation model to explain the evolution of modern humans

2.3.2 South African population

Immigration from Africa, Europe and Asia to South Africa happened during the past three and a half centuries. The South African population encompasses different African groups, for example the Xhosa, Khoi, San, Zulu, Tswana and Pedi. The white South African

people descended largely from British, German, French and Dutch immigrants. The admixed population groups include the South African Coloureds and Asian people (Mountain, 2003; Nurse *et al.*, 1985). It was previously indicated that the San people were the first to have inhabited South Africa and evidence supporting this can be found in the fact that San people exhibit the most ancient mitochondrial DNA lineages (L1) (Tishkoff & Williams, 2002) and the most ancient Y-chromosome (Underhill *et al.*, 2001). The Indian population in South Africa is considered to be a hybrid population which originated from different regions of the Indian subcontinent including South India, Gujerat and the Tamil community (Nurse *et al.*, 1985). The traditional distribution of these populations can be seen in Fig. 2.3, however due to large scale migration towards developed areas these traditional areas have become much less defined (Lombard *et al.*, 2006). The genetic diversity between different South African populations can clearly be seen in the different frequencies of the human leukocyte antigen (HLA) class I (Cao *et al.*, 2004) and II alleles (Lombard *et al.*, 2006) in the different population groups.

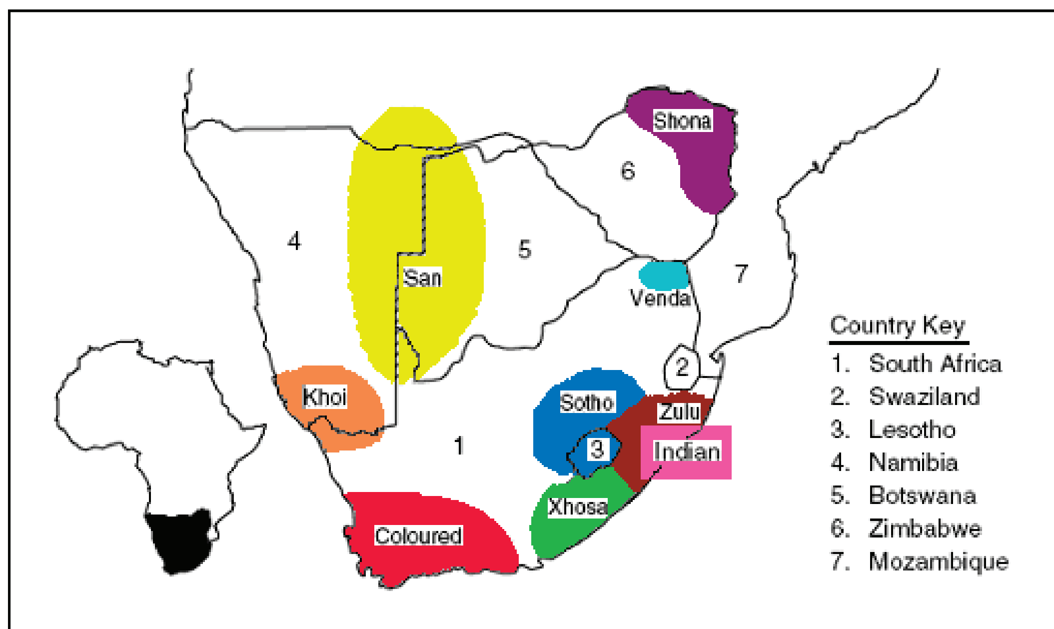


Fig. 2.3 The traditional distribution of the different population groups in South Africa (Adapted from Lombard *et al.*, 2006).

The population of South Africa, because of its geographical situation and turbulent history, is at present one of the most cosmopolitan in the world. The amount of diversity within Africa has been recognized (Tishkoff & Williams, 2002), but South Africa was the destination of previously continentally divided groups, (Europe, Asia and the rest of Africa)

all of which added to the complex society of the present day. The authorities during the apartheid era and the preceding colonial periods formalised and attempted to legislate the separation of ethnic groups in society. A highly admixed population formed in the early days of Western Cape history, and subsequent restrictive practices had the effect of creating cohesion and moulding the mixed ancestry people into a group known today as the South African Coloured population (SAC) (Adhikari, 2005; Nurse *et al.*, 1985; Van der Ross, 1993). This is the official term used by Statistics South Africa in the 1996 and 2001 national census (http://www.statssa.gov.za/census01/HTML/default.asp). According to the census, the population of South Africa in 2001 consisted of 79.0% black African, 9.6% white, 8.9% Coloured, and 2.5% Indian or Asian, with Coloureds constituting 53.9% of the population of the Western Cape province.

This unique and quintessentially South African population has suffered a great deal from marginalisation and stereotyping, under the old regime and the new (Adhikari, 2005). The population has its roots in the indigenous Khoesan living in the area that is now Cape Town and its surrounds at the time of the first colonization by European settlers of the Dutch East India Company (VOC) in 1652 (Mountain, 2003; Shell, 1994). Although the Khoekhoen (herders) and the San (hunter-gatherers) were originally distinct groups (Boonzaaier *et al.,* 1996; Elphick, 1985; Mountain, 2003), the lines became blurred in the colonial era (Elphick, 1985; Mountain, 2003), and the term Khoesan is often used to denote both groups. The VOC subsequently imported political exiles from the Dutch East Indies (Indonesia and Malaysia) (Mountain, 2004). Other slaves, from many locations, mainly around the Indian Ocean (Nurse *et al.*, 1985), including the Indian subcontinent (25.9%), Africa, mainly the east coast (26.4%), Madagascar (25.1%) and Indonesia (22.7%) (Shell, 1994) were brought in chiefly to provide farm labour. This active slave trade began in 1658 and continued until the banning of the seaborne slave trade in 1806, and the last recorded illegal imports in 1822 (Shell, 1994). In the early 1700's, the slave population in the Cape regularly outnumbered the settlers (Mountain, 2003), and men outnumbered women in both the slave and free populations (Shell, 1994). People brought in as slaves would have been incorporated in the Coloured community after emancipation of the slaves in 1838 (Shell, 1994).

The Khoekhoen were not enslaved, but frequently served as indentured labourers or serfs on the farms (Mountain, 2003; Shell, 1994). A small but significant number of women of Khoekhoe or slave descent and their children were integrated into the colonial household, often by marriage (Mountain, 2003; Shell, 1994). The early Cape society was fairly open and mixed marriages were socially acceptable, usually between European men and women who were either Khoekhoen, manumitted (freed) slaves or of mixed parentage (Keegan, 1996), and between Khoekhoen and slave (Mountain, 2003). However, in the majority of cases, and particularly after 1700, the progeny of mixed marriages and liaisons were assimilated into the growing group known as the "Cape Coloureds" (Keegan, 1996; Mountain, 2003; Nurse *et al.*, 1985), a term used since the mid nineteenth century (Keegan, 1996). These unions were more common in the farming areas (Mountain, 2003; Shell, 1994). By the late 1700's, racial restrictions were common, and these were formalised under the British administration from 1806 (Mountain, 2003), when class was more easily overcome in society than race and ancestry (Keegan, 1996).

Several missionary stations were set up for the indigenous people, with the first at Genadendal in 1738 (Mountain, 2004). The missionary influences amongst the aboriginal and Coloured people led to Christianization (Keegan, 1996), except among the small Cape Malay group that remained Muslim. After the emancipation of the slaves by the British administration in 1834-1838, many indigent people and slaves that had left their owners, settled at these mission stations due to the availability of land and the prospect of living as free people in an established environment (Mountain, 2004).

It is noteworthy that the mission stations set up among and for the Khoekhoen were subsequently found to be populated by people describing themselves as "Coloured" (Nurse *et al.*, 1985), which was regarded as a more politically expedient designation (Nurse *et al.*, 1985). Some mission stations formed the nucleus of a "coloured group area" according to legislation in 1950 (Boonzaaier *et al.,* 1996; Mountain, 2003). Many of the Khoesan had European or African (particularly Xhosa) ancestry (Keegan, 1996). The formalisation of the racial order in society began in the late 1700's and early 1800's. During the apartheid era from 1910 and particularly 1948-1994, the regulation of life according to perceived race reached its peak, with choice of marriage partners and place of residence determined by

law. Official segregation served to maintain the population make-up of the Coloured group.

People designated as SAC, living in other parts of South Africa may well have slightly different proportions of ancestral admixture, due to different population dynamics, but the Western Cape is still home to the majority of the SAC population (61.1% according to the 2001 census) and it is said that the Western Cape is the traditional centre of concentration of the Coloured people (Adhikari, 2005; Cilliers, 1963). The popular perception of this population group is that it is made up of black, white and Khoesan contributions. While a fair amount of genetic research has been done on the Bantu-speaking black ethnic groups of South Africa (Lombard *et al.,* 2006; Nurse *et al.*, 1985), the SAC group is under–researched.

We investigated a population that self-identifies as SAC, living predominantly in the suburbs of Ravensmead and Uitsig in metropolitan Cape Town. The population of Ravensmead/Uitsig is 91% Christian, and only 1.5% Muslim (2001 SA census), which raises an important distinction with another group in South Africa known as the Cape Malays. This group has its origins in the political exiles brought from the Dutch East Indies (now mainly Indonesia) in the 1700's (Mountain, 2004). They brought the religion of Islam to South Africa, which served as a unifying force in the community and may have created a genetic subgroup. This group was not investigated here. The name "Malay" is at present often used by members of the group to denote chiefly the affiliation with Islam, and the group is also known as the Cape Muslims (Mountain, 2004). It is a minority group which has not been incorporated into the core structure of the Coloured people (Nurse *et al.*, 1985).

One aim of our study was to assess whether population substructure exists within the admixed SAC population from Ravensmead/Uitsig. In order to do this, 209 unlinked randomly selected microsatellites were genotyped in 46 unrelated individuals with an unknown clinical status.

A further aim was to investigate population structure in the SAC population. Population structure can be described as the population being subdivided into different populations

and to identify the actual subpopulations (ancestral populations) that exist. Once the make-up of the study population is determined future studies such as admixture mapping can be performed. A large genome-wide analysis was done, consisting of 959 individuals from the SAC group, genotyped with a panel of 500,000 single nucleotide polymorphism (SNP) markers, of which nearly 75,000 markers are shared with the International HapMap consortium (Frazer *et al.,* 2007; The International HapMap Consortium, 2005) and Human Genome Diversity Project (HGDP) (Cann *et al.*, 2002). This is the first genome-wide investigation of this admixed South African population. Whilst understanding the demographic history of the SAC population is of interest in its own right, it is anticipated that genome level studies will facilitate the identification of genes associated with complex diseases in this population. Therefore, characterizing the pattern of genetic variation in this study population will provide valuable baseline data for subsequent analysis of disease association, and continent-wide genomic and population structure.

2.4 MATERIALS AND METHODS

2.4.1 Study population

All the subjects used in this study originated from the Ravensmead and Uitsig suburbs near Tygerberg Hospital, Parow, Western Cape, South Africa. Under the Group Areas Act of the apartheid government, the suburb of Ravensmead/Uitsig was declared an area for habitation by Coloureds only in 1962 and was rapidly settled. Although this act has not been in force since a fully democratic government was elected in 1994, 98% of people in this suburb self-identified themselves as “Coloured” in the 2001 South African census. Informed consent was obtained from all study participants. The study was approved by the Ethics Review Committee of the Faculty of Health Sciences, Stellenbosch University, Tygerberg, South Africa.

i) Population structure using microsatellite data

Forty-six unrelated individuals of unknown clinical status from the Ravensmead-Uitsig community were genotyped for 209 unlinked randomly selected autosomal microsatellites (data from Graham Cooke and Adrian Hill from the University of Oxford) to test for population substructure in the SAC population. These individuals were genotyped using a commercially available set of microsatellite markers (LMSv2 MD10; Applied Biosystems, Foster City, CA) used for a genomewide linkage analysis (Cooke *et al.,* 2008).

ii) Population structure using SNP data

An independent sample set consisting of 959 individuals (856 cases and 94 controls) from the SAC group residing in Ravensmead-Uitsig was genotyped with a panel of 500,000 markers using the Affymetrix 500K SNP chip, of which nearly 75,000 markers could be used for comparison with those in the HGDP database.

2.4.2 Population substructure

All the microsatellites were in HWE and analyzed using STRUCTURE (version 2.2). This program can be obtained from the website (http://pritch.bsd.uchicago.edu) (Pritchard, Stephens & Donnelly 2000). The Pr (X|K) obtained in this study as well as guidelines on how to use STRUCTURE were obtained from the user manual set up by Pritchard, Wen & Falush, April 2007. STRUCTURE is based on a Bayesian model clustering algorithm that identifies subgroups with distinctive allele frequencies (Pritchard & Donnelly, 2001). It makes use of unlinked randomly selected markers across the genome to infer population structure (Pritchard *et al.*, 2000).

When setting up STRUCTURE, preliminary runs were done using different models and parameters in order to determine the optimal parameter set to use. The admixture model was used as a default in all parameter sets as the population is known to be admixed (section 2.3.2). Both the allele frequency correlated and the allele frequency independent model was tested. The parameter set used in this analysis was the admixture model together with the allele frequency independent model. In the advanced parameter section we also included the print Q-hat option in order to see the estimated fraction of the individual's genome that is assigned to each *K* cluster. The burn-in period was 5×10^5 and the length burn in/reps were 5×10^6. These settings were used to run STRUCTURE for K= 1-6 with 10 replicates per *K*.

The Pr (X|K) were calculated for *K*=1-4. These values point out the probability that individuals will form part of a particular cluster based on their genotype data. Note that *K* should be treated with caution due to reasons covered in section 2.2.

2.4.3 Population structure

i) Sampling, genotyping and genotype calling

All samples, and CEU controls from the International HapMap Project (Frazer *et al.,* 2007; The International HapMap Consortium, 2005), were genotyped using the Affymetrix 500k SNP genotyping platform (http://www.affymetric.com) used in several studies. SNP genotypes were called using the Affymetrix Power Tools pipeline (V1.10.0). Firstly, samples that had a reported NSP/STY concordance rate of less than 90% were discarded. The Dynamic Model (DM) algorithm's call-rate was used as an initial quality control measure. CEL files with a call-rate of 93% or higher were selected, and used to train probe specific models using the BRLMM algorithm (http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf). These models were then saved, and used to call all samples with STY and NSP CEL files DM call rate of 70% or higher. All these SNPs were in HWE. In addition to data generated in this study we obtained genome wide SNP data from two additional public data repositories, The International HapMap Project (Frazer *et al.,* 2007; The International HapMap Consortium, 2005) (http://www.hapmap.org) and the Human Genome Diversity Project (Cann *et al.*, 2002) (HGDP; http://hagsc.org/hgdp/files.html). Populations were chosen from these public data sources to represent putative ancestral populations that may have contributed through admixture to the SAC population. The populations chosen were representative of four major groups, namely 1) European, 2) non-Khoesan African (including Bantu and Pygmy populations), 3) Khoesan and 4) Asian (Table 2.3). We reduced the SNPs genotyped in this study to a subset (n = 74,889) shared between SAC and the public data sources (Table 2.3).

Population structure analyses were performed in order to identify the major genetic contributions to this uniquely admixed population. We used STRUCTURE (Falush *et al.,* 2003; Pritchard *et al.*, 2000), which identifies population structure without prior assignment of individuals to populations. STRUCTURE has an upper limit on the number of SNPs that can be analysed, and assumes both Hardy-Weinberg and complete linkage equilibrium between adjacent markers (Falush *et al.,* 2003; Pritchard *et al.*, 2000). The selection of highly informative markers reduces the number of genotypes required for the accurate inference of ancestry. Therefore, we selected SNPs from the set of shared markers (*n* = 74,889) that were ancestry informative for the putative contributions to the SAC, and that

were putatively unlinked. We used Rosenberg's Ancestry Informative Markers (AIMs) selection method (Rosenberg *et al.*, 2003), yet taking potential linkage into account by selecting AIMs separated by a physical distance of at least 1MB. Alternate marker selection strategies including random selection, random selection accounting for linkage disequilibium and AIMs not accounting for linkage disequilibrium were also tested (Table 2.4).

Table 2.3 Putative ancestral populations included in population structure analysis of South African Coloureds (SAC).

Population	Description	n	Source
European			
CEU	Utah residents with Northern and Western European ancestry from the CEPH collection	109	HapMap 3
African: non Khoesan			
LWK	Luhya in Webuye, Kenya	83	HapMap 3
MKK	Maasai in Kinyawe, Kenya	143	HapMap 3
YRI	Yoruba in Ibadan, Nigeria	108	HapMap 3
BAN	Bantu from Africa	19	HGDP
BPG	Biaka Pygmies from Africa	21	HGDP
MAN	Mandenka from Africa	22	HGDP
MPG	Mbuti Pygmies from Africa	12	HGDP
African: Khoesan			
SAN	San from Africa	5	HGDP
Asian			
CHB	Han Chinese in Beijing, China	79	HapMap 3
GIH	Gujarati Indians in Houston, Texas	83	HapMap 3
JPT	Japanese in Tokyo, Japan	82	HapMap 3
LAH	Lahu from East Asia	8	HGDP
MEL	Melanesian from Oceania	10	HGDP
PAP	Papuan from Oceania	17	HGDP

Table 2.4 Alternate SNP marker selection strategies employed in the STRUCTURE analysis.

	Minimum inter-SNP distance	Number of SNPs
Random	Not applicable	2500
Random (unlinked)	1 Mb	2431
AIMs	Not applicable	2732
AIMs (unlinked)	1 Mb	1121

We first estimated the number of ancestral populations (K) in the combined SAC and public dataset listed in Table 2.3. We used an admixture model with correlated allele frequencies. Convergence of MCMC chains was assessed with five independent runs (burn-in = 1000, chain length = 2500) for each K between one and eight. The number of ancestral subpopulations (K) was estimated as that which maximized the probability of the data, and minimized the variance in this probability over successive iterations (Pritchard *et al.*, 2000). For each SNP subset we estimated the proportions of inferred ancestry for each individual at the optimal number of ancestral populations (K), and plotted these proportions using DISTRUCT (Rosenberg, 2004). A potential limitation in estimating proportions of ancestry for SAC in these analyses is that the admixture model used does not account for linkage disequilibrium due to admixture (Falush *et al.*, 2003), known to occur in our study population (Nurse *et al.*, 1985). An alternate linkage model accounts for linkage disequilibrium due to admixture and has more accurate estimates of statistical uncertainty in admixed populations, yet has runtimes that scale exponentially with the number of ancestral subpopulations (Falush *et al.*, 2003). Therefore, we estimated ancestral proportions using the linkage model in STRUCTURE only for the optimal number of ancestral subpopulations identified in the previous analyses. In this case, we used a larger sample of 10 000 SNPs since linkage due to admixture is incorporated into the model, but we still maintained a physical distance of at least 10 Kb between adjacent SNPs to limit the effect of background linkage disequilibrium. This linkage model has the added benefit of estimating the average rate of decay of admixture linkage disequilibrium, and the time since admixture occurred. Finally, we performed principal component analysis using SMARTPCA in the EIGENSOFT pakage (Patterson *et al.,* 2006; Price *et al.*, 2006) and included all SNP markers shared between the populations analyzed (n = 74889).

2.5 RESULTS

2.5.1 Population substructure

In this study the posterior probability (P[KlX])[a] for K=1 was 1 where K=2-4 was zero (Table 2.5). Therefore the conclusion that can be drawn from this result is indicative of an absence of population structure in the South African Coloured population used in this study. This can also be seen in the bar plots (Fig. 2.4).

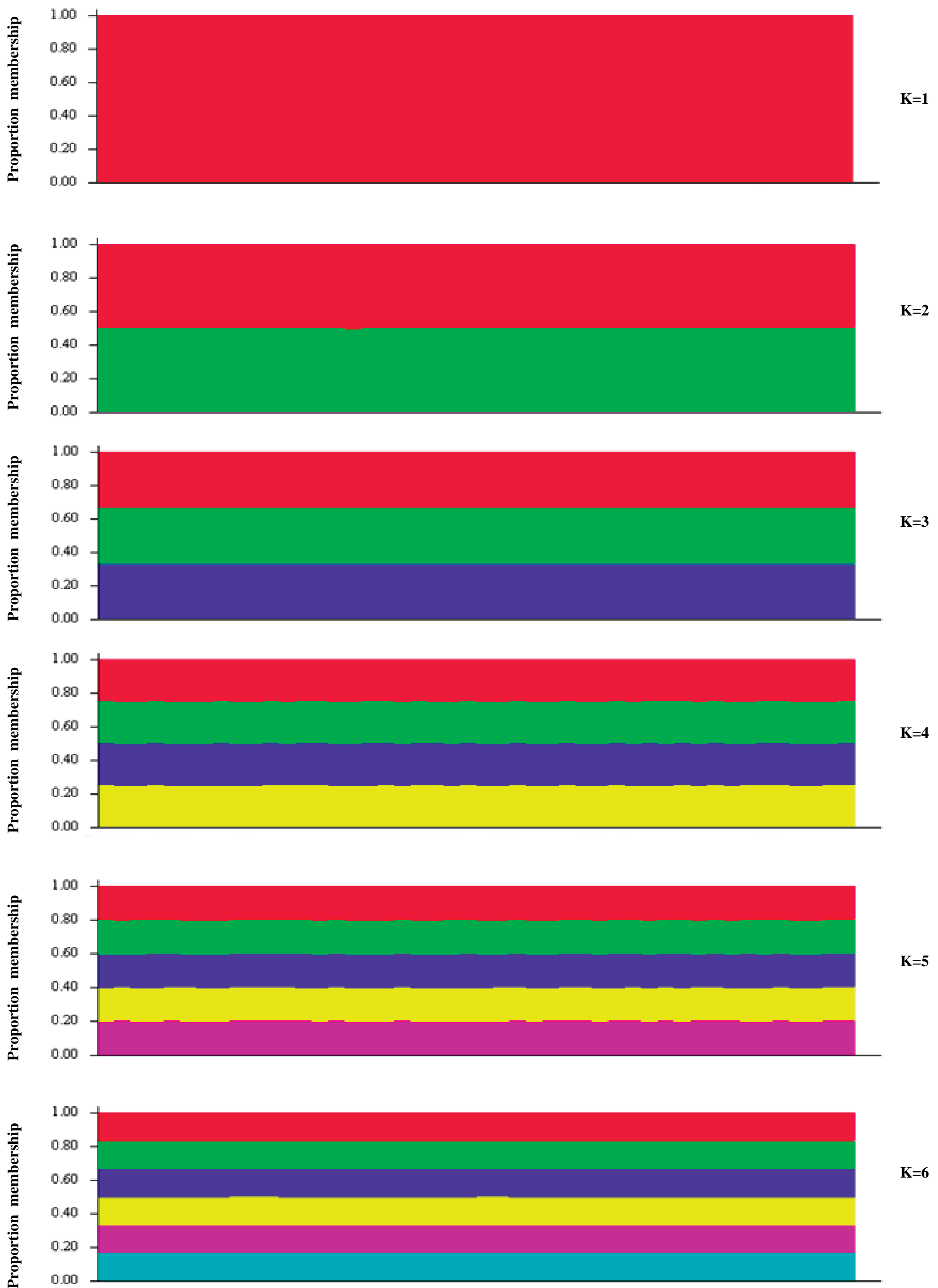


Fig. 2.4 Bar plots for estimated membership from structure for K = 1- 6 obtained with the microsatellite data. Each vertical line represents an individual on the X-axis while the Y-axis indicated the individual's estimated membership fraction in the different clusters (K=1-6).

Each vertical line represents an individual and the six different colours are representative of the six different clusters (*K*=1-6). These coloured segments represent the individual's estimated membership fraction in the different clusters. All the individuals are almost equally assigned to *K*=1–6. The estimated posterior probabilities for *K* are only shown for *K*=1-4 as the rest would also be zero.

Table 2.5 The estimated posterior probabilities for *K* in the South African Coloured samples.

K	Total Sample	
	In P (X|K)[a]	Posterior Probability P (K | X)
1	-34210.95	1.00
2	-34244.72	0.00
3	-34380.41	0.00
4	-34434.56	0.00

[a] The probability that an individual will occupy a specific cluster given the genotype data

2.5.2 Population structure

Genotype calling performance was determined by measuring concordance of the included HapMap cell line samples with the known genotypes of these individuals, and was found to be >99%. Furthermore, four duplicate SAC samples were included in the SNP genotyping experiment, which allowed for validation of the genotype-calling algorithm on SAC samples. Genotype concordance of these SAC samples was >97%.

In the STRUCTURE analyses, including both SAC data and potential ancestral populations derived from public data sources; the number of populations was estimated as between 4 and 7 (Fig. 2.5). Datasets not accounting for putative linkage between SNPs provided higher estimates of K. Although the estimate for the number of ancestral populations varied with each subset of SNPs, the inferred major contributions to the southern African coloured population were consistent (Figs. 2.6 - 2.9). The four major contributions to SAC include Europeans, non-Khoesan Africans, Khoesan Africans and an Asian contribution. Of these, the Khoesan contribution is substantially the largest, followed by European, African and Asian (Table 2.6). It is surprising that only a small sample of Khoesan individuals (*n* = 5) suggested such a substantial Khoesan contribution to the SAC population. However, it is likely that a large sample of coloureds allowed for this

distinction, given that several individuals within the South African Coloureds have high proportions of Khoesan ancestry (Fig. 2.10).

Estimates of the proportion of ancestry derived from the linkage model indicate the Khoesan contribution to SAC to be lower than in the admixture model (Table 2.6). These differences are likely the result of Khoesan having approximately 30% shared ancestry with non-Khoesan African populations in the linkage model analysis (Fig. 2.11). The European contribution is also substantially lower in the linkage model analysis (Table 2.6). The linkage model also enabled the estimation of *r* = 0.0001, the product of time (in generations) since admixture and the recombination rate (Falush *et al.,* 2003). Assuming recombination rates on the order of 10^{-5} per base pair per meiosis, the estimated time since admixture is 10 generations, broadly corresponding to 200 years (assuming generation time is 20 years). Principal component analyses indicated that SAC span variation between Africans and non-Africans along the first eigenvector (Fig. 2.12) and eigenvector three pulled out the different African populations like San and Bantu. The PCA results suggest that the ancestral Asian population that contributed to SAC is more closely related to the contemporary Gujarati Indian population, than to the Chinese (CHB) and Japanese (JPT) populations from HapMap 3 (Fig. 2.12), as also found recently (Tishkoff *et al.*, 2009).

Table 2.6 Mean and standard error on proportion of ancestry for each of four populations contributing to South African Coloureds (SAC), for admixture and linkage models. The unlinked AIMS were used for the admixture model. Mean and standard errors are from multiple STRUCTURE analysis (*n*=5) with the number of ancestral populations (K) equal to 4.

	Asian	**African#**	**Khoesan**	**European**
Admixture model	0.0892 ± 0.034	0.1966 ± 0.081	0.4334 ± 0.083	0.2808 ± 0.075
Linkage model	0.1082 ± 0.002	0.3593 ± 0.005	0.3230 ± 0.004	0.2095 ± 0.004

#non-Khoesan African

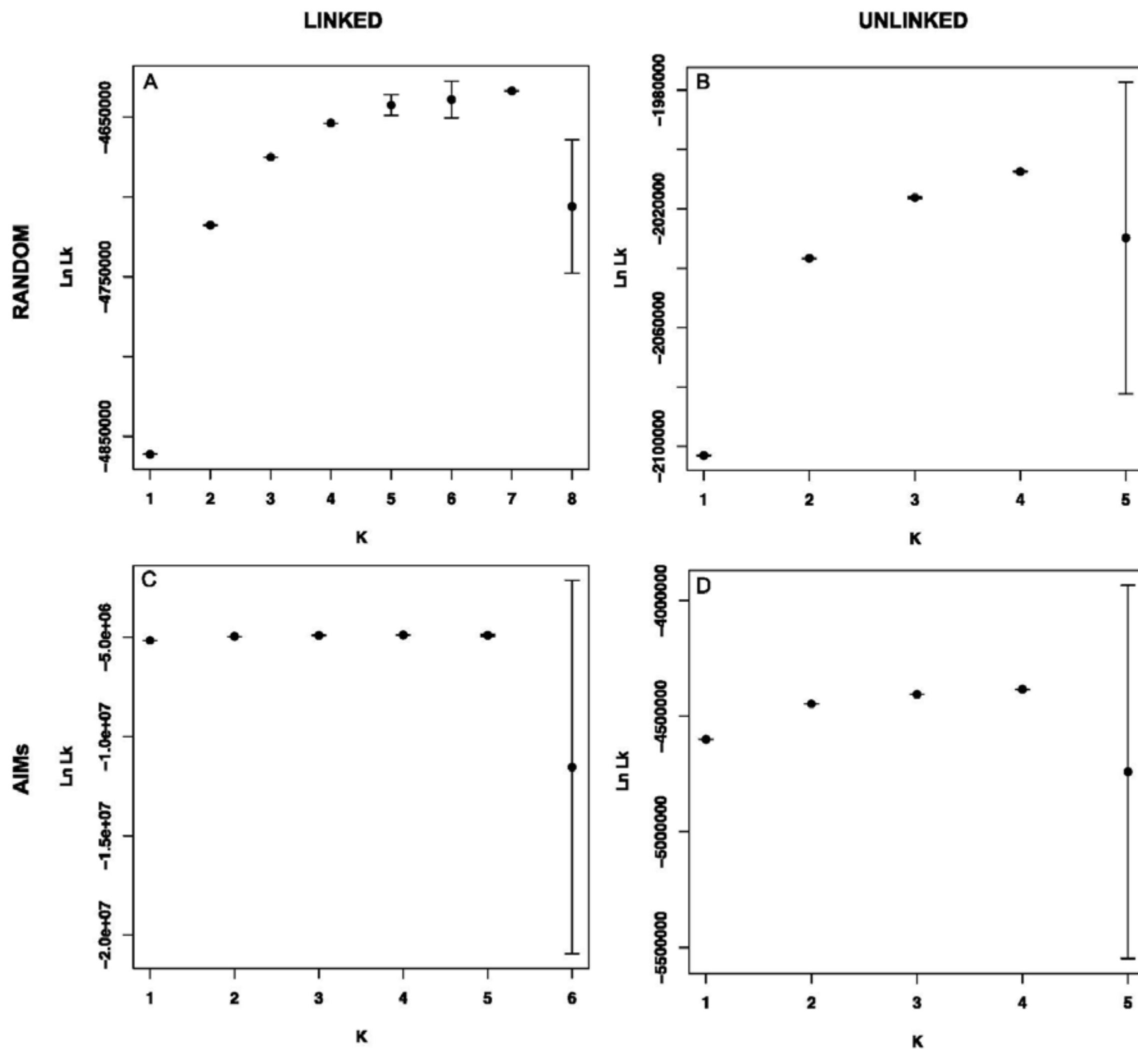


Fig. 2.5 Estimates of the number of ancestral populations (K) for the SAC and combined HapMap and HGDP samples under an admixture model using the STRUCTURE software. The estimated probability of the data given the model is plotted against increasing K for each of the subsets of SNP data used (see Methods).

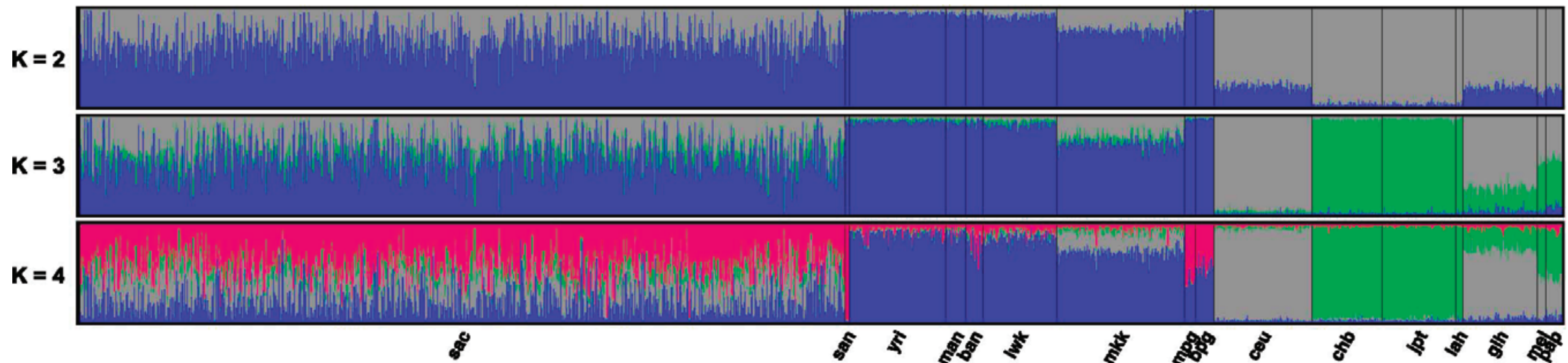


Fig. 2.6 Proportion of each individual's ancestry for the number of ancestral populations from k = 2 to the estimated number of ancestral populations with greatest probability (Fig. 2.5D). Plots shown are for unlinked Ancestry Informative Markers. Population labels are as in Table 2.1.

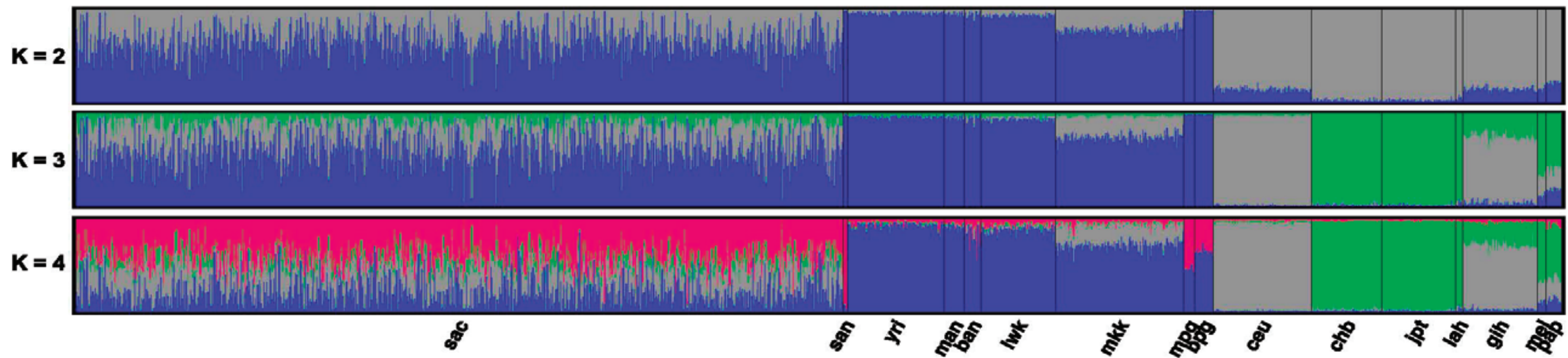


Fig. 2.7 Proportion of each individual's ancestry for the number of ancestral populations from k = 2 to the estimated number of ancestral populations with greatest probability (Fig. 2.5B). Plots shown are for unlinked random SNPs.

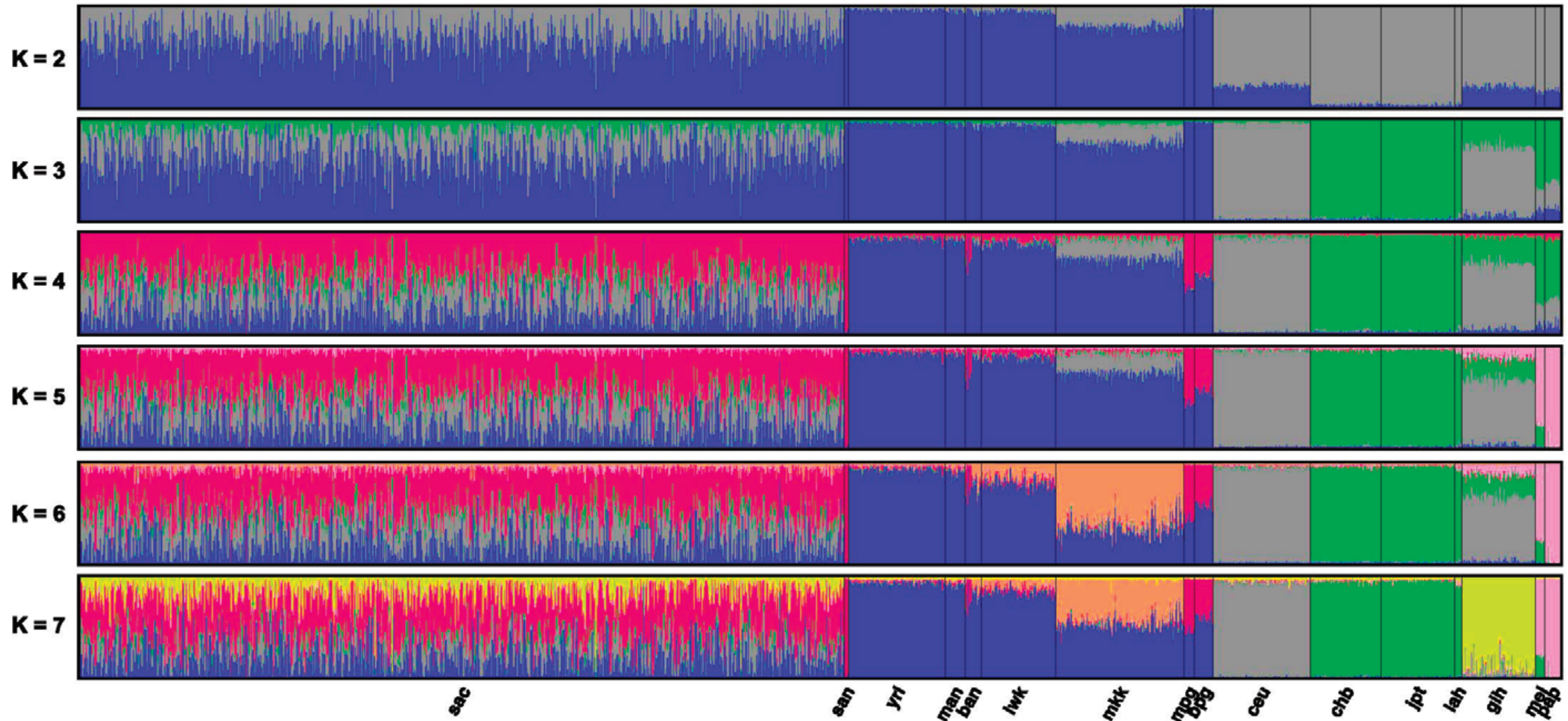


Fig. 2.8 Proportion of each individual's ancestry for the number of ancestral populations from k = 2 to the estimated number of ancestral populations with greatest probability (Fig. 2.5A). Plots shown are for random linked SNPs.

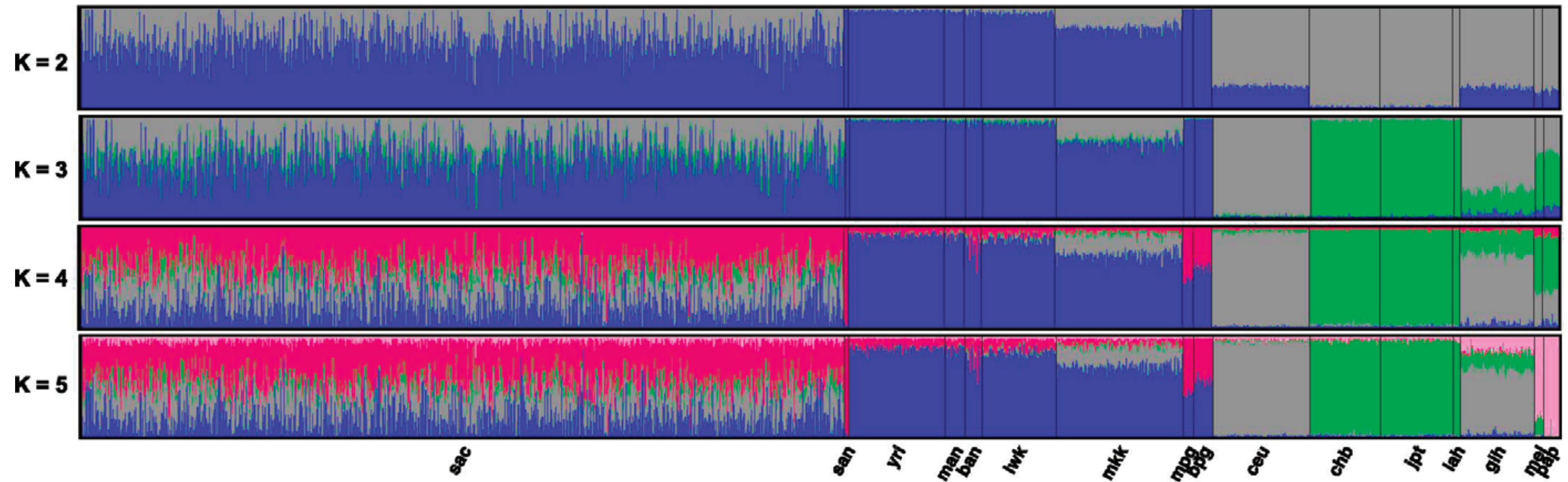


Fig. 2.9 Proportion of each individual's ancestry for the number of ancestral populations from K = 2 to the estimated number of ancestral populations with greatest probability (Fig. 2.5C). Plots shown are for linked Ancestry Informative Markers.

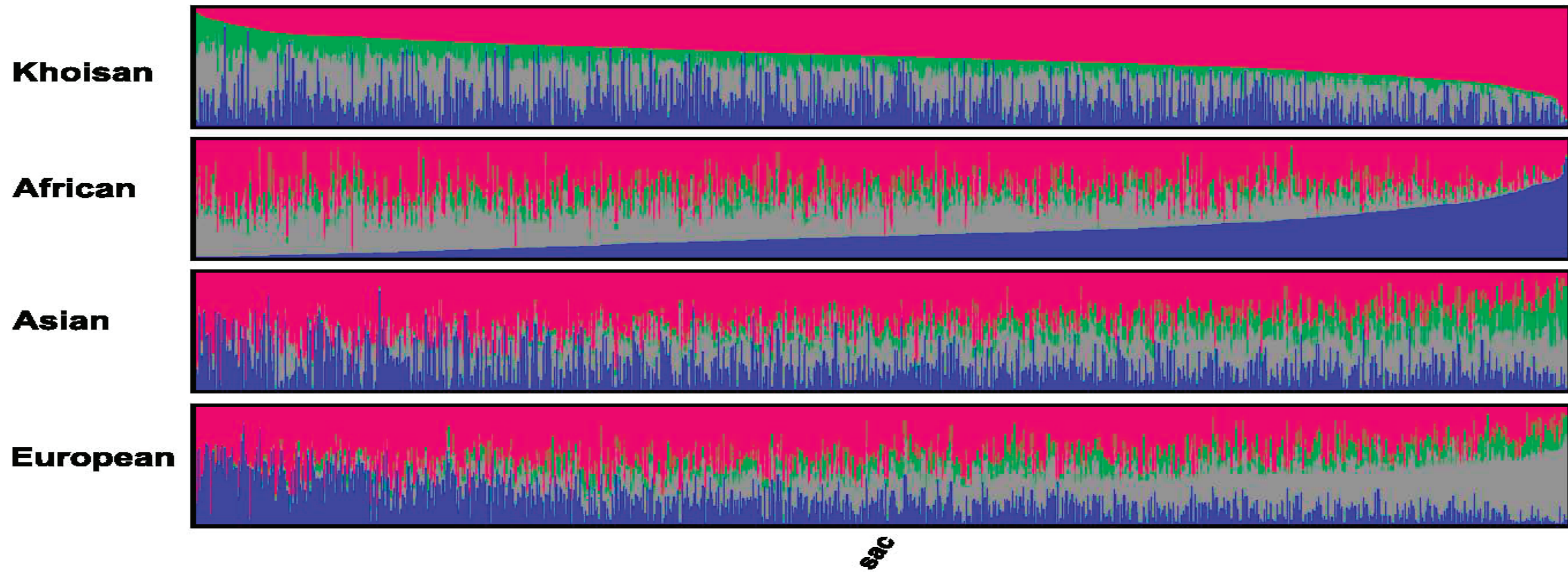


Fig. 2.10 Proportion of each individual's ancestry (K = 4) sorted (in ascending order from left to right) by the proportion of ancestry for each of the major contributions to the SAC. Khoisan red, African blue, Asian green and European grey.

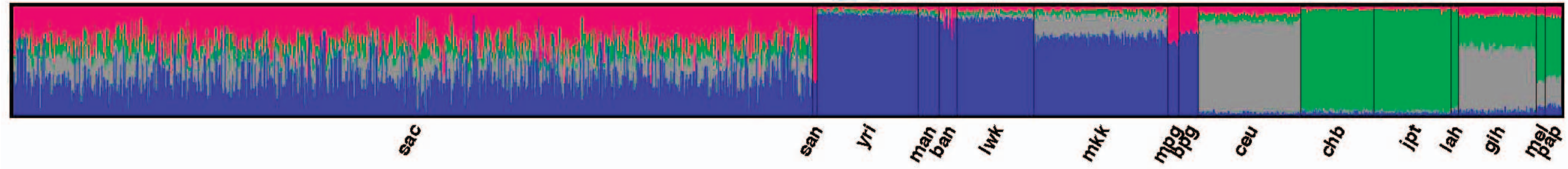


Fig. 2.11 Proportion of each individual's ancestry derived using the linkage model in STRUCTURE for the optimal number of ancestral populations (K = 4).

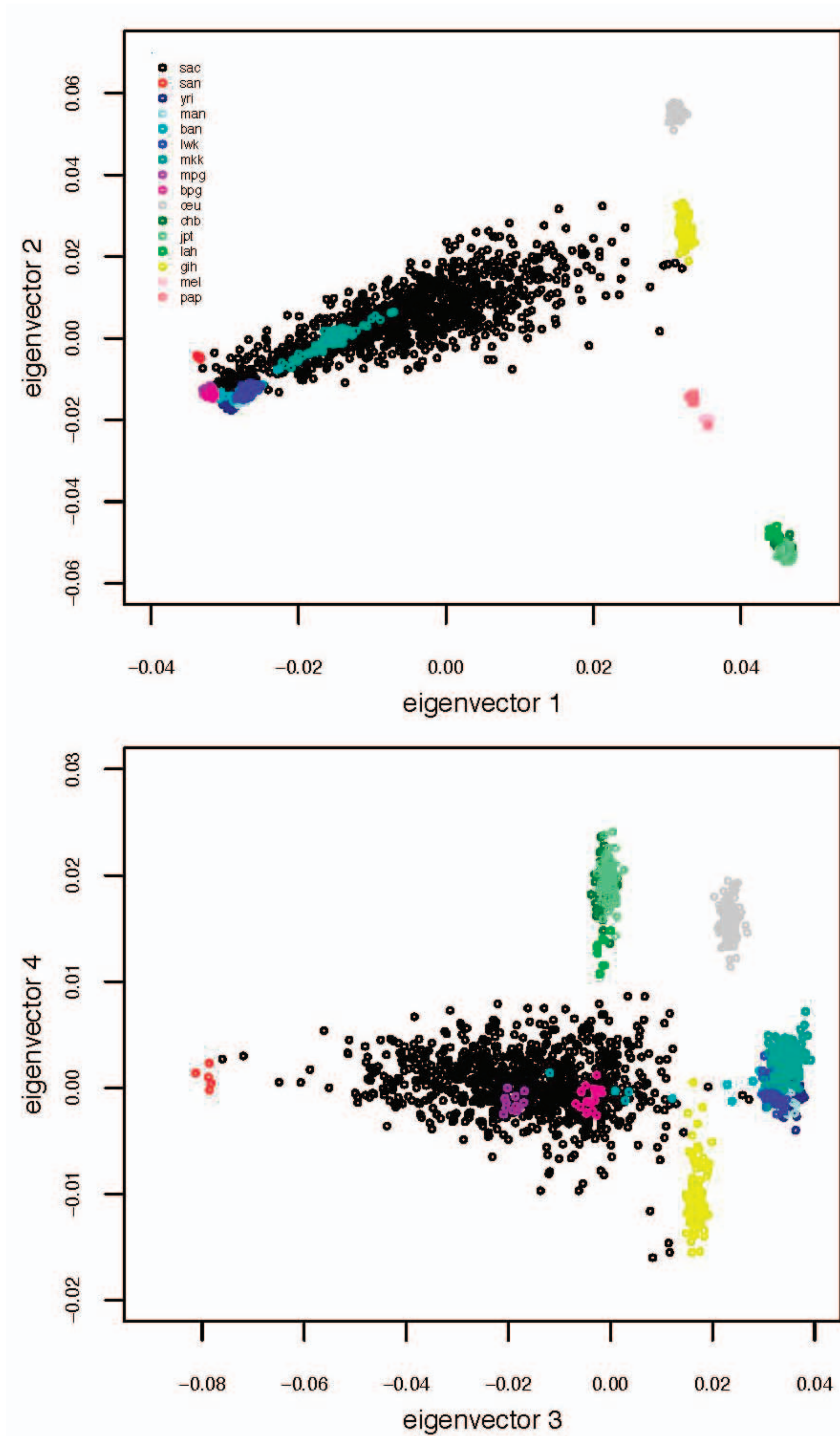


Fig. 2.12 Plot of the first four eigen vectors in the PCA analysis of SAC, HapMap 3 and HGDP populations selected as putative ancestral populations for the South African Coloured population.

2.6 DISCUSSION

i) Population substructure

The historical roots of the SAC populations are complex and diverse with contributions from various groups. The genetic depth of this population is estimated at approximately 10-15 generations which ties in with historical evidence that the population was founded between 150 and 350 years ago. Since this population dates back more than 2 centuries and was kept separate by powerful social forces, one would expect that it should be relatively homogenous. Our results finding no substructure in the population were congruent with those of Barreiro *et al.,* (2006), who found no population stratification

ii) Population structure

Our results, and those recently reported (Tishkoff *et al.*, 2009), bear testament to the statement that the SAC population has the highest degree of intercontinental admixture of any global population (Tishkoff *et al.*, 2009). We have genotyped 959 individuals from the SAC population, and selected 75000 markers for population structure analyses. We found that their genetic background agrees with the historical record and furthermore, reveals detailed insights into the events that resulted in the present-day SAC population. We show that the SAC population group has 4 major ancestral components: the Khoesan, European, African and Asian (Indian), the proportions of which are dependent on the statistical model used in inference (Table 2.6). Differences between the admixture and linkage model are to be expected since each accounts for different components of linkage disequilibrium. The admixture model ignores linkage disequilibrium along chromosomes as a result of admixture, whereas the linkage model does not (Falush *et al.*, 2003), and thus the latter is a better approximation of the population history of SAC. Nonetheless, the inferred ancestry proportions indicate a substantial contribution from the Khoesan, and considerable variation in ancestry proportions between individuals (Fig. 2.6, 2.10). The degree of Khoesan ancestry reflects the role of indigenous Khoesan in the early establishment of the SAC population (Mountain, 2003). It could be argued that the rather small Khoesan sample size contributes to uncertainty with respect to estimating ancestral proportions. However, these results are consistent with an independent study with a slightly larger population of Khoesan (Tishkoff *et al.*, 2009). Some authors have proposed that the Khoesan people in South Africa are becoming extinct (Mountain, 2003). The San in particular endured bouts of genocide from all other groups (Mountain, 2003; Shell, 1994),

and the Khoekhoe society had collapsed completely by 1713 (Elphick, 1985), the time of a devastating smallpox epidemic (Nurse *et al.*, 1985). Although many members of the Khoesan existed on the fringes of colonial society (Mountain, 2003), many others, particularly the women, were part of the household of the pioneer farmers, in the patriarchal or paternalistic societal system that had elements of slavery, indentured labour and authoritarian family life (Shell, 1994). Often Khoekhoe men were bonded labourers on the farms (Keegan, 1996; Shell, 1994) and integrated into western society (Elphick, 1985). The putative near extinction of the Khoisan, however, is not apparent from our results; given that some SAC individuals harbour large proportions of Khoesan ancestry (Fig. 2.6, 2.10), and assuming the HGDP Khoesan population is a sufficiently pure source of ancestral Khoesan diversity.

In addition to the strong Khoesan contribution to SAC, a large proportion of their ancestry is derived from non-Khoesan Africans (Fig. 2.6), in particular Bantu-speaking populations. Although we did not have samples for southern African Bantu, these groups are themselves admixed with the Khoesan (Nurse *et al.,* 1985; Thorp, 2000; Tishkoff *et al.*, 2009), which is evidenced by click consonants in the Xhosa language (Tishkoff *et al.*, 2009). Furthermore, substantial input from the European settlers (Dutch, German, British and French), and a smaller contribution from Asia is evident in the SAC (Fig. 2.6). This Asian contribution may be consistent with 26% of imported slaves originating from East India, mainly Bengal (Shell, 1994), and the apparent shared ancestry between East Asians and Gujarati Indians (Fig. 2.6). The use of HGDP Gujarati Indians as a proxy for the Indian populations that were the actual ancestral populations in the SAC is supported by the genetic homogeneity of Gujarati and Bengali populations (Tishkoff *et al.*, 2009). One analysis we performed, which included a random subset of SNPs, did detect the Gujarati Indian contribution to SAC (Fig. 2.8), however, we suspect this result may be influenced by the larger proportion of linked SNPs in that analysis, and the inability of STRUCTURE to account for background linkage disequilibrium. Nonetheless, the Indian contribution to SAC is supported by PCA analysis of all 75000 markers (Fig. 2.12). Low levels of ancestry from East Asia (CHB/JPT in HapMap) may be ascribed partly to the Chinese who formed part of the free blacks (Keegan, 1996), a group forming 9% of the Cape Town population by 1821 (Shell, 1994). "Free blacks" were free persons not of European origin, and comprised manumitted slaves, a few political exiles, and several hundred Chinese convicts

(Mountain, 2004; Shell, 1994). Interestingly, apart from the predominant European component, Chinese, Indian and Cape-born slaves have also been found to contribute to the Afrikaner population (Greeff, 2007; Heese, 1971).

The 959 individuals investigated in our study show a greater proportion of Khoesan ancestry and lower proportion of both European and Indian ancestry than the 39 individuals genotyped by Tishkoff *et al.,* 2009, who showed approximately equal ancestries of Khoesan, European, black African and Indian (19-25% each). It is possible that this group of 39 contained a reasonable proportion of people from the Cape Malay group, who could be expected to have a genetic makeup higher in Indian and lower in Khoesan ancestry, due to greater Indonesian or Malaysian ancestry. The samples used by Tishkoff *et al.,* 2009 were collected from blood donors in the Western Cape (MJ Kotze, personal communication), and not from a specific area. An early study of blood group gene frequencies in Cape Town found similar ancestral contributions from European, black and Asian, but the criteria for inclusion were not clear (Botha, 1972).

The resolution of inferred ancestral contributions could certainly be improved with the addition of both more suitable ancestral population samples from Malaysia and Indonesia, and with a larger sample of San. It is encouraging that the genotype results concur with the historical record, but in addition provide quantitative information of the extent of the contribution of putative ancestral groups, not obtainable by conventional historical research.

In addition to the results presented here being of historical anthropological interest, the inferred ancestral contributions are highly relevant for mapping of disease genes. The SAC population in the Western Cape suffers from one of the highest incident rates of tuberculosis (TB) ever recorded (Kritzinger *et al.*, 2009), and thus knowledge of their population structure, and ancestry, is critical to disease-gene mapping. One approach for the genome-wide identification of genetic risk factors is admixture mapping (McKeigue, 1997; Montana & Pritchard, 2004; Seldin, 2007; Zhu *et al.*, 2006; Zhu *et al.*, 2008). The SAC are an excellent subject for mapping of disease genes in complex admixed populations since their ancestral populations have substantially different rates of TB infection and disease (Stead *et al.*, 1990). An essential requirement for admixture mapping is the

elucidation of ancestral proportions of the populations involved. Thus the results reported here will enable the investigation of the impact of admixture on TB susceptibility, and potentially explain the apparent high vulnerability of this population to disease. Furthermore, given the high number of private alleles found in this population (Tishkoff *et al.*, 2009), novel TB susceptibility alleles could be identified.

2.7 ACKNOWLEDGEMENTS

I am grateful to Dr. Eugenia D'Amato (UWC) for her advice and assistance with the programme STRUCTURE. I would like to thank Dr. Marlo Möller, Dr. Wayne Delport, Emile Chimusa Rugamika and Ayton Meintjes (UCT) for analysing the 500k SNP chip data. Thank you to Prof. Cathal Seoighe (NUI Galway) for comments, suggestions and guidance with the population structure analysis. I would also like to thank Dr Adrian Hill and Prof. Graham Cooke (University of Oxford) for providing the microsatellite data.

CHAPTER 3

ASSOCIATION STUDIES IN EIGHT CANDIDATE GENES

3.1 INTRODUCTION

Eight genes, previously indicated to play an important role in TB susceptibility, were chosen as candidate genes for this study and will be discussed in this chapter.

The polymorphisms in these genes include a three SNP haplotype in mannose-binding lectin (*MBL)* at codons 52 (Arg-Cys), 54 (Gly-Asp) and 57 (Gly-Glu) in exon 1, natural resistance-associated macrophage protein (*NRAMP1)* $(GT)_n$ repeat*, NRAMP1* (TGTG) deletion, interleukin-1 receptor antagonist (*IL-1Ra)* (86 bp VNTR), the C-C chemokine regulated upon activation, normal T-cell expressed and secreted (*RANTES)* also known as *CCL5* that is a ligand for chemokine receptor 5 (-403A/G), interferon-gamma (*IFN-γ)* (+874 T/A), interleukin 8 (*IL-8)* (+100 C/T), surfactant protein D (*SP-D)* (Met11Thr) and interferon-gamma receptor 1 (*IFNGR1)* (1050 T/G).

Samples were collected from the suburb of Ravensmead-Uitsig in Cape Town, Western Cape, South Africa where the incidence of TB has been reported as up to ± 1000 per 100 000 population together with a low HIV prevalence (Munch *et al.*, 2003). The study population, known as the South African Coloureds (SAC), has a unique mixed ancestry dating back over 300 years, and consist of different ancestral population groups such as the European Caucasoid, Asian, Khoi, San and Black Africans (Nurse *et al.* 1985). Since the admixture occurred several generations ago the population should be relative homogenous. This statement has previously been investigated in the SAC population by using 25 unlinked SNP markers (Barreiro *et al.*, 2006) and by a more in-depth investigation in this study where no significant substructure was observed (see chapter 2).

Blood samples were collected with ethics approval from the Ethics Committee of the Faculty of Health Sciences, Stellenbosch University, South Africa whereafter DNA was extracted using standard methods. This was a case-control study design in which all the cases (*n*=505) had culture proven TB and all the controls (*n*=318) had never been diagnosed with TB before although they were presumably exposed to the mycobacterium as they were in close contact with other TB patients in the same community. The sample sizes for these association studies were determined by power calculations. Tuberculous meningitis (TBM) cases (*n*=50) were included as an extreme phenotype and also as an

association has previously been found in the SAC population (Hoal-van Helden *et al.*, 1999). All the participants were unrelated, HIV negative and the cases formed part of the bigger cohort discussed in Chapter 2. Some but not all of the controls were also part of the samples discussed in Chapter 2.

3.1.1 Surfactant protein D (SP-D)

The gene encoding the collectin surfactant protein D (SP-D) is located on chromosome 10q22.2-23.1 and consists of seven exons (Fig. 3.1). SP-D participates in the pulmonary innate immune response by opsonization or lysis of inhaled pulmonary pathogens and interaction with the macrophage cell surface molecules to modulate phagocytosis of *Mycobacterim tuberculosis* (Ferguson *et al.,* 2002; Gardai *et al.*, 2003; Wu *et al.*, 2003).

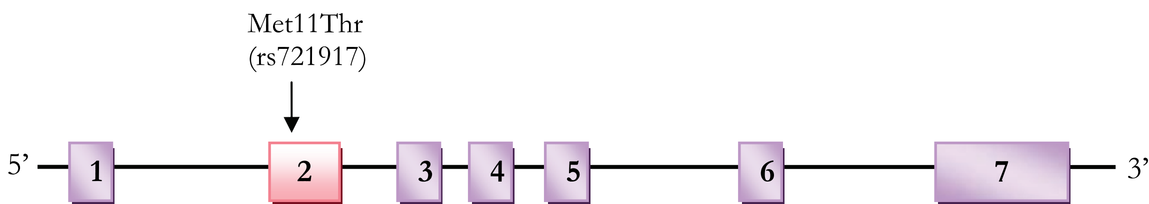


Fig. 3.1 *SP-D* and the polymorphism investigated in this study.

SP-D consists of 12 polypeptide chains of 43 kDa. The single peptide is composed of an NH_2-terminal, a collagen region, a coiled-coil neck region and a C-type lectin carbohydrate recognition domain (CRD) (Crouch & Wright, 2001; Leth-Larsen *et al.*, 2005). SP-D binds to *M. tuberculosis* in a manner consistent with an interaction between the CRD of the protein and the surface of the bacterium (Ferguson *et al.*, 2002). The main function of SP-D is thought to be the agglutination of bacteria as it binds to *M. tuberculosis*, but it is unknown whether bacterial agglutination promotes uptake into the host cell or whether the agglutinated *M. tuberculosis* are cleared. Therefore SP-D may be either a protective or risk factor for *M. tuberculosis* infection (Ferguson *et al.*, 1999). SP-D has been used as a biomarker for pulmonary disease states as it is present on the extrapulmonary epithelial surfaces and in serum (Sørensen *et al.*, 2006).

Polymorphisms in *SP-D* have previously been associated with several diseases (Diangelo *et al.*, 1999). Recent studies in children indicated that serum SP-D concentrations are genetically determined and that the (Met11Thr) SNP, localized in the amino-terminal region of the protein, is significantly associated with serum SP-D levels (Sørensen *et al.*,

2006). The Met11Thr SNP (dbSNP rs721917) is one of the most important SNPs in *SP-D* as it alters the codon corresponding to amino acid 11 where a Methionine (Met) is exchanged for a Threonine (Thr). This SNP has been implicated in TB in a Mexican population where individuals with the Thr variant are more susceptible to TB (Floros *et al.*, 2000). This study aimed to determine whether a correlation exists between this polymorphism (Met11Thr) and susceptibility to TB in the South African Coloured population.

3.1.2 Mannose binding lectin (MBL)

Mannose binding lectin (MBL), also known in the past as mannan-binding protein (MBP) is a C-type lectin in the lectin pathway which plays a major role in the pulmonary innate immune system (Ji *et al.*, 2005).

MBL is a multimeric molecule made up of six subunits each consisting of three identical polypeptide chains that contain a collagen-like region, hydrophobic neck region and a carbohydrate-recognition domain (CRD) (Fig. 3.2) (Presanis *et al.*, 2003). Microbial carbohydrate structures are recognized by pathogen associated molecular pattern (PAMP) receptors such as MBL (Buzás *et al.*, 2006). MBL plays an important role in the modulation of inflammation as it binds to the pathogen to initiate the lectin pathway which acts as the first line of defense against microbes (Holmskov *et al.,* 2003; Turner *et al.*, 2003).

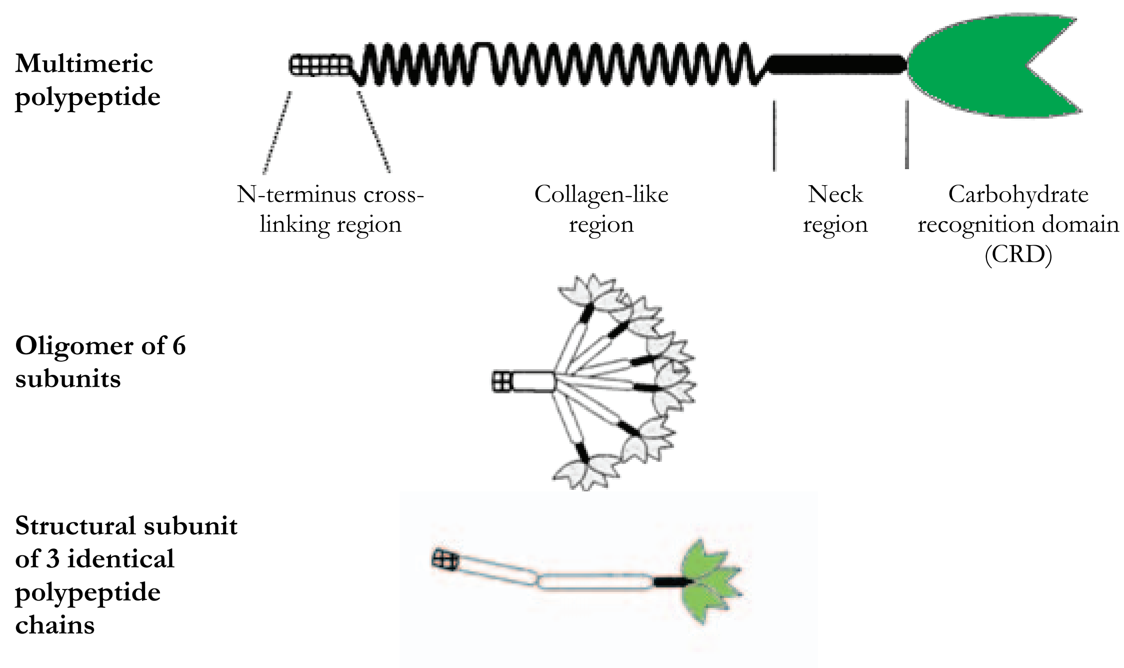


Fig. 3.2 The structure of MBL. Adapted from (Turner & Hamvas, 2000)

The exact mechanisms whereby MBL modulates inflammation are not well-known but it has been proposed that MBL exerts a complex effect on cytokine release by monocytes (Jack *et al.*, 2001). Phagocytes represent the first cellular defense in the alveoli of which the surface is rich in C-type lectin pattern recognition receptors.

The gene encoding the collectin mannose-binding lectin (*MBL*) maps to chromosome 10q11.2-q21 and consists of 4 exons (Fig. 3.3). Four allelic forms resulting in three SNPs within the collagen-like region of *MBL* exon 1 have been described. The A allele = wildtype, B allele = SNP within codon 54 (Gly → Asp), C allele = SNP within codon 57 (Gly → Glu) and D allele = SNP in codon 52 (Arg → Cys) (Lipscombe *et al.,* 1992; Madsen *et al.*, 1994; Sumiya *et al.*, 1991). These polymorphisms disrupt the assembly of the MBL peptide trimers or accelerate MBL degradation, resulting in profoundly reduced serum levels of functional polymeric MBL (Mullighan *et al.*, 2000). It has become evident that MBL plays a multifaceted role in many diseases and a deficiency of the protein has been associated with increased susceptibility to some infectious diseases (Summerfield *et al.*, 1997). Equivocal results have been found in the case of TB (Bellamy *et al.,* 1998; Selvaraj *et al.*, 1999), where promoting the uptake of bacteria into macrophages may be advantageous to the bacterium, and the variant alleles have been associated with protection against TB in Denmark (Soborg *et al.*, 2003), Turkey (Cosar *et al.*, 2008) and particularly tuberculous meningitis (TBM) in South Africa (Hoal-van Helden *et al.*, 1999).

.

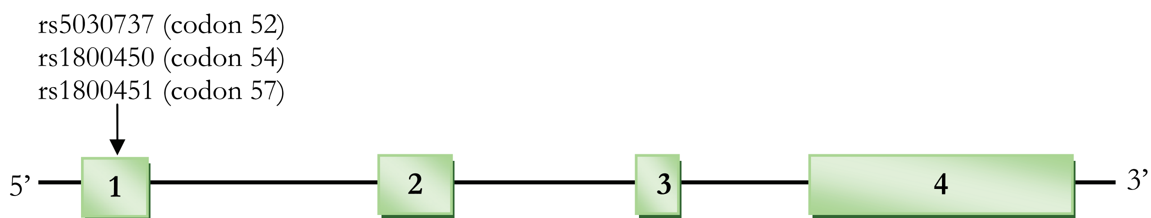


Fig. 3.3 *MBL* and the polymorphisms investigated in this study.

3.1.3 Interferon-gamma (IFN-γ) and Interferon-gamma receptor 1 (IFNGR1)

The interferon-gamma (IFN-γ) pathway is one of the most researched in TB since it is a flagship Th1 cytokine and plays a vital role in the protective immune response against *M.*

tuberculosis infection (Vidyarani *et al.*, 2006). IFN-γ is mainly secreted by CD4+ T cells, CD8+ T cells, natural killer (NK) cells and macrophages. The functional IFN-γ and IFNGR form a complex essential for containing the growth of *M. tuberculosis* (Dorman *et al.*, 2004). A previous study showed that defects in either of these two genes influence the production of IFN-γ and signaling and therefore individuals were more prone to mycobacterial infections (Casanova & Abel, 2002; Ottenhoff *et al.*, 1998).

Several polymorphisms have been indentified in *IFN-γ* and in the α and β chains of the IFN-γ receptor (*IFNGR*) gene, that were mapped to chromosome 12 and 6 respectively (Papanicolaou *et al.,* 1997; Zimonjic *et al.*, 1995). The structure of these two genes (*IFN-γ* and *IFNGR1*) regarding exons and introns can be seen in Fig. 3.4.

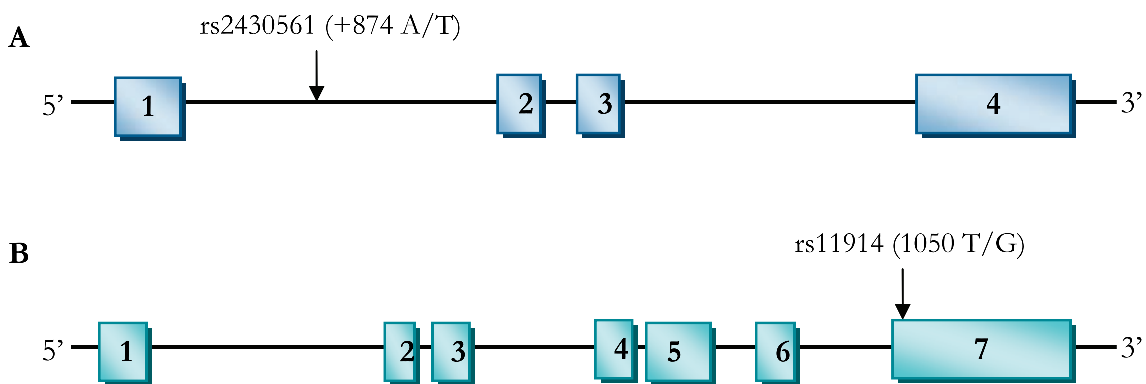


Fig. 3.4 [A] *IFN-γ* and [B] *IFNGR1* and the polymorphisms investigated in this study.

One of the most studied polymorphisms in *IFNG* is located in the first intron (+874 T/A) and has been associated with TB susceptibility in several populations such as Spanish (Lopez-Maderuelo *et al.*, 2003), Chinese (Tso *et al.*, 2005) and South African, where it was found in both a case-control and TDT study (Rossouw *et al.*, 2003). A meta-analysis performed on this particular SNP indicated a significant protection conferred by the T allele, to TB in 12 different population groups (Pacheco *et al.*, 2008). The +874 T allele provides a binding site for nuclear factor-κB, the transcription factor that induces IFN-γ expression, and the +874 AA genotype was predictive of a lower likelihood of sputum conversion in Japanese patients (Shibasaki *et al.*, 2009).

Most of the polymorphisms identified in the *IFNGR1* gene in patients with Mendelian susceptibility to mycobacterial infections, result in either a complete absence of protein expression or the expression of an abnormal protein that does not bind IFN-y. If IFN-γ signaling is completely abolished, the effects for the patient will be detrimental (Awomoyi *et al.*, 2004). It is therefore clear that the receptor complex plays a crucial role in the IFN-γ pathway. The SNP investigated in this study, previously shown to be associated with TB susceptibility in the South African Coloured population, is located in exon 7 (rs11914 1050 T/G).

This study aimed to determine whether a correlation existed between the *IFN-γ* rs2430561 (+874 T/A) polymorphism, the *IFNGR1* rs11914 (1050 T/G) polymorphism and susceptibility to TB in the South African Coloured population.

3.1.4 Interleukin 8 (IL-8)

An effective response by the host during *M. tuberculosis* infection in the lung is the recruitment and activation of inflammatory cells to the lung through chemokine activity. Interleukin 8 (IL-8) is a tissue-derived peptide and one of the major 13 human CXC chemokines that is secreted by several cell types, such as leukocytes and airway epithelial cells, in response to inflammatory stimuli such as *M. tuberculosis* (Ma *et al.*, 2003). It functions as a neutrophil chemoattractant for the recruitment of leukocytes to inflammatory sites.

In tuberculosis patients, increased levels of IL-8 have been reported in the bronchoalveolar lavage fluid (BALF), pleural exudates or cerebrospinal fluid compared to the controls (Abula *et al.,* 2008; Dlugovitzky *et al.*, 1997; Mastroianni *et al.*, 1994; Sadek *et al.*, 1998). In addition, other studies also showed increased IL-8 plasma concentration levels in patients who died from TB compared to those who survived (Friedland *et al.,* 1995; Pace *et al.*, 1999) and in vivo, as the anti-IL-8 antibodies were found to inhibit granuloma formation in rabbits (Larsen *et al.*, 1995).

Due to the major role that IL-8 plays in the immune system as a chemoattractant, variants in *IL-8* may predispose individuals to infection (Hull *et al.*, 2000). *IL-8* consists of 4 exons and is located at chromosome position 4q12-q21 (Modi *et al.*, 1990). Several genetic

polymorphisms (90) have been identified in this gene (Fig. 3.5). An important *IL-8* promoter variant, -251 A/T (rs4073), has been associated with susceptibility to TB in racially mixed American populations (Ma *et al.*, 2003). However this association was not replicated in either the Gambian population (Cooke *et al.*, 2004) or our South African Coloured population (unpublished data). In a pilot study previously conducted in the SAC population a more unfamiliar SNP, +100 C/T (rs2227538), located 1 base pair upstream from the ATG start codon showed a trend of significant association with TB (unpublished data). Therefore this study focused on this particular SNP.

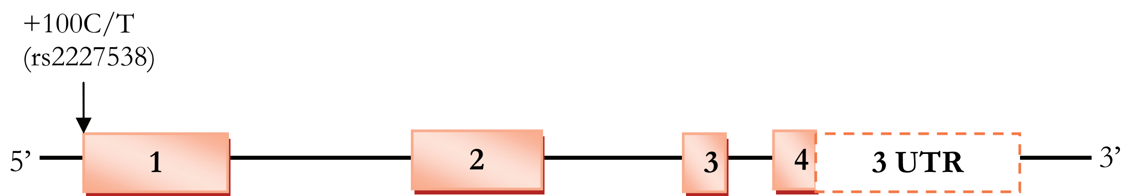


Fig. 3.5 *IL-8* and the polymorphism (+100 C/T) investigated in this study.

3.1.5 Regulated upon Activation, Normal T-cell Expressed and Secreted chemokine (RANTES), also known as CCL5

The gene encoding the chemokine Regulated upon Activation, Normal T-cell Expressed and Secreted (RANTES), also known as CCL5 and the ligand CCR5, consists of three exons and is located on chromosome 17q11.2 –q12 (Fig. 3.6) in the chemokine cluster region that has been shown to be important in susceptibility to TB (Jamieson *et al.*, 2004).

Fig. 3.6 *RANTES* and the polymorphism investigated in this study.

RANTES is a potent chemokine that forms part of a group of chemoattractant molecules that functions in the development of protective TB granulomas (Roach *et al.*, 2002). It recruits several types of inflammatory cells including monocytes and T lymphocytes, induces a Th1 response, is expressed in T-cells, lymphocytes and macrophages and is

induced after *M. tuberculosis* infection (Simeoni *et al.*, 2004). RANTES plays a major role in the three-step model of inflammation as described by Butcher and Springer (Fig. 3.7) (Butcher & Picker, 1996; Springer, 1995).

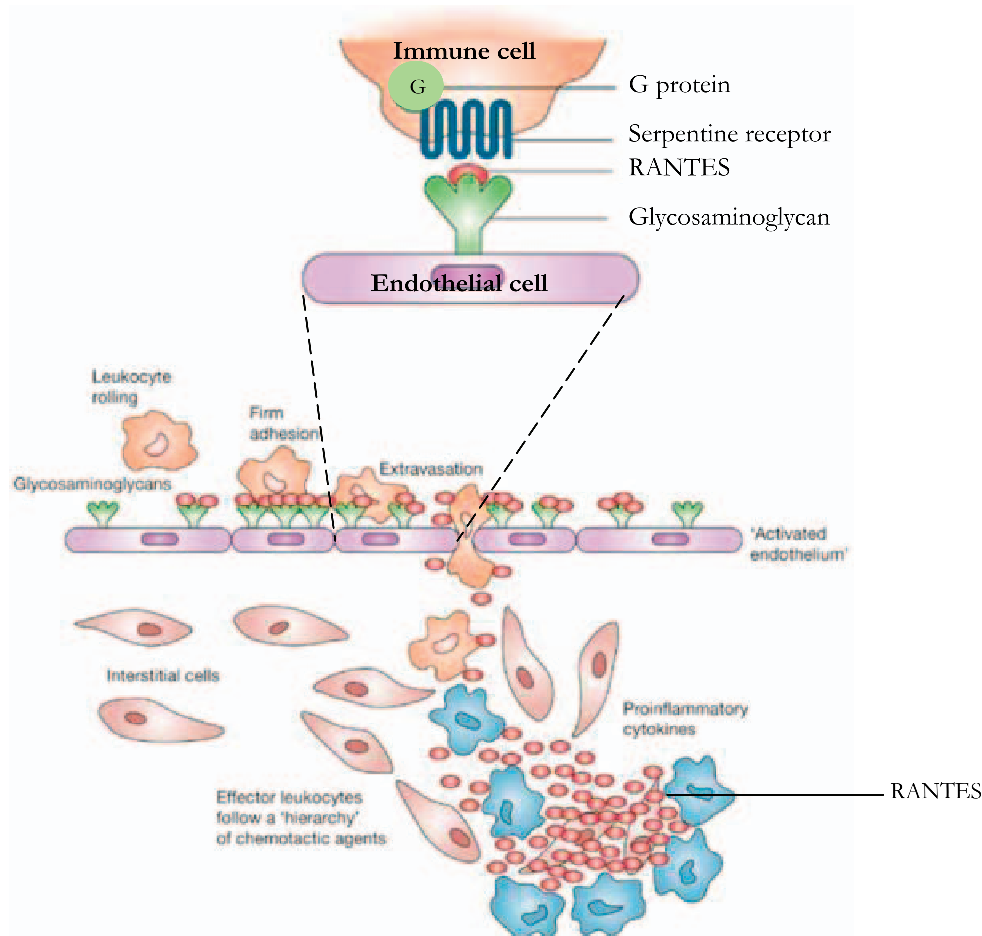


Fig. 3.7 RANTES attracts inflammatory cells from the bloodstream to the site of infection. Leukocytes roll along the endothelium via interactions with selectins. Upon inflammation, macrophages in the bloodstream release interleukin-1 (IL-1) and tumour necrosis factor (TNF) which in turn promotes the release of chemokines (including RANTES). The various cell types follow the chemotactic gradient to the infection site. In addition RANTES binds to glycoaminoglycans (GAG) on the endothelial cell surface. Once the T lymphocytes reach the injury site they produce large amount of RANTES. Adapted from (Krensky & Ahn, 2007).

In tuberculosis patients, an increased level of RANTES has been found in bronchoalveolar lavage fluid and in lung alveoli during active pulmonary tuberculosis (Sadek *et al.*, 1998). The promoter region of the gene has been extensively screened for genetic variants because of its suspected importance in protein expression (Azzawi *et al.,* 2001; Gonzalez *et al.*, 2001; Liu *et al.*, 1999; Nickel *et al.*, 2000). The three most frequent SNPs that were found in this

region include -403 A/G (rs2107538), -109 T/C (rs1800825) and -28 C/G (rs2280788). It has been hypothesized that these polymorphisms play an important role in TB susceptibility. RANTES recruits *M. tuberculosis*-infected macrophages and suppresses intracellular growth of the bacterium. The -403 A/G polymorphism is one of the most important SNPs in *RANTES* as the A allele showed an 8-fold increased transcription activity compared to that of the G allele (Nickel *et al.*, 2000). Therefore this study focused on whether a correlation exists between this polymorphism and susceptibility to TB.

3.1.6 Natural resistance-associated macrophage protein 1 (NRAMP1) also known as Solute carrier family 11a member 1 (SLC11A1)

The first susceptibility gene to be identified from a mouse model of mycobacterial disease was *NRAMP1* (*SLC11A1*), which is involved in macrophage activation. The gene encoding the natural resistance-associated macrophage protein (*NRAMP1*) is located on chromosome 2q35 and contains 15 exons that span 12kb of DNA and consist of 550 amino acids. NRAMP1 plays an important role in the early innate immune response to mycobacterial infection by activating the microbial responses in infected macrophages. It is a divalent transporter localized to the late endosomal membrane that regulates cytoplasmic cation levels by specifically regulating the iron metabolism in the macrophages, leading to possible containment of early mycobacterial infections (Wyllie *et al.*, 2002). It has also been hypothesized that variants in the *NRAMP1* gene may alter the function of the protein (Fig. 3.8) (Gruenheid *et al.*, 1997).

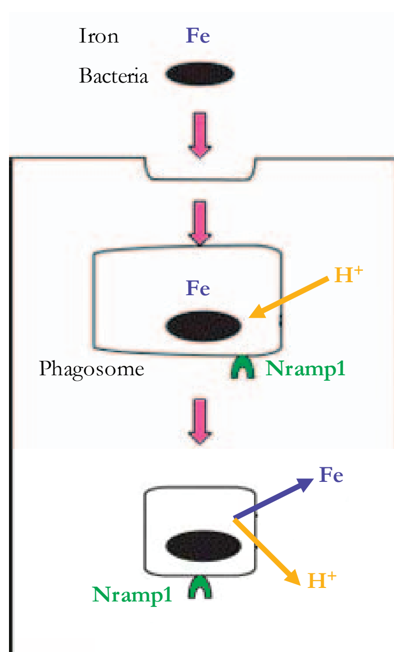


Fig. 3.8 The NRAMP1 protein is a transmembrane cation iron transporter. It is recruited to the phagosome of the macrophage after both bacteria and iron are ingested. NRAMP1 transports iron and protons such as H^+ out of the phagosome into the macrophage cytoplasm. Adjusted from (Gruenheid *et al.*, 1997).

Numerous genetic variants of *NRAMP1* have been associated with TB in several studies of which the 3'UTR and 5' $(GT)_n$ variants were shown to be the most important (Bellamy *et al.,* 1998; Hoal *et al.*, 2004). The 5' $(GT)_n$ variant, a microsatellite, is a repeat in the promoter region and the 3'UTR variant is a TGTG deletion located 55nt downstream of the last codon in exon 15 (1729 +55del4) (Fig. 3.9).

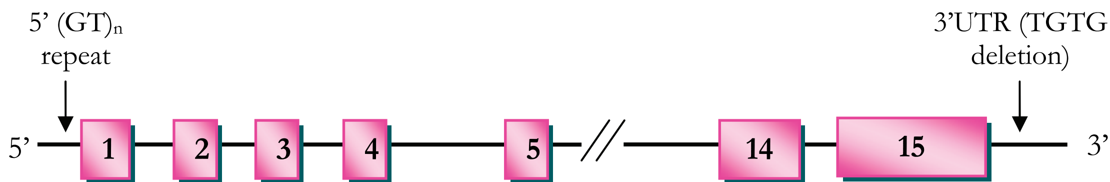


Fig. 3.9 *NRAMP1* and the polymorphisms investigated in this study.

The 5' $(GT)_n$ variant has been shown to be of functional importance as it influenced NRAMP1 expression in a luciferase reporter assay (Searle & Blackwell, 1999). The 118 allele ($(GT)_9$) was associated with higher expression and resistance to TB whereas the 120 allele ($(GT)_{10}$) was associated with lower promoter activity and susceptibility to TB (Awomoyi *et al.,* 2002; Hoal *et al.*, 2004). In a Japanese study it was found that the 5' $(GT)_9$ variant was found to be significantly associated with TB (Gao *et al.*, 2000), however a Brazilian study did not find this variant to be associated with TB (Shaw *et al.*, 1997).

The 3'UTR variant has also been linked to a significantly increased risk of pulmonary tuberculosis in West Africa (Bellamy *et al.*, 1998), Asia (Ryu *et al.*, 2000) and South Africa (Hoal *et al.*, 2004). The 3'UTR allelic frequencies differed significantly amongst the different Gambian ethnic groups in West Africa, once again highlighting the importance of the make-up of the population groups used in genetic case-control studies (Bellamy *et al.*, 1998).

Due to all these conflicting results, a meta-analysis of *NRAMP1* polymorphisms was carried out confirming the importance of the 3'UTR and 5' $(GT)_n$ variants and their involvement in tuberculosis susceptibility (Li *et al.*, 2006). It is unclear whether any of the other polymorphisms in *NRAMP1*, such as the 3'UTR TGTG deletion affects the function or whether they are in linkage disequilibrium with another functional polymorphism that has not yet been described.

3.1.7 Interleukin-1 receptor antagonist (IL-1Ra)

The interleukin-1 receptor antagonist (*IL-1Ra*) gene is located on chromosome 2q14-q21 in close proximity to the *IL-1a* and *IL-1β* genes (Fig. 3.10). Both these genes (*IL-1a* and *IL-1β*) are involved in regulation of immunological and inflammatory reactions as they bind to the IL-1 receptor and set off the recruitment and activation of macrophages leading to potent pro-inflammatory immune responses (Dinarello, 1988). IL-1Ra also binds to the receptor but acts as a competitive inhibitor of IL-1 biological activity, therefore the interaction between IL-1 and IL-1Ra determines the initiation, persistence or termination of the pro-inflammatory response (McIntyre *et al.*, 1991). Polymorphisms in *IL-1Ra* have previously been associated with TB susceptibility (Bellamy *et al.,* 1998; Gomez *et al.*, 2006; Wilkinson *et al.*, 1999).

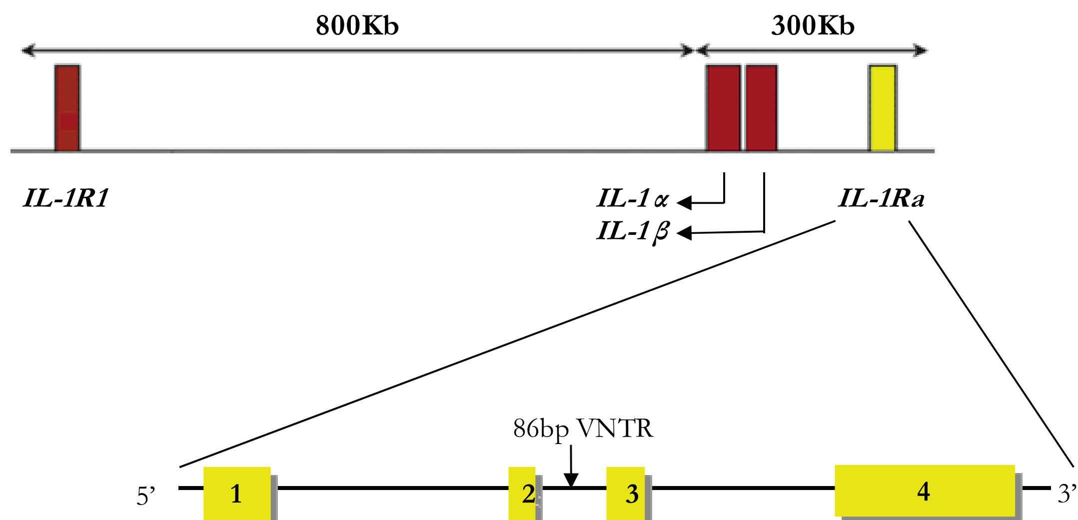


Fig. 3.10 The *IL-1* gene cluster: Interleukin -1 receptor 1 (*IL-1R1*), interleukin-1 alpha and beta (*IL-1a* & *IL-1β*) and interleukin-1 receptor antagonist (*IL-1Ra*). *IL-1Ra* is expanded to show the polymorphism investigated in this study. Adapted from (Walley et al., 2004).

One of the most important *IL-1Ra* variants is the 86bp tandemly repeated sequence, also known as a minisatellite, consisting of two to six copies in intron 2 (Lennard *et al.*, 1992). This variable number of tandem repeat (VNTR) variant has previously been associated with several clinical conditions such as sepis and increased risk of death due to sepsis (Ma *et al.*, 2002), ankylosing spondylitis (McGarry *et al.*, 2001), febrile seizures (Tsai *et al.*, 2002) and susceptibility to TB (Bellamy *et al.*, 1998).

The *IL-1Ra* allele 2 has been associated with several auto-immune diseases (Blakemore *et al.,* 1995; Clay *et al.*, 1994) and the macrophages of allele 2 carriers produced increased IL-1Ra and decreased IL-1-α levels when stimulated in vitro with granulocyte-macrophage colony stimulating factor (Danis *et al.*, 1995). The Gambian study in West Africa found the *IL-1Ra* allele 2 heterozygote's to be less frequent among TB patients than controls (Bellamy *et al.*, 1998). It is difficult to predict the effect in *IL-1Ra* allele 2 homozygote's since the prevalence was reported to be considerably lower in Black Africans and African Americans than in the white population (Mwantembe *et al.,* 2001; Rider *et al.*, 2000).

3.2 MATERIALS AND METHODS

3.2.1 SP-D

Genomic DNA from all patients was analysed using the polymerase chain reaction (PCR) to identify the presence of the wild type and/or variant allele at the Met11Thr SNP in *SP-D*. The primers used to analyse this variant are shown in Table 3.1 (Diangelo *et al.*, 1999). PCR was performed in a total reaction volume of 25µl containing 100ng template DNA, 1.5µl of each primer (10µM stock), 2.5µl dNTPs (2.5mM stock) (Bioline), 2µl MgCl (25mM stock), 1U HotStarTaq DNA polymerase (Qiagen) and PCR buffer. The PCR amplification conditions were as follows: initial denaturing step at 95^0C for 15 min, followed by 5 cycles of 95^0C for 30 s, 62^0C for 45 s and 72^0C for 30 s, decreasing the melting temperature (T_m) by 1^0C every cycle until the optimal T_m (57^0C) was reached. After these five cycles the reaction was carried out for a further 35 cycles of 95^0C for 30 s, 57^0C for 30 s and 72^0C for 30 s and a final extension at 72^0C for 5 min.

This Met11Thr (rs721917) SNP was detected by restriction enzyme digestion following PCR. Of the PCR product (59bp), 10.5µl was used in a total volume of 15µl containing the appropriate buffer and 1U of the restriction enzyme *Fsp*I (New England Biolabs). Samples were incubated overnight at 37^0C. Fragment sizes produced upon enzyme digestion were 59bp, 40 bp and 19 bp. The restriction enzyme digested PCR products were analysed on a 15% (w/v) non-denaturing polyacrylamide gel (15% polyacrylamide, 10% (w/v) ammoniumpersulfate and TEMED) at room temperature at 200V for 45 min. For electrophoresis 1x TBE buffer (0.089M Tris, 0.089M Boric acid and 20mM EDTA, pH

8.0) was used and DNA fragments were visualized with UV light following staining in 0.5μg/ml ethidium bromide.

Hardy-Weinberg equilibrium (HWE) was assessed in the control group. Logistic regression was used for all modeling in this study and all models included age and gender, in order to adjust for the differences in age (usually small) and gender (usually large) between the groups. We assessed both genotype and allelic association for individual variants. The genotypes were treated as categorical variables for the genotype association and numerical for the allelic association. For variants with more than 2 alleles, each allele's counts were modeled separately, in order to identify the specific allele having a significant effect. All the analyses were done in R (www.r-project.org) and R package DGCgenetics (http://www-gene.cimr.cam.ac.uk/clayton/software/DGCgenetics.

3.2.2 MBL

Genomic DNA from all patients was analysed using the polymerase chain reaction (PCR) to identify the presence of the wild type and/or variant alleles in *MBL*. The two sets of primers (MBLexon1For/Rev and MBLFor/Rev D allele) used to analyse these variants (Madsen *et al.*, 1995) are shown in Table 3.1. Two separate PCRs were performed in a total reaction volume of 25μl each containing 100ng template DNA, 1.5μl of each primer (10μM stock), 2.5μl dNTPs (2.5mM stock) (Bioline), 2μl MgCl (25mM stock), 1U HotStarTaq DNA polymerase (Qiagen) and PCR buffer. The PCR amplification conditions for the first primer set (MBLexon1For/Rev) were as follows: initial denaturing step at 95^0C for 15 min, followed by 35 cycles of 95^0C for 30 s, 58^0C for 1 min and 72^0C for 2 min and a final extension at 72^0C for 5 min. The PCR conditions were exactly the same for the second primer set (MBLFor/Rev D allele) except for the T_m that was at 55^0C for 30 s.

These SNPs (rs5030737, rs1800450 and rs1800451) were detected by restriction fragment length polymorphism (RFLP) following PCR. PCR amplification with the MBLexon1For/Rev primer set yielded a fragment size of 349bp. This PCR product was digested with two different enzymes, *Ban*I and *Mbo*II (New England Biolabs), to reveal the presence or absence of the B and C allele respectively. PCR amplification with the MBLFor/Rev D allele primer set yielded a fragment size of 125bp and was digested with the *Mlu*I (New England Biolabs) restriction enzyme to reveal the presence or absence of

Table 3.1. Primers used in the PCR amplification of variants investigated in this study

Variant	Location	Primer sequence (5'-3')		PCR product
SP-D (Met11Thr) rs721917	Exon 1	*For:*	CTGGAAGCAGAAATGAAGAC	59 bp
		Rev:	ACCAGGGTGCAAGCACTGCG	
IFN-γ (874 A/T) rs2430561	Intron 1	*PrimerA:*	TTCTTACAACACAAAATCAAATCA	261 bp
		PrimerT:	TTCTTACAACACAAAATCAAATCT	
		Rev:	TCAACAAAGCTGATACTCCA	
		IntCFor:	CCTTCCAACCATTCCCTTA	426 bp
		IntCRev:	TCACGGATTTCTGTTGTGTTTC	
NRAMP1 $(GT)_n$ repeat	Promoter region	*For:*	ACTCGCATTAGGCCAACGAG	118 bp, 120 bp or 122 bp
		Rev:	HEX-TTCTGTGCCTCCCAAGTTAGC	
NRAMP1 (TGTG) deletion	Exon 15	*For:*	GCATCTCCCCAATTCATGGT	Deletion = 240 bp
		Rev:	6FAM-AACTGTCCCACTCTATCCTG	No deletion = 244 bp
RANTES (403 A/G) rs2107538	Promoter region	*PrimerA:*	CATGGATGAGGGAAAGGCGG	464 bp
		PrimerG:	CATGGATGAGGGAAAGGCGA	
		IntCfor:	GGTCGCTTAGCAAGTAAATGG	681 bp
		SharedRev:	GTCCACGTGCTGTCTTGATC	
MBP (codon 52 (Arg-Cys) rs5030737 (codon 54 (Gly-Asp) rs1800450 (codon 57 (Gly-Glu) rs1800451	Exon 1	*For:*	AGTCGACCCAGATTGTAGGACAGAG	349 bp
		Rev:	AGGATCCAGGCAGTTTCCTCTGGAAGG	
		ForD:	CATCAACGGCTTCCCAGGCAAAGACGCG	125 bp
		RevD:	AGGATCCAGGCAGTTTCCTCTGGAAGG	
IFNGR1 (1050 T/G) rs11914	Exon 7	*For:*	CTTTAATCAATTTTTCTCCC	381 bp
		Rev:	TCACTCTCAGAACAATTTCT	
IL1Ra (86bp VNTR)	Intron 2	*For:*	HEX-CTCAGCAACACTCCTAT	410 bp, 240 bp, 500 bp, 325 bp or 595 bp
		Rev:	CTCATCTTCCTGGTCTGC	
IL-8 (100 C/T) rs2227538	Promoter region	*PrimerC:*	ACCATCTCACTGTGTGTAGAC	426 bp
		PrimerT:	ACCATCTCACTGTGTGTAGAT	
		Rev:	ATTGTCTGAGTTACCTTGCTC	
		IntCFor:	GGTCGCTTAGCAAGTAAATGG	618 bp
		IntCRev:	GTCCACGTGCTGTCTTGATC	

the D allele. Of the two PCR products 10.5µl were used in a total volume of 15µl containing the appropriate buffer and 1U of the suitable restriction enzyme. Samples were incubated overnight at 37^{0}C. The restriction enzyme-digested PCR products were analysed on a 2% (w/v) agarose gel at 200V for 1 hour. For electrophoresis 1x Sodium Boric (SB) buffer was used and DNA fragments were visualized with UV light following staining in 0.5µg/ml ethidium bromide. See section 3.2.1 for statistical analysis.

3.2.3 IFN-γ and IFNGR1

i) *IFN-γ* (+874 T/A)

Genomic DNA from all patients was analysed using the Amplification Refractory Mutation System (ARMS). ARMS is an oligonucleotide ligation technique where two sets of primers are designed for both SNP alleles (wildtype and variant) (Newton *et al.*, 1989). This was a two (one for each allele) tube reaction. An internal control is critical when using the ARMS to confirm the robustness of the PCR reaction in both tubes (Fig. 3.11).

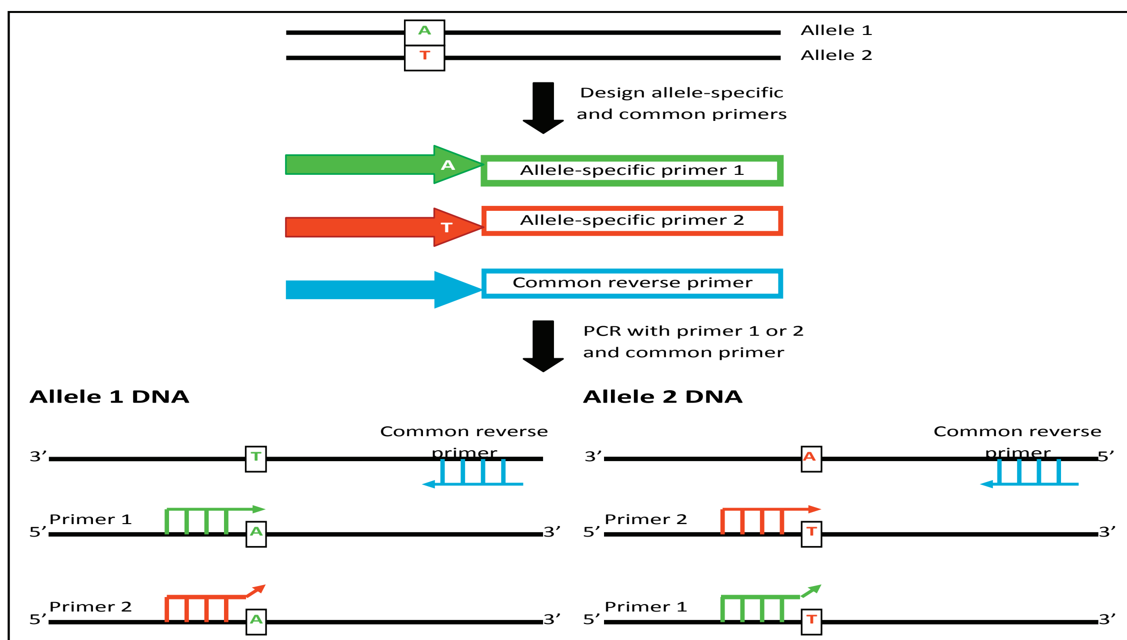


Fig. 3.11 An example of how an ARMS-PCR works (Strachan & Read, 1999)

The ARMS primers used to analyse this variant (Pravica *et al.*, 2000) can be seen in Table 3.1. PCR was done in a total reaction volume of 30µl containing 100ng template DNA, 1.5µl of each primer (10µM stock), 3µl dNTPs (2.5mM stock) (Bioline), 2µl MgCl (25mM stock), 1U HotStarTaq DNA polymerase (Qiagen) and PCR buffer. The two-step PCR amplification conditions were as follows: initial denaturing step at 95^{0}C for 15 min,

followed by 10 cycles of 95^0C for 15 s, 62^0C for 50 s and 72^0C for 1 min and then 25 cycles of 95^0C for 20 s, 56^0C for 50 s and 72^0C for 1 min. Final extension was at 72^0C for 10 min.

The PCR products were 426bp (internal control) and 261bp (SNP) and analysed on a 2% (w/v) agarose gel at 200V for 1 hour. For electrophoresis 1x SB buffer was used and DNA fragments were visualized with UV light following staining in 0.5µg/ml ethidium bromide. See section 3.2.1 for statistical analysis.

ii) *IFNGR1* (1050 T/G)

Primers used for this PCR can be seen in Table 3.1. PCR was performed in a total reaction volume of 25µl containing 100ng template DNA, 1.5µl of each primer (10µM stock), 2.5µl dNTPs (2.5mM stock) (Bioline), 2µl MgCl (25mM stock), 1U HotStarTaq DNA polymerase (Qiagen) and PCR buffer. The PCR amplification conditions were as follows: initial denaturing step at 95^0C for 15 min, followed by 35 cycles of 95^0C for 1 min, 53^0C for 1 min and 72^0C for 1 min and a final extension at 72^0C for 4 min.

This SNP was detected by restriction enzyme digestion following PCR. Of the PCR product (381bp), 10.5µl was used in a total volume of 15µl containing the appropriate buffer and 1U of the restriction enzyme *Bfa*I (New England Biolabs). Samples were incubated overnight at 37^0C. Fragment sizes produced upon enzyme digestion were 381bp, 213bp and 168bp. The restriction enzyme digested PCR products were analysed on a 2% (w/v) agarose gel at 200V for 1 hour. For electrophoresis 1x SB buffer was used and DNA fragments were visualized with UV light following staining in 0.5µg/ml ethidium bromide. See 3.2.1 for statistical analysis.

3.2.4 IL-8

Genomic DNA from all patients was analysed using the ARMS two (one for each allele) tube reaction including an internal control (see section 3.2.3). The ARMS primers used to analyse this variant were designed according to the recommendations of Newton *et al.* (Newton *et al.*, 1989) and can be seen in Table 3.1. PCR was performed in a total reaction volume of 25µl containing 100ng template DNA, 1.5µl of each primer (10µM stock), 2.5µl dNTPs (2.5mM stock) (Bioline), 2µl MgCl (25mM stock), 1U HotStarTaq DNA polymerase (Qiagen) and PCR buffer. The PCR amplification conditions were as follows:

initial denaturing step at 95^0C for 15 min, followed by 35 cycles of 95^0C for 30 s, 58^0C for 30 s and 72^0C for 1 min and a final extension at 72^0C for 10 min.

The PCR products were 618bp (internal control) and 426bp (SNP) and analysed on a 2% (w/v) agarose gel at 200V for 1 hour. For electrophoresis 1x SB buffer was used and DNA fragments were visualized with UV light following staining in 0.5µg/ml ethidium bromide. See section 3.2.1 for statistical analysis.

3.2.5 RANTES

Genomic DNA from all patients was analysed using the two tube ARMS including an internal control (see section 3.2.3). The ARMS primers used to analyse this variant can be seen in Table 3.1. PCR was done in a total reaction volume of 25µl containing 100ng template DNA, 1.5µl of each primer (10µM stock), 2.5µl dNTPs (2.5mM stock) (Bioline), 2µl MgCl (25mM stock), 1U HotStarTaq DNA polymerase (Qiagen) and PCR buffer. The PCR amplification conditions were as follows: initial denaturing step at 95^0C for 15 min, followed by 30 cycles of 94^0C for 30 s, 60^0C for 30 s and 72^0C for 1 min followed by a final extension at 72^0C for 10 min.

The PCR products were 681bp (internal control) and 464bp (SNP) and analysed on a 2% (w/v) agarose gel at 200V for 1 hour. For electrophoresis 1x SB buffer was used and DNA fragments were visualized with UV light following staining in 0.5µg/ml ethidium bromide. See section 3.2.1 for statistical analysis.

3.2.6 NRAMP1

Genomic DNA from all patients was analysed using the polymerase chain reaction (PCR) followed by microsatellite analysis using the ABI PRISM 3130xl Genetic Analyser (Perkin Elmer, Applied Biosystems, Warrington WA, Great Britain) and the analysis software GeneMapper version 3.7. Microsatellites, also known as short tandem repeats (STRs), are polymorphic DNA loci that contain a repeat sequence of 2-7 nucleotides. The primers used to analyse both these variants (Bellamy *et al.*, 1998) can be seen in Table 3.1. The two forward primers of the 5' $(GT)_n$ repeat and the 3'UTR deletion were fluorescently labeled (HEX: green and 6FAM: blue respectively) to permit multiple loci analysis in the same capillary injection. PCR was performed in a total reaction volume of 25µl containing

100ng template DNA, 0.25µl of each *NRAMP1* 5'(GT)$_n$ primer (10µM stock), 0.6µl of each *NRAMP1* 3'UTR primer (10µM stock), 2.5µl dNTPs (2.5mM stock) (Bioline), 2µl MgCl (25mM stock), 1U HotStarTaq DNA polymerase (Qiagen) and PCR buffer. The PCR amplification conditions were as follows: initial denaturing step at 95^{0}C for 15 min, followed by 25 cycles of 94^{0}C for 30 s, 62^{0}C for 30 s and 72^{0}C for 1 min followed by a final extension at 72^{0}C for 5 min. See section 3.2.1 for statistical analysis.

3.2.7 IL-1Ra

Genomic DNA from all patients was analysed using the polymerase chain reaction (PCR) followed by microsatellite analysis using the ABI PRISM 3130xl Genetic Analyser (Perkin Elmer, Applied Biosystems, Warrington WA, Great Britain) and the analysis software GeneMapper version 3.7. The primers used to analyse this variant (Murtha *et al.*, 2006) can be seen in Table 3.1. The forward primer was fluorescently labeled (HEX) to permit electrophoretic analysis of the VNTR. PCR was performed in a total reaction volume of 25µl containing 100ng template DNA, 0.25µl of each primer (10µM stock), 2.5µl dNTPs (2.5mM stock) (Bioline), 2µl MgCl (25mM stock), 1U HotStarTaq DNA polymerase (Qiagen) and PCR buffer. The PCR amplification conditions were as follows: initial denaturing step at 95^{0}C for 15 min, followed by 25 cycles of 94^{0}C for 30 s, 56^{0}C for 30 s and 72^{0}C for 1 min followed by a final extension at 72^{0}C for 5 min. See section 3.2.1 for statistical analysis.

3.3 RESULTS

3.3.1 SP-D

This Met11Thr SNP was detected, and was seen in both cases and controls (Fig. 3.12). The controls were in Hardy-Weinberg equilibrium (HWE) for this polymorphism (P=0.351), and no significant difference was found between the TB cases or TBM cases versus controls with respect to genotypic or allelic frequencies (Table 3.2).

3.3.2 MBL

These three SNPs (rs5030737, rs1800450 and rs1800451) were detected, and seen in both groups (Fig. 3.13). The controls were in HWE (P=0.239). No significant difference was found between the TB cases or TBM cases versus controls with respect to genotypic frequencies (P=0.368, P=0.116 respectively). The borderline significant association found

with the D allele when comparing TBM cases to controls (P=0.045) might be an artifact due to the low frequency of the D allele and the small sample size of the TBM group (Table 3.3).

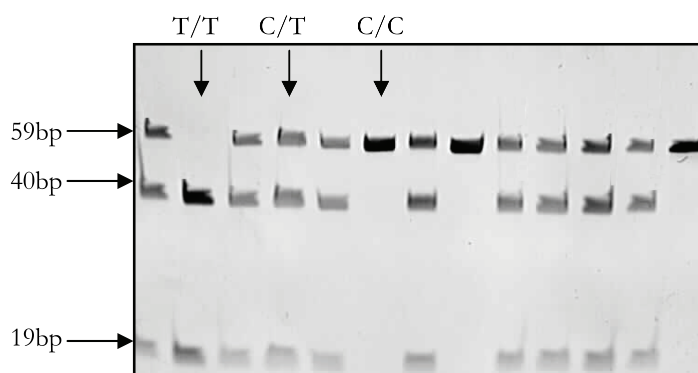


Fig. 3.12 Restriction enzyme analysis of the Met11Thr SNP in *SP-D*

Table 3.2 Genotypic and allelic distribution of the Met11Thr SNP in South African Coloured TB patients compared to controls. A single *P*-value is given for genotype and allele association, adjusted for age and gender.

	Genotypic distribution no. of patients (%)				Allelic distribution		
SP-D	T/T	C/T	C/C	*P*-value	T	C	*P*-value
TB patients, n=494	201 (40%)	196 (40%)	97 (20%)	0.315	598 (61%)	390 (39%)	0.234
Controls, n=304	103 (34%)	141 (46%)	60 (20%)		347 (57%)	261 (43%)	
	T/T	C/T	C/C	*P*-value	T	C	*P*-value
TBM patients, n=46	16 (35%)	20 (43%)	10 (22%)	0.887	40 (43%)	52 (57%)	0.778
Controls, n=304	103 (34%)	141 (46%)	60 (20%)		347 (57%)	261 (43%)	

Table 3.3 Genotypic and allelic distribution of the *MBL* exon 1 variant.

Genotypic distribution no. of patients (%)				Allelic distribution			
Variant	TB patients (n = 499)	Controls (n = 313)	*P*-value[1]		TB patients	Controls	*P*-value[2]
MBL							
A/A	363 (73%)	211 (67%)		A	860 (86%)	524 (84%)	0.241
A/B	63 (13%)	50 (16%)	0.368	B	67 (7%)	50 (8%)	0.369
A/C	56 (11%)	39 (13%)		C	56 (6%)	39 (6%)	0.401
A/D	15 (3%)	13 (4%)		D	15 (1%)	13 (2%)	0.173
B/B	2 (0%)	0 (0%)					
Variant	TBM patients (n = 49)	Controls (n = 313)	*P*-value		TBM patients	Controls	*P*-value
MBL							
A/A	39 (80%)	211 (67%)		A	88 (90%)	524 (84%)	0.242
A/B	4 (8%)	50 (16%)	0.116	B	4 (4%)	50 (8%)	0.215
A/C	6 (12%)	39 (13%)		C	6 (6%)	39 (6%)	0.572
A/D	0 (0%)	13 (4%)		D	0 (0%)	13 (2%)	0.045*

* Association is significant P≤0.05
[1]A single *P*-value is given for genotype association, adjusted for age and gender.
[2]Separate tests were done for each allele versus all others.

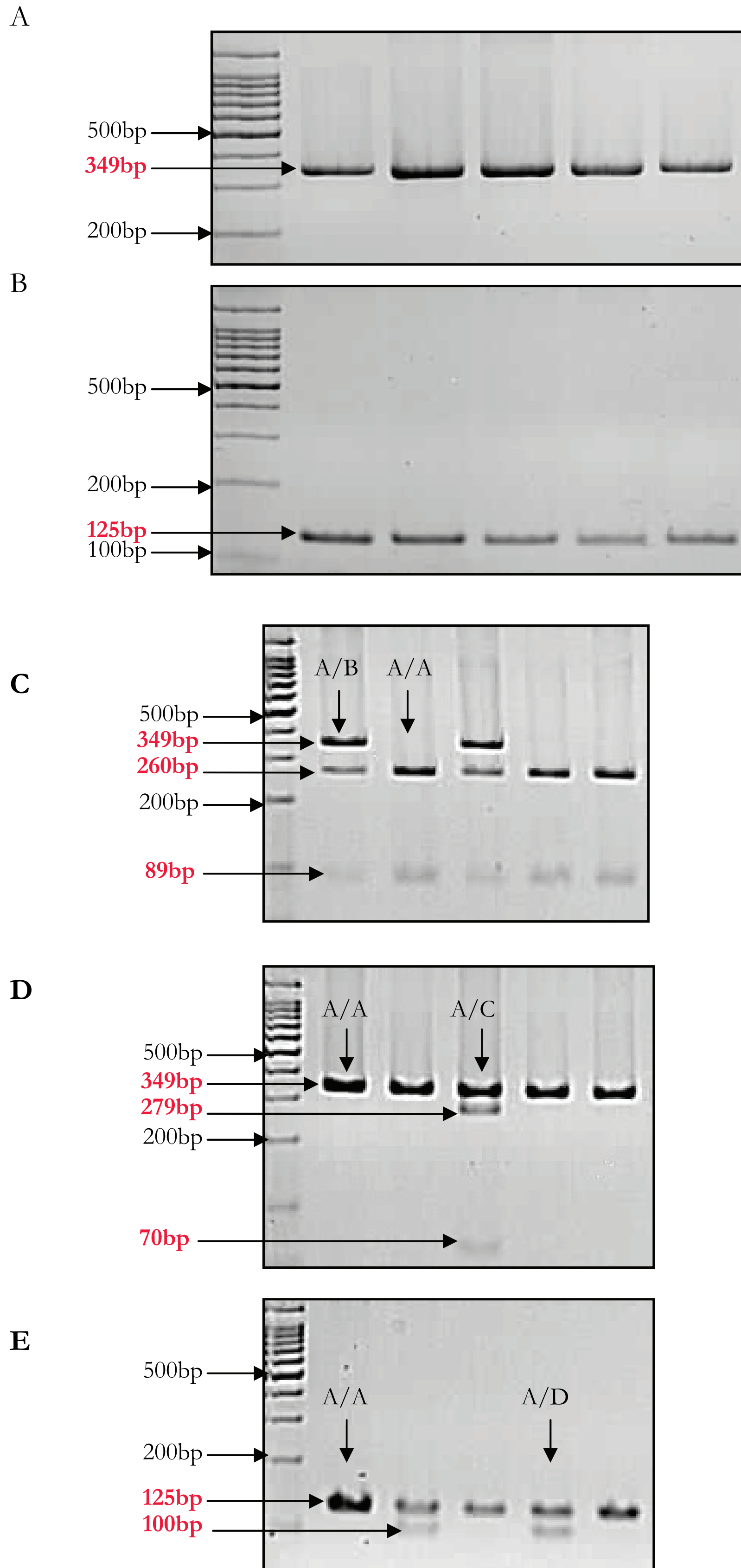


Fig. 3.13 Restriction enzyme analysis of the 3 SNPs located in exon 1 of *MBL*, [A] PCR product with MBLexon1 primer set, [B] PCR product with MBL D allele primer set, [C] Digestion with *Ban*I (B allele), [D] Digestion with *Mbo*II (C allele), [E] Digestion with *Mlu*I (D allele).

3.3.3 IFN-γ and IFNGR1

The +874 A/T and +1050 G/T SNPs in *IFN-γ* and *IFNGR1* respectively, were detected, and seen in both groups (Fig. 3.14 and Fig. 3.15).

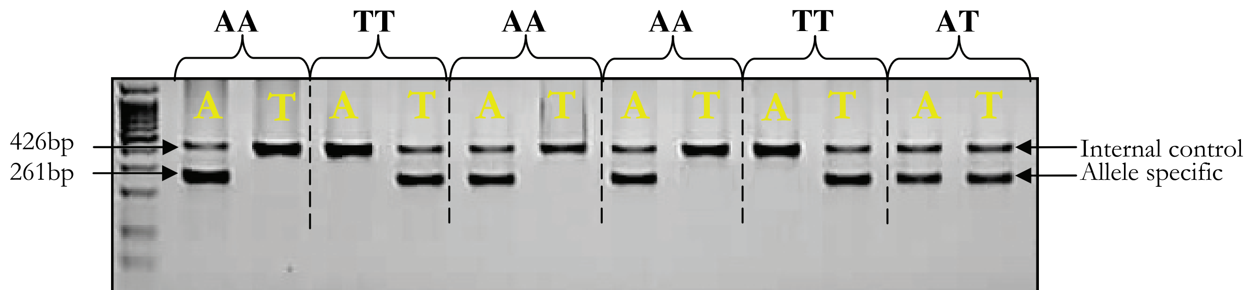


Fig. 3.14 The *IFN-γ* +874 A/T polymorphism detected using the ARMS method

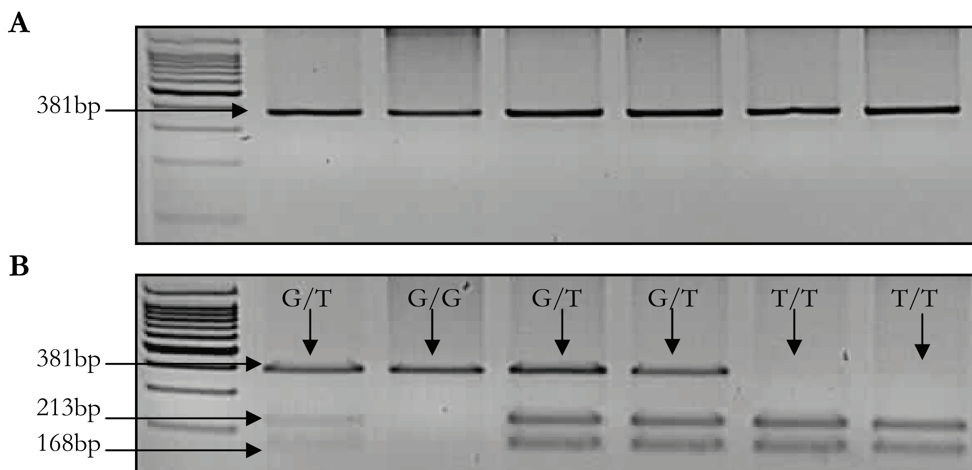


Fig. 3.15 Detection of the polymorphism +1050 G/T in *IFNGR1*: [A] PCR product followed by, [B] restriction enzyme analysis.

The controls were in HWE for both the *IFN-γ* (P=0.284) and the *IFNGR1* (P=0.257) SNPs. For the *IFN-γ* polymorphism, no significant difference was found between either the TB cases or controls with respect to genotype (P=0.448) or allele frequencies (P=0.244) nor the TBM cases (P=0.593, P=0.311 respectively) (Table 3.4). Significant differences were found with the *IFNGR1* polymorphism regarding allele frequencies in both TB and TBM cases versus controls (P=0.022, P=0.037 respectively) but not with genotype frequency (Table 3.5).

Table 3.4 Genotypic and allelic distribution of the *IFN-γ* variant in TB cases *vs* controls and TBM cases *vs* controls, adjusted for age and gender.

	Genotypic distribution no. of patients (%)				Allelic distribution		
IFN-γ	A/A	A/T	T/T	*P*-value	A	T	*P*-value
TB patients, n=500	266 (53%)	194 (39%)	40 (8%)	0.448	726 (73%)	274 (27%)	0.244
Controls, n=315	158 (50%)	124 (40%)	33 (10%)		440 (70%)	190 (30%)	
	A/A	A/T	T/T	*P*-value	A	T	*P*-value
TBM patients, n=47	27 (57%)	16 (34%)	4 (9%)	0.593	70 (74%)	24 (26%)	0.311
Controls, n=315	158 (50%)	124 (40%)	33 (10%)		440 (70%)	190 (30%)	

Table 3.5 Genotypic and allelic distribution of the *IFNGR1* variant in TB cases *vs* controls and TBM cases *vs* controls, adjusted for age en gender.

	Genotypic distribution no. of patients (%)				Allelic distribution		
IFNGR1	T/T	T/G	G/G	*P*-value	T	G	*P*-value
TB patients, n=492	418 (85%)	69 (14%)	5 (1%)	0.064	905 (92%)	79 (8%)	0.022*
Controls, n=315	250 (79%)	59 (19%)	6 (2%)		559 (89%)	71 (11%)	
	T/T	T/G	G/G	*P*-value	T	G	*P*-value
TBM patients, n=44	40 (91%)	4 (9%)	0 (0%)	0.105	84 (95%)	4 (5%)	0.037*
Controls, n=315	250 (79%)	59 (19%)	6 (2%)		559 (89%)	71 (11%)	

*Association is significant P≤0.05

3.3.4 IL-8

This +100 C/T (rs2227538) SNP was detected in both groups. An example of the SNP detected is depicted in Fig. 3.16. The controls were in HWE for this polymorphism (*P*=0.305). No significant difference was found between the TB or TBM cases versus controls with respect to genotypic frequencies (*P*=0.062, *P*=0.680 respectively), however a significant association was found between the TB cases and controls for the allelic frequency (*P*=0.019) (Table 3.6).

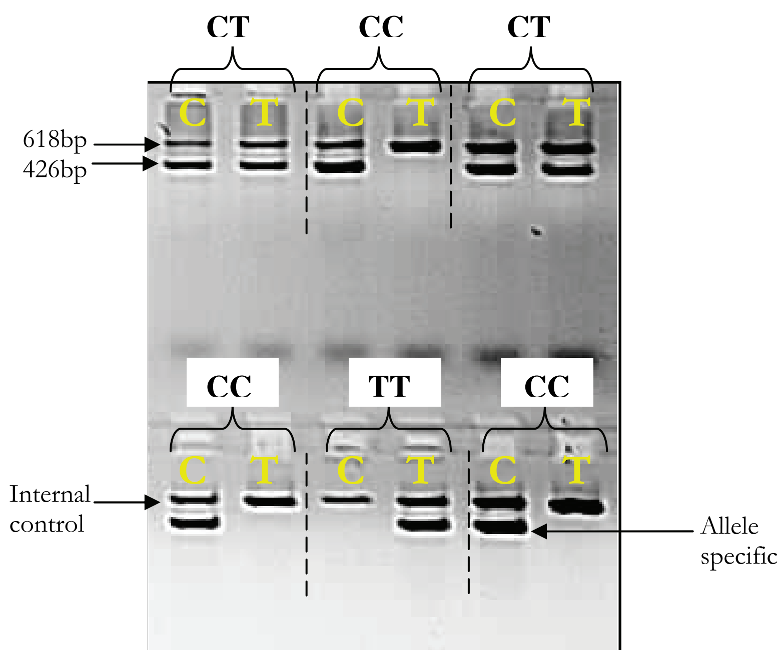


Fig. 3.16 Detection of all three genotypes of the +100 C/T *IL-8* polymorphism.

Table 3.6 Genotypic and allelic distribution of the *IL-8* variant in TB cases *vs* controls and TBM cases *vs* controls, adjusted for age and gender.

IL-8	Genotypic distribution no. of patients (%)				Allelic distribution		
	C/C	C/T	T/T	*P*-value	C	T	*P*-value
TB patients, n=496	379 (76%)	105 (21%)	12 (3%)	0.062	863 (87%)	129 (13%)	0.019*
Controls, n=312	259 (83%)	49(16%)	4 (1%)		567 (91%)	57 (9%)	
	C/C	C/T	T/T	*P*-value	C	T	*P*-value
TBM patients, n=48	38 (79%)	9 (19%)	1 (2%)	0.680	85 (89%)	11 (11%)	0.518
Controls, n=312	259 (83%)	49(16%)	4 (1%)		567 (91%)	57 (9%)	

*Association is significant P≤0.05

3.3.5 RANTES

The-103 A/G (rs2107538) SNP in *RANTES* was detected in both groups and is depicted in Fig. 3.17. The controls were in HWE (*P*=0.909). No significant difference was found between the TB cases and controls with respect to genotype (*P*=0.451) or allele frequencies (*P*=0.347). However, the allele frequency was significantly different between the TBM cases and controls (*P*=0.037) (Table 3.7), although not significantly different between TB cases and controls (*P* = 0.347).

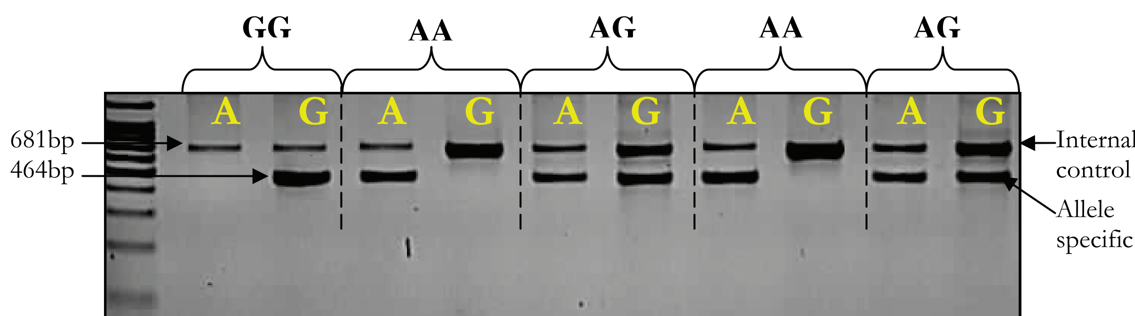


Fig. 3.17 The detection of all three genotypes of the *RANTES* - 403 A/G polymorphism.

Table 3.7 Genotypic and allelic distribution of the -403 A/G *RANTES* variant in TB cases *vs* controls and TBM cases *vs* controls, adjusted for age and gender.

	Genotypic distribution no. of patients (%)				Allelic distribution		
RANTES	A/A	A/G	G/G	*P*-value	A	G	*P*-value
TB patients, n=493	141 (29%)	228 (46%)	122 (25%)	0.451	514 (52%)	472 (48%)	0.347
Controls, n=309	81 (26%)	153 (50%)	75 (24%)		315 (51%)	303 (49%)	
	A/A	A/G	G/G	*P*-value	A	G	*P*-value
TBM patients, n=46	17 (37%)	21 (46%)	8 (17%)	0.110	55 (60%)	37 (40%)	0.037*
Controls, n=309	81 (26%)	153 (50%)	75 (24%)		315 (51%)	303 (49%)	

*Association is significant P≤0.05

3.3.6 NRAMP1

Both the 5' $(GT)_n$ repeat and the 3'UTR deletion variants in *NRAMP1* were detected in both groups (Fig. 3.18). The control group was in HWE for both the 5' $(GT)_n$ repeat (*P*=1.00) and the 3'UTR deletion (*P*=0.634). No significant difference was found between the 5' $(GT)_n$ repeat variant and the TB or TBM cases and controls with respect to genotype frequency (*P*=0.077, *P*=0.979 respectively). The 118 and 120 allele frequencies were significantly different between the TB cases and controls (*P*=0.011, *P*=0.011 respectively) (Table 3.8). No significant association was found between the 3'UTR deletion and TB or TBM in this study (Table 3.9).

Table 3.8 Genotypic and allelic distribution of the *NRAMP1* 5' $(GT)_n$ repeat in TB cases *vs* controls and TBM cases *vs* controls.

	Genotypic distribution no. of patients (%)			Allelic distribution			
Variant	TB patients (n = 498)	Controls (n = 315)	*P*-value[1]		TB patients	Controls	*P*-value[2]
NRAMP1 repeat							
118/118	301 (61%)	216 (69%)		118	776 (78%)	523 (83%)	0.011*
118/120	173 (35%)	91 (29%)	0.077	120	217 (22%)	106 (17%)	0.011*
118/122	1 (0%)	0 (0%)		122	3 (0%)	1 (0%)	0.686
120/120	21 (4%)	7 (2%)					
120/122	2 (0%)	1 (0%)					
Variant	TBM patients (n = 49)	Controls (n = 315)	*P*-value		TBM patients	Controls	*P*-value
NRAMP1 repeat							
118/118	34 (69%)	216 (69%)		118	82 (84%)	523 (83%)	0.947
118/120	14 (29%)	91 (29%)	0.979	120	16 (16%)	106 (17%)	0.926
118/122	0 (0%)	0 (0%)		122	0 (0%)	1 (0%)	0.675
120/120	1 (2%)	7 (2%)					
120/122	0 (0%)	1 (0%)					

*Association is significant P≤0.05
[1]A single *P*-value is given for genotype association, adjusted for age and gender.
[2]Separate tests were done for each allele versus all others.

Table 3.9 Genotypic and allelic distribution of the *NRAMP1* 3'UTR deletion in TB cases *vs* controls and TBM cases *vs* controls, adjusted for age and gender.

	Genotypic distribution no. of individuals (%)				Allelic distribution		
NRAMP del	N/N	N/D	D/D	*P*-value	N	D	*P*-value
TB patients, n=492	393 (80%)	85 (17%)	14 (3%)	0.202	871 (89%)	113 (11%)	0.170
Controls, n=312	261 (84%)	48 (15%)	3 (1%)		570 (91%)	54 (9%)	
	N/N	N/D	D/D	*P*-value	N	D	*P*-value
TBM patients, n=49	37 (76%)	12 (24%)	0 (0%)	0.221	86 (88%)	12 (12%)	0.195
Controls, n=312	261 (84%)	48 (15%)	3 (1%)		570 (91%)	54 (9%)	

3.3.7 IL-1Ra

All five alleles previously reported for the *IL-1Ra* 86bp VNTR were detected and an example of the different alleles obtained can be seen in Fig. 3.19. The control group was in HWE (*P*=0.721). The five different alleles represent 2 to 6 repeats (Table 3.10). Allele 1 (4 repeats) and allele 2 (2 repeats) were the most common, which corresponds with previous studies (Tarlow *et al.,* 1993; Witkin *et al.*, 2002).

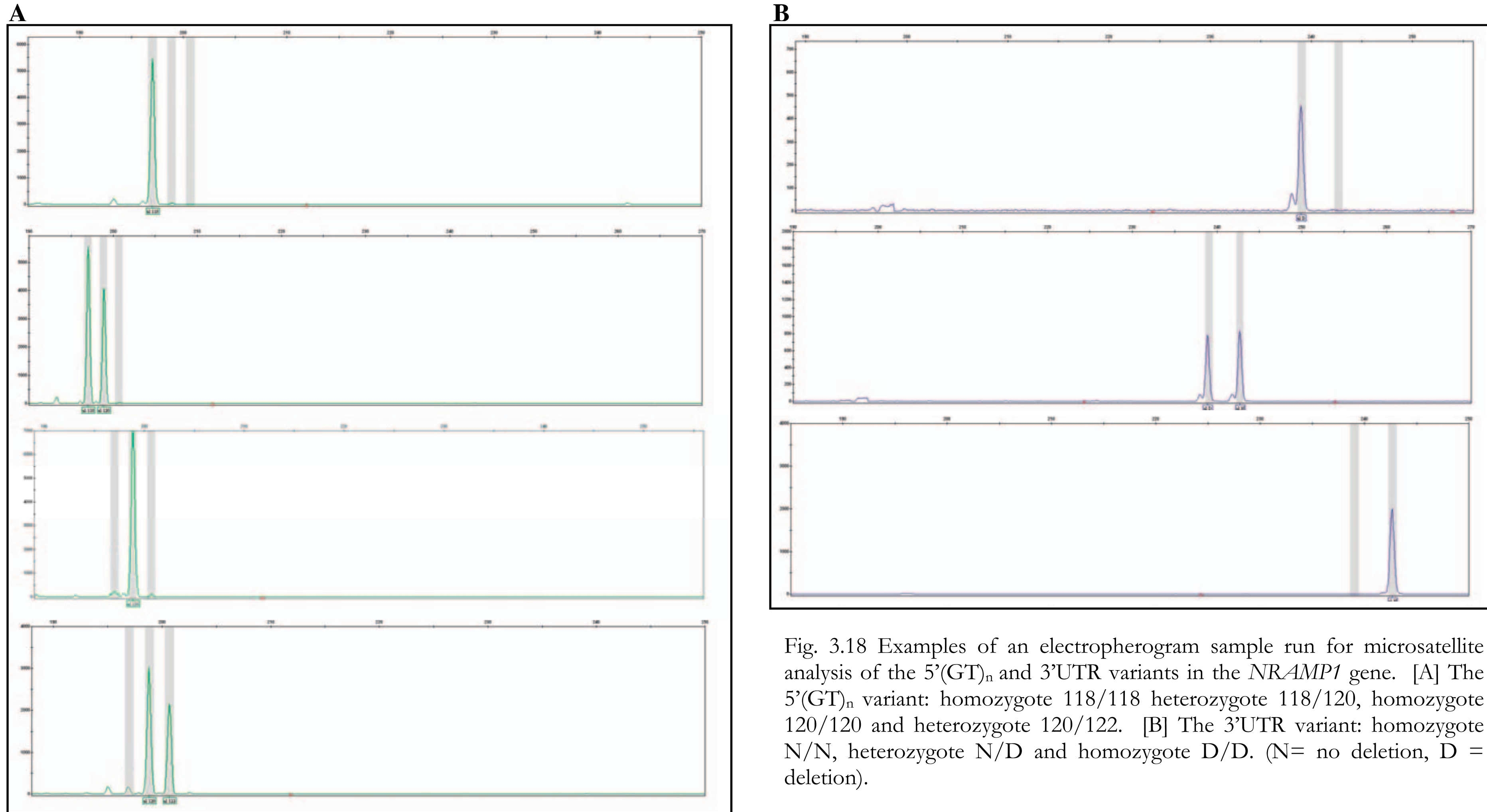


Fig. 3.18 Examples of an electropherogram sample run for microsatellite analysis of the 5'$(GT)_n$ and 3'UTR variants in the *NRAMP1* gene. [A] The 5'$(GT)_n$ variant: homozygote 118/118 heterozygote 118/120, homozygote 120/120 and heterozygote 120/122. [B] The 3'UTR variant: homozygote N/N, heterozygote N/D and homozygote D/D. (N= no deletion, D = deletion).

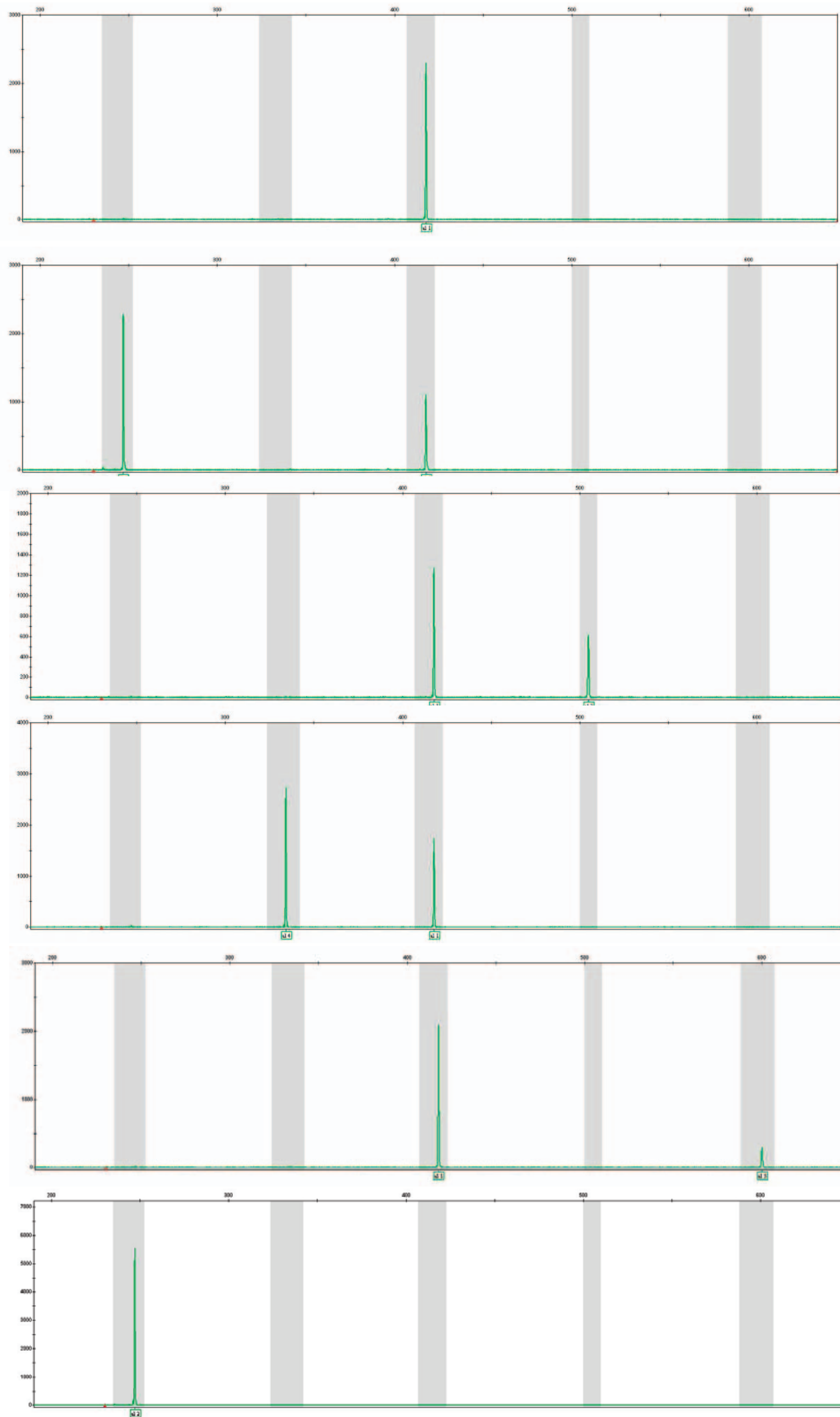


Fig. 3.19 All alleles obtained with the microsatellite analysis of the 86bp VNTR in the *IL-1Ra* gene. The grey areas are the bins set up for the different alleles.

Table 3.10 *IL-1Ra* allele frequencies among TB cases and controls in this South African population

Allele	Frequency in controls	Frequency in cases	PCR product size	Number of repeats
1	0.76	0.80	410 bp	4
2	0.20	0.16	240 bp	2
3	0.01	0.02	500 bp	5
4	0.03	0.02	325 bp	3
5	0.01	0.01	595 bp	6

Statistically significant associations were observed with TB status for the genotypes (P=0.037) and alleles 1 and 2 (P=0.019, P=0.025 respectively). No statistically significant difference was found when comparing the TBM cases to the controls (Table 3.11).

Table 3.11 *IL-1Ra* 86bp VNTR genotype and allele frequencies among TB and TBM cases and controls.

Genotypic distribution no. of patients (%)				Allelic distribution			
Variant	TB patients (n=465)	Controls (n=281)	*P*-value[1]		TB patients	Controls	*P*-value[2]
IL-1Ra							
1/1	308 (67%)	173 (62%)	0.037*	1	742 (80%)	425 (75%)	0.019*
1/2	94 (21%)	62 (22%)		2	146 (15%)	112 (20%)	0.025*
1/3	13 (3%)	1 (0%)		3	15 (2%)	3 (1%)	0.102
1/4	14 (3%)	11 (4%)		4	20 (2%)	17 (3%)	0.152
1/5	5 (1%)	5 (2%)		5	7 (1%)	5 (1%)	0.685
2/2	24 (5%)	25 (9%)					
2/3	2 (0%)	0 (0%)					
2/4	2 (0%)	0 (0%)					
3/3	0 (0%)	1 (0%)					
4/4	2 (0%)	3 (1%)					
5/5	1 (0%)	0 (0%)					
Variant	TBM patients (n=41)	Controls (n=281)	*P*-value		TBM patients	Controls	*P*-value
IL-1Ra							
1/1	26 (63%)	173 (62%)	0.817	1	65 (79%)	425 (75%)	0.332
1/2	11 (27%)	62 (22%)		2	15 (18%)	112 (20%)	0.441
1/3	0 (0%)	1 (0%)		3	0 (0%)	3 (1%)	0.554
1/4	2 (5%)	11 (4%)		4	2 (3%)	17 (3%)	0.896
1/5	0 (0%)	5 (2%)		5	0 (0%)	5 (1%)	0.348
2/2	2 (5%)	25 (9%)					
2/3	0 (0%)	0 (0%)					
2/4	0 (0%)	0 (0%)					
3/3	0 (0%)	1 (0%)					
4/4	0 (0%)	3 (1%)					
5/5	0 (0%)	0 (0%)					

*Association is significant P≤0.05
[1]A single *P*-value is given for genotype association, adjusted for age and gender.
[2]Separate tests were done for each allele versus all others.

3.9 DISCUSSION

Susceptibility to tuberculosis was investigated by means of case-control association studies in the SAC population. Polymorphisms in *SP-D, MBL, IFN-γ*, *IFNGR1, IL-8, RANTES*, *IL-1Ra* and *NRAMP1* were investigated. All these genes play an important role in either the uptake of *M. tuberculosis* in the alveolar macrophages or in the inflammatory response and have previously been associated with TB.

We found no statistically significant associations between TB susceptibility in the SAC population and the polymorphisms in *SP-D*, *MBL*, *IFN-γ* or *RANTES*. These polymorphisms have previously been associated with TB susceptibility in other population groups and in smaller studies in our SAC population unpublished data obtained by several colleagues in the TB host genetics group). This, once again, demonstrates the common phenomenon in association studies where the first report can show a positive association and subsequent studies are often negative as a result of genetic differences between populations (allele based) or a difference in sample size. If a positive association was detected in a small sample size but did not hold when the sample size was increased (in the same population group) the first association may have been due to chance therefore unable to be replicated.

For the *IFNGR1* 1050 T/G (rs11914) polymorphism, a statistically significant difference in allelic frequencies was found between the TB cases and TBM cases, and controls (P=0.022, P=0.037 respectively) and a trend towards significance was observed comparing the genotypic frequencies in TB cases and controls (P=0.064). The T allele was more prevalent in the cases than controls. However, this is a synonymous SNP (Ser/Ser) therefore it will be unlikely to have an effect on the protein expression. This SNP might be in linkage disequilibrium with another genetic marker or there may be an interaction between this particular SNP and other polymorphisms in the IFN-γ pathway.

A statistically significant difference in allelic frequency for the *IL-8* +100 C/T (rs2227538) SNP was found between the TB cases and controls (P=0.019) while a trend towards significance was observed with the genotypic frequency (P=0.062). The T allele was more prevalent in the TB cases while the C allele was more prevalent in the controls suggesting that the T allele might be a risk factor for susceptibility to TB. In light of the fact that this

polymorphism is located one base pair upstream from the ATG start codon of the *IL-8* gene, any nucleotide change here might affect the transcriptional level of the gene. Increased levels of IL-8 have been reported in the bronchoalveolar lavage fluid (BALF) of TB patients compared to controls (Dlugovitzky *et al.,* 1997; Mastroianni *et al.*, 1994; Sadek *et al.*, 1998) as well as increased IL-8 plasma concentration levels in patients who die from TB compared to those who survive (Friedland *et al.,* 1995; Pace *et al.*, 1999). Therefore it can be hypothesized that having the T allele leads to an increased transcription rate and therefore increased IL-8 expression. However, the functional work to support this hypothesis has not been done.

Of the two *NRAMP1* variants, the 5' $(GT)_n$ repeat and 3'UTR (TGTG) deletion investigated in this study, a statistically significant difference between the cases and controls was only observed for the 5' $(GT)_n$ repeat. The allelic frequencies of the 118 and 120 alleles were significantly different between the TB cases and controls (P=0.011, P=0.011 respectively). In our SAC population, the 118 allele, which is the most common allele, has a higher frequency in the controls than the cases whereas the 120 allele is more frequent among the TB cases than controls. Our findings concur with those of others where the 118 allele has been associated with higher expression and resistance to TB in a number of widely disparate populations and the 120 allele with lower promoter activity and susceptibility to TB (Awomoyi *et al.,* 2002; Hoal *et al.*, 2004). The mechanisms of how NRAMP1 influences resistance or susceptibility to TB are still speculative, but it has been shown that the 120 allele is associated with a higher induced level of the anti-inflammatory cytokine interleukin 10 (IL-10) (Awomoyi *et al.*, 2002). One of the major roles of NRAMP1 is that it acts as a divalent transporter that regulates cytoplasmic cation levels by specifically regulating the iron metabolism in the macrophages leading to possible containment of early mycobacterial infections (Wyllie *et al.*, 2002) whereas one of the major roles of IL-10 is to deactivate macrophages. Therefore the association observed with the 120 allele of the 5' $(GT)_n$ *NRAMP1* variant might be due to an interactive effect with IL-10.

In all of the above reports, statistically significant differences were found with allelic frequencies only, but not with genotypic frequencies between the TB cases and TBM cases, and controls. Not much can be deduced from allelic associations, as an individual has two

alleles at any locus, making up an individual's genotype. These two alleles can act additively or interact to influence susceptibility to a complex disease. Therefore, associations between genotypic frequencies and the disease may provide more information than the allelic association.

For the *IL-1Ra* 86bp VNTR variant, statistically significant differences in genotypic frequency (P=0.037) and allelic frequencies (P=0.019, P=0.025 for allele 1 and 2 respectively) were found between the TB cases and controls. Possibly due to the low frequency of alleles 3, 4 and 5 no significant association with TB was found in this study. In our study and in concordance with the Gambian study (Bellamy *et al.*, 1998), allele 1 was associated with an increase in probability of developing TB whereas allele 2 was associated with a decreased probability. A previous study reported that macrophages of *IL-1Ra* allele 2 carriers produced increased IL-1Ra, IL-1β and decreased IL-1α levels when stimulated in vitro with granulocyte-macrophage colony stimulating factor (Danis *et al.*, 1995). It has been proposed that imbalanced levels of IL-1Ra and IL-1β proteins may play a major role in the determination of the severity of an inflammatory reaction (Santtila *et al.*, 1998).

In the past decade, several candidate genes have been investigated regarding host genetic susceptibility to TB. Since TB is a complex disease the genetic susceptibility component of TB seems to be divided amongst many genes and is probably not due to one or even a few SNPs. Although we found significant associations between single polymorphisms in different genes and TB susceptibility, it is possible that the effect of one gene might not be revealed if the effect of another gene is not considered and therefore, focusing on single polymorphisms at a time may generate false negatives. It has become increasingly evident that gene-gene interactions might play a vital role in an individual's susceptibility to a complex disease.

In conclusion, data presented here suggests and confirms previous pilot studies of associations between the *IFNGR1, IL-8, IL-1Ra* (unpublished data) and *NRAMP1* (Bellamy *et al.,* 1998; Hoal *et al.*, 2004) polymorphisms and TB susceptibility in the SAC population.

CHAPTER 4

GENE-GENE AND GENE- (*M. TUBERCULOSIS)* STRAIN INTERACTIONS IN TUBERCULOSIS SUSCEPTIBILITY

4.1 INTRODUCTION

TB is a complex disease and the susceptibility component of TB seems to be made up of many genes as well as different genes being implicated in different populations (Maartens & Wilkinson, 2007). Many polymorphic genes have been associated with TB susceptibility (Table 4.1) of which meta-analysis has been done on interferon-gamma (*IFN-γ)* (Pacheco *et al.*, 2008) and natural resistance-associated macrophage protein (*NRAMP1* / *SLC11A1*) (Li *et al.*, 2006), confirming their role in tuberculosis susceptibility. Due to the complex nature of the immune system and the polygenic nature of complex diseases it has become increasingly evident that gene-gene interactions play a far more important part in an individual's susceptibility to a complex disease than single polymorphisms would on their own (Ritchie *et al.,* 2001; Tsai *et al.*, 2003; Williams *et al.*, 2000). Gene-gene interaction, also known as epistasis, can be defined as the interaction between genes that are not on the same locus and which can be very distant from each other in the genome. In a gene-gene interaction, one gene can enhance (synergistic epistasis) or hinder (antagonistic epistasis) the expression of another gene (Nagel, 2005).

Gene-gene interaction analysis is still a fairly new approach to elucidate susceptibility to complex diseases, therefore little has been published regarding gene-gene interactions in infectious diseases such as TB. Gene-gene interactions have been found in non-infectious diseases such as asthma (Kim *et al.,* 2006; Lee *et al.*, 2004), Parkinson's disease (Deng *et al.,* 2004; Pankratz *et al.*, 2003), diabetes (Bergholdt *et al.,* 2005; Koeleman *et al.*, 2004) and rheumatoid arthritis (Martinez *et al.*, 2006). Therefore, one aim of our study was to assess the possibility of gene-gene interactions, focusing on 11 polymorphisms in nine genes that have previously been associated with TB in our studies and / or others.

These include two SNPs in the dendritic cell-specific ICAM-3 grabbing nonintegrin (*DC-SIGN)* (-366 A/G) and *DC-SIGN* (-871 A/G); *IFN-γ* (+874 T/A); interferon-gamma receptor 1 (*IFNGR1)* (1050 T/G); interleukin 8 (*IL-8)* (+100 C/T); interleukin-1 receptor antagonist (*IL-1Ra)* (86 bp VNTR); a three-SNP haplotype in mannose-binding lectin (*MBL)* at codons 52 (Arg-Cys), 54 (Gly-Asp) and 57 (Gly-Glu) in exon 1; *NRAMP1*, also known as *SLC11A1* $(GT)_n$ repeat*, NRAMP1* (TGTG) deletion; the C-C chemokine regulated upon activation, normal T-cell expressed and secreted (*RANTES)*, also known as

CCL5, that is a ligand for chemokine receptor 5 (-403A/G); and surfactant protein D (*SP-D)* (Met11Thr).

Table 4.1. Genes associated with TB

Genes	References
CR1	(Fitness *et al.*, 2004)
DC-SIGN	(Barreiro *et al.*, 2006)
HLA DR2	(Rajalingam *et al.*, 1996) (Hawkins *et al.,* 1988; Sanjeevi *et al.*, 1992)
IFN-γ	(Tso *et al.*, 2005) (Lopez-Maderuelo *et al.,* 2003; Rossouw *et al.*, 2003)
IFN-γ-R1	(Awomoyi *et al.*, 2004) (Bulat-Kardum *et al.,* 2006; Cooke *et al.*, 2006)
IL-1 complex	(Wilkinson *et al.*, 1999) (Gomez *et al.*, 2006)
IL-1RA	(Bellamy *et al.*, 1998)
IL-8	(Ma *et al.*, 2003)
MBL	(Hoal-van Helden *et al.*, 1999) (Garcia-Laorden *et al.,* 2006; Soborg *et al.*, 2003)
RANTES (CCL5)	(Chu *et al.*, 2007)
SLC11A1/ NRAMP1	(Delgado *et al.*, 2002) (Liu *et al.*, 2004) (Hoal *et al.,* 2004; Malik *et al.*, 2005)
SP-D	(Floros *et al.*, 2000)
Vitamin D receptor	(Bellamy *et al.*, 1999) (Babb *et al.,* 2007; Roth *et al.*, 2004; Wilkinson *et al.*, 2000)

A combined data set consisting of 106 variants was constructed from our database and also used for gene-gene interaction analysis. All of these genes had previously been selected as candidate genes for susceptibility to TB based on their biological function and proposed role in the multistep course of this complex disease, or were selected according to similar work in the literature.

In addition to host genotype conferring susceptibility, it has been hypothesized that the interaction between the genotype of the human host, and the bacterial strain genotype, could influence the susceptibility of the host in terms of both the progression to disease and perhaps the type of disease seen. Several *M. tuberculosis* strains, such as LAM (Latin American and Mediterranean), Haarlem, C strain and X strain exist all over the world, and the W-Beijing family of strains has been studied extensively (Malik & Godfrey-Faussett, 2005). An investigation into 875 strains from 80 different countries led to the hypothesis that ancestral *M. tuberculosis* originated and drifted together with humans "out of Africa" (Gagneux *et al.*, 2006). Remarkably, evidence exists that six lineages of *M. tuberculosis* have adapted to specific populations (Maartens & Wilkinson, 2007) and Hanekom *et al.* showed that strains from a defined sublineage may have been selected by a human population in a defined geographical setting (Hanekom *et al.*, 2007). This argument is supported by the fact that the HLA allele frequencies, vary widely between human populations with different historical backgrounds, with some alleles completely absent in certain populations (Lombard *et al.*, 2006) and that the HLA genotype has been associated with susceptibility to *M. tuberculosis*. To our knowledge, only one study has looked at the interaction between bacterial strains and host genotypes thus far and found an association between the C allele of the Toll-like receptor 2 (*TLR2*) 597 T/C SNP of the host and the bacterial genotype in *M. tuberculosis* disease (Caws *et al.*, 2008).

The Beijing family of strains, most easily recognised by its distinctive spoligotyping pattern (Bifani *et al.*, 2002), was first reported in 1995 (van Soolingen *et al.*, 1995). This is probably the most dominant or distinctive strain worldwide and has been reported in many countries, such as Vietnam, Thailand, Hong Kong, Russia and South Africa (Anh *et al.,* 2000; Chan *et al.*, 2001; Drobniewski *et al.*, 2002; Glynn *et al.*, 2006; Prodinger *et al.*, 2001; Warren *et al.*, 2004). It has been suggested that Beijing strains are hypermutable and have greater virulence compared to other strains (Rad *et al.*, 2003) and they have also been associated with a higher frequency of drug resistance (Dale *et al.,* 2005; Rad *et al.*, 2003; van der Spuy *et al.*, 2009).

From the original Mycobacterial complex, *M. tuberculosis* has evolved into three principle genetic groups (PGG 1, 2 and 3). Based on the combinations of two polymorphisms at the *kat*G codon 463 and *gyr*A codon 95 site, isolates of *M. tuberculosis* were assigned to one of

these three groups (Sreevatsan *et al.*, 1997). PGG 1, known as the ancestral group, is genetically more diverse than PGG 2 and 3, since it is older and therefore has accumulated variation. It consists mainly (95%) of the Beijing family, whereas PGG2 consists of LAM, LCC and Haarlem families and PGG3 of other less frequent strains. Therefore comparing Beijing to non Beijing strains will be essentially the same as comparing PGG 1 to other PGGs.

4.2 MATERIALS AND METHODS

4.2.1 Study population

i) Condensed set

The same culture proven or smear positive TB cases (*n*=505) and controls (*n*=318) that were used for the case-control association study (chapter 3) were used for the gene-gene interactions between the 11 variants genotyped.

ii) Combined set

Culture proven or smear positive TB cases (*n* = 339) and controls (*n* = 206) were used for the combined dataset of 106 variants scattered across many genes. The number of participants is less than for the original gene-gene interaction study (11 variants) as they were selected to maximize overlapping genotype information and minimize missing values. All these variants have previously been genotyped by our group in the same study cohort. This combined dataset included well-known and well-characterised genes that have previously been shown to possibly play a role in susceptibility to TB.

In Ravensmead-Uitsig, in which we performed the majority of the DNA collection, others in our department have also extensively researched the strains of *M. tuberculosis* present, and the relative virulence of the strains in terms of their spread and prevalence in the community (Warren *et al.*, 2004). Over the past decade several thousand individuals with culture proven TB have been characterised with respect to the *M. tuberculosis* strain harboured. The overlap between patients with bacterial strain information and a genetic profile was about 50%. Therefore only those patients (*n* = 229) could be used to study this interaction.

4.2.2 Genotyping

See chapter three for detailed genotyping information and primer sequences for polymorphisms in *MBL, NRAMP1, IL-1Ra, RANTES, IFN-γ, IL-8, SP-D* and *IFNGR1*. Based on the fact that two *DC-SIGN* polymorphisms (-366 A/G (rs4804803) and -871 A/G (rs735239)) had previously been genotyped in the same study cohort and had been associated with TB susceptibility (Barreiro *et al.*, 2006), it was decided to include these two SNPs in the condensed gene-gene interaction analysis. These two promoter polymorphisms were genotyped using the template-directed dye-terminator incorporation with fluorescence polarization detection (TDI-FP) using Victor PCR. See (Barreiro *et al.*, 2006) for detailed genotyping information. Genotyping information regarding all the other TB susceptibility genes used in the combined dataset, is not shown here, but can be found in our department.

4.2.3 Bacterial strains

Sputum samples were collected from suspect TB patients followed, by a process of decontamination and inoculation (Fig. 4.1).

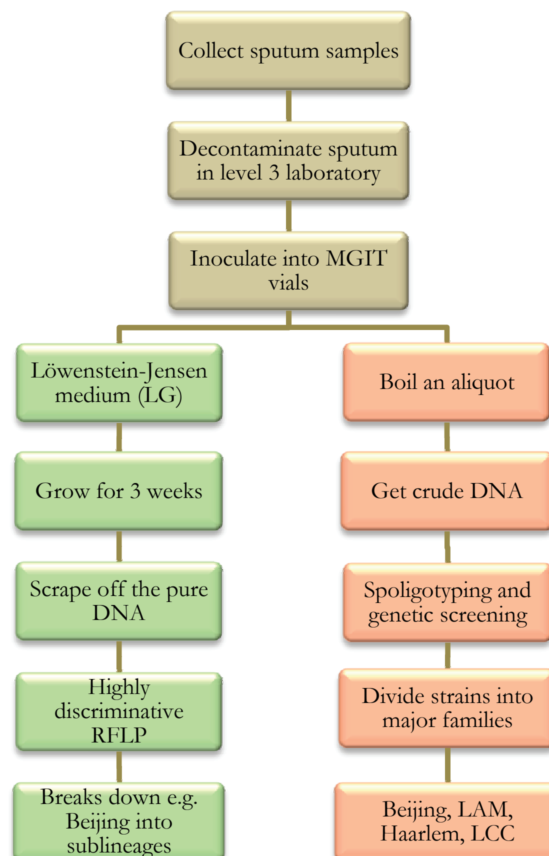


Fig. 4.1 A simplified illustration, explaining the process from sputum collection to the *M. tuberculosis* strain information.

The WHO recommended certain media to be used of which the Löwenstein-Jensen (LG) medium was one (World Health Organization, 2001). Thereafter, genotyping by RFLP analysis or spoligotyping was done (Warren *et al.*, 2002). Bacterial strains can be divided into two groups, namely clustered or non-clustered, also known as unique. The clustered group can be divided into big clusters (defined as five or more individuals harbouring the same strains) and small clusters (defined as less than five patients with the same strains).

4.2.4 Statistical analysis

Hardy-Weinberg equilibrium (HWE) was assessed in the control group for all the genetic variants tested. Logistic regression was used for all modeling in this study and all models included age and gender, in order to adjust for the differences in age (usually small) and gender (usually large) between the groups. We assessed both genotype and allelic association for individual variants. The genotypes were treated as categorical variables for the genotype association and numerical for the allelic association. For variants with more than 2 alleles, each allele's counts were modeled separately, in order to identify the specific allele having a significant effect. For SNPs, there is no need to model both alleles, because the effect of one will necessarily be the converse of that of the other. (If allele X increases one's likelihood by a factor of 2, allele not-X will decrease it by a factor of 2.) In order to assess the interaction between genes, the alleles from pairs of variants were coded numerically. The interaction between all pairs of variants described in the combined set above, as well as two SNPs (-871 and -336) in the DC-SIGN-encoding gene *CD209,* that were previously associated with TB in our study population (Barreiro *et al.*, 2006) was assessed. The outcome modeled was logit(TB), the natural log of the odds ratio (OR) of TB versus no TB, so that single effect sizes can be converted to odds ratios, and interpreted as such. Unfortunately interaction terms are not so straightforward to interpret.

We also created a model for case-control status (logit(TB)) as a function of age, gender and all eleven candidates as main effects (not interactions) and selected an optimal model from it, using a stepwise backwards procedure, by repeatedly discarding the least important term (by AIC, Akaike's criterion) until the optimal selection of terms is reached.

A dichotomous genotype-trait association test was performed to assess the interactions between the human genotype and the bacterial strain. This test was adjusted for age. All

the analyses were done in R (www.r-project.org) and R package DGCgenetics (http://www-gene.cimr.cam.ac.uk/clayton/software/DGCgenetics).

4.3 RESULTS

There were 55 (17%) males in the control group and 227 (45%) in the cases of the condensed set. The median age was 30 in the control group and 31 for the TB cases. The odds of having TB decreases significantly with increasing age (OR=0.987; *P*=0.021) after adjusting for gender. Therefore all our models and all results reported here were adjusted for age and gender. All variants genotyped in this study were in HWE in the control group.

4.3.1 Association studies

None of the five SNPs (rs721917, rs2430561, rs2107538, rs11914 and rs2227538) in *SP-D*, *IFN-γ*, *RANTES*, *IFNGR1* and *IL-8* respectively, or the *MBL* three SNP haplotype showed a statistically significant difference between the TB cases and controls at the genotype level, but significant trends in allele frequency were detected for the SNPs in *IFNGR1* and *IL-8* respectively. This implies that there is an additive effect, but we did not have enough power to detect and confirm it with the genotype test. Allele T of the *IL-8* SNP increased the odds of TB by 48% (OR=1.48; *P*=0.019), whereas the G allele of the *IFNGR1* SNP decreased the odds of TB by 34% (OR=0.66; *P*=0.022) (Table 4.2).

The 118 allele of the *NRAMP1* $(GT)_n$ repeat showed a statistically significant decrease of almost 30 % in odds of having TB (versus no TB) associated with each 118 allele compared to the other alleles (OR=0.71; *P* = 0.011). The frequency of allele 122 was 0.003 and 0.002 in the TB cases and controls respectively, confirming previous reports regarding the rarity of this allele. The 122 allele frequency was too low for any conclusions to be drawn. The *NRAMP1* (TGTG) deletion did not show a statistically significant difference between the TB cases and controls at the genotype or allele level (Table 4.2).

Table 4.2. Case-control analysis of polymorphisms and TB in the South African Coloured population.

Genotype				Allele			
Variant (rs number)	**TB patients n (frequency)**	**Controls n (frequency)**	***P*-value[1]**		**TB patients n (frequency)**	**Controls n (frequency)**	***P*-value[1,2,3]**
SP-D (rs721917)							
C/C	97 (0.20)	60 (0.20)		C	390 (0.39)	261 (0.43)	0.234
C/T	196 (0.40)	141 (0.46)	0.315	T	598 (0.61)	347 (0.57)	
T/T	201 (0.40)	103 (0.34)					
IFN-γ (rs2430561)							
A/A	266 (0.53)	158 (0.50)		A	726 (0.73)	440 (0.70)	0.244
A/T	194 (0.39)	124 (0.39)	0.448	T	274 (0.27)	190 (0.30)	
T/T	40 (0.08)	33 (0.10)					
RANTES (rs2107538)							
A/A	143 (0.29)	81 (0.26)		A	514 (0.52)	315 (0.51)	0.347
A/G	228 (0.46)	153 (0.50)	0.451	G	472 (0.48)	303 (0.49)	
G/G	122 (0.25)	75 (0.24)					
IFNGR1 (rs11914)							
T/T	418 (0.85)	250 (0.79)		T	905 (0.92)	559 (0.89)	**0.022**
T/G	69 (0.14)	59 (0.19)	0.064	G	79 (0.08)	71 (0.11)	
G/G	5 (0.01)	6 (0.02)					
IL-8 (rs2227538)							
C/C	379 (0.76)	259 (0.83)		C	863 (0.87)	567 (0.91)	**0.019**
C/T	105 (0.21)	49 (0.16)	0.062	T	129 (0.13)	57 (0.09)	
T/T	12 (0.02)	4 (0.01)					
NRAMP1 (TGTG) deletion							
D/D	14 (0.03)	3 (0.01)		D	113 (0.11)	54 (0.09)	0.170
D/N	85 (0.17)	48 (0.15)	0.202	N	871 (0.89)	570 (0.91)	
N/N	393 (0.80)	261 (0.84)					
NRAMP1 $(GT)_n$ repeat							
118/118	301 (0.60)	216 (0.69)		118	776 (0.78)	523 (0.83)	**0.011**
118/120	173 (0.35)	91 (0.29)	0.077	120	217 (0.22)	106 (0.17)	**0.011**
118/122	1 (0.00)	0 (0.00)		122	3 (0.00)	1 (0.00)	0.686
120/120	21 (0.04)	7 (0.02)					
120/122	2 (0.00)	1 (0.00)					
MBP (Exon 1)							
A/A	363 (0.73)	211 (0.67)		A	860 (0.86)	524 (0.84)	0.241
A/B	63 (0.13)	50 (0.16)	0.368	B	67 (0.07)	50 (0.08)	0.369
A/C	56 (0.11)	39 (0.12)		C	56 (0.06)	39 (0.06)	0.401
A/D	15 (0.03)	13 (0.04)		D	15 (0.02)	13 (0.02)	
B/B	2 (0.004)	0 (0.00)					

[1] *P*-values are adjusted for age and gender.
[2] *P*-values < 0.05 are bold.
[3] Each allele is tested against all other alleles combined

All five alleles previously reported for the *IL-1Ra* 86 bp VNTR were present in our study (see chapter 3). Allele 1 (4 repeats) and allele 2 (2 repeats) were the most common, which corresponds with previous studies (Tarlow *et al.,* 1993; Witkin *et al.*, 2002). Statistically significant associations with TB were observed for alleles 1 and 2. Allele 2 was associated with a decrease in the OR (OR=0.74; *P*=0.025) and allele 1 with a corresponding increase (OR=1.34; *P*=0.019). Possibly due to the low frequency of alleles 3, 4 and 5, no significant association with TB was found in this study (Table 4.3), or in any other study that we are aware of.

Table 4.3. *IL-1Ra* 86bp VNTR genotype and allele frequencies among TB cases and controls.

Genotype				Allele			
Variant	TB patients n (frequency)	Controls n (frequency)	*P*-value[1]		TB patients n (frequency)	Controls n (frequency)	*P*-value[2]
IL-1Ra 86bp VNTR							
1/1	308 (0.66)	173 (0.62)	**0.037**	1	742 (0.80)	425 (0.76)	**0.019**
1/2	94 (0.20)	62 (0.22)		2	146 (0.16)	112 (0.20)	**0.025**
1/3	13 (0.03)	1 (0.00)		3	15 (0.02)	3 (0.01)	0.102
1/4	14 (0.03)	11 (0.04)		4	20 (0.02)	17 (0.03)	0.152
1/5	5 (0.01)	5 (0.02)		5	7 (0.01)	5 (0.01)	0.685
2/2	24 (0.05)	25 (0.09)					
2/3	2 (0.00)	0 (0.00)					
2/4	2 (0.00)	0 (0.00)					
3/3	0 (0.00)	1 (0.00)					
4/4	2 (0.00)	3 (0.01)					
5/5	1 (0.00)	0 (0.00)					

[1] A single *P*-value is given for genotype association, adjusted for age and gender.
[2] Separate tests adjusted for age and gender were done for each allele versus all others.
Significant *P*-values are bold.

4.3.2 Optimal model

The optimal model, which best describes and predicts TB case-control status, contained age, gender and four variants, *NRAMP1* $(GT)_n$ repeat, *IFNGR1, IL-1Ra*, IFN-γ and *IL-8* of which the following were individually statistically significant: gender ($P < 0.001$), *NRAMP1* ($P = 0.014$), *IFNGR1* ($P = 0.016$) and *IL-1Ra* allele 2 ($P = 0.026$)(Table 4.4).

Table 4.4. Summary of the optimal model for logit(TB), selected from the model containing age, gender and alleles from 11 variants ($P < 0.001$).

Variables		Effect	SE	*P*-value
Age:	no. of years	-0.01	0.01	0.127
Gender:	male	1.47	0.19	**<0.001**
NRAMP1 $(GT)_n$ repeat:	no. of 120 alleles	0.37	0.15	**0.014**
IFNGR1 (1050 T/G):	no. of G alleles	-0.47	0.19	**0.016**
IFN-γ (+874 T/A):	no. of T alleles	-0.19	0.13	0.134
IL-1Ra (86bp VNTR):	no. of 2 alleles	-0.30	0.14	**0.026**
IL-1Ra (86bp VNTR):	no. of 4 alleles	-0.63	0.33	0.053
IL-8 (+100 C/T):	no. of T alleles	0.34	0.18	0.060

4.3.3 Gene-gene interactions

i) Condensed set

Of the 103 possible interacting pairs we detected significant interactions between eight pairs of variants. The models containing the interactions are awkward to interpret. Therefore we list the pairs of variants and illustrate the modelled interactions graphically (Table 4.5; Fig. 4.2).

Table 4.5. Pairs of variants showing significant interactions in models adjusted for age and gender. For variants with more than 2 alleles, the specific allele in the model is given.

Variant	Allele	Variant	Allele	*P*-value
IFNGR1 (1050 T/G)	G	NRAMP1 (TGTG) deletion	Deletion	**0.0078**
NRAMP1 $(GT)_n$ repeat	120	NRAMP1 (TGTG) deletion	Deletion	**0.0184**
NRAMP1 $(GT)_n$ repeat	120	DC-SIGN (-871 A/G)	G	**0.0108**
MBP (Exon 1)	C	DC-SIGN (-871 A/G)	G	**0.0495**
MBP (Exon 1)	C	*IL-1Ra* (86bp VNTR)	2	**0.0343**
MBP (Exon 1)	D	RANTES (-403 A/G)	G	**0.0477**
IFN-γ (+874 T/A)	T	RANTES (-403 A/G)	G	**0.0312**
IFN-γ (+874 T/A)	T	*IL-8* (+100 C/T)	T	**0.0493**

The vertical axis on the graphs in Fig. 4.2 is the logit of the probability of having TB, and indicates the risk of having TB. We will interpret it as such. In graphs from non-SNPs, the effect of a single allele is illustrated, and the others are combined and denoted N. Each graph shows the risk of TB for nine possible genotype combinations for a 30 year old male. The modelled graphs for females and other ages will show the same pattern with only the

starting points differing. Women start lower, and the risk gets slightly lower with each year of age in both sexes. If the lines on the graphs were parallel, that would indicate no interaction between the variants. The differences between the slopes determine the extent of the interaction. The nature of each of the eight interactions is described below. The models fitted the observed data extremely well, with *P*-values below 0.0001 for all eight models.

A highly significant interaction was detected between *INFGR1* and the *NRAMP1* (TGTG) deletion. With the *INFGR1* T/T genotype, the deletion had no effect on the risk of having TB, but for *INFGR1* T/G heterozygotes, the logit(TB) increased slightly and for the *INFGR1* G/G homozygotes the risk increased dramatically with the number of deletions (Fig. 4.2A)

The impact of the interaction between *NRAMP1* $(GT)_n$ repeat and the *NRAMP1* (TGTG) deletion on the risk of developing TB was as follows: In the absence of the 120 allele the odds of getting TB decreased slightly with each deletion. In the presence of a single 120 allele the risk of TB increased with each deletion, while the 120/120 homozygotes showed a remarkable increase in logit(TB) with each deletion. The pattern was reversed for the 118 allele (Fig. 4.2B).

A significant interaction was detected between *NRAMP1* $(GT)_n$ repeat and *DC-SIGN* (-871 A/G). With the *DC-SIGN* A/A genotype, each 120 allele increased the logit(TB), whereas with each *DC-SIGN* G allele, the risk of TB decreased with the number of 120 alleles. The decrease in the risk of developing TB was more prominent for the *DC-SIGN* G/G homozygotes than in the heterozygotes. This pattern is reversed for the *NRAMP1* $(GT)_n$ repeat 118 allele (Fig. 4.2C).

The impact of the interaction between *MBL* and *DC-SIGN* (-871 A/G) on the risk of developing TB was as follows: In the absence of the *MBL* C haplotype the risk of getting TB decreased slightly with each *DC-SIGN G* allele; in the presence of a single C haplotype the risk of TB increased with each *DC-SIGN G* allele, while the *MBL* C/C homozygotes showed a strong increase in logit(TB) with each *DC-SIGN G* allele (Fig. 4.2D). This interaction was of borderline significance.

A significant interaction was detected between *MBL* and *IL-1Ra*. In the absence of the *MBL* C allele haplotype the odds of getting TB decreased slightly with each *IL-1Ra* allele 2. In the presence of a single C allele the risk of TB increased with each *IL-1Ra* allele 2, while the *MBL* C/C homozygotes showed a large increase in risk of TB with each *IL-1Ra* allele 2. The pattern was reversed for *IL-1Ra* allele 1 (Fig. 4.2E).

An individual with *RANTES* G/G showed an increase in risk of developing TB with increasing number of *MBL* D haplotypes, whereas the *RANTES* A/G showed a decrease. The most prominent decrease with increase in number of D alleles was in *RANTES* A/A homozygotes. The pattern was reversed with *MBL* A haplotypes (Fig. 4.2F).

A significant interaction was detected between *IFN-γ* and *RANTES*. The risk of developing TB was the highest with *IFN-γ* A/A and *RANTES* A/A genotypes. IFN-γ heterozygotes were not affected by *RANTES* 403 genotype. *IFN-γ* +874 homozygotes showed strong trends with increasing number of G alleles, increasing for *IFN-γ* +874 T/T and decreasing for *IFN-γ* +874 A/A (Fig. 4.2G).

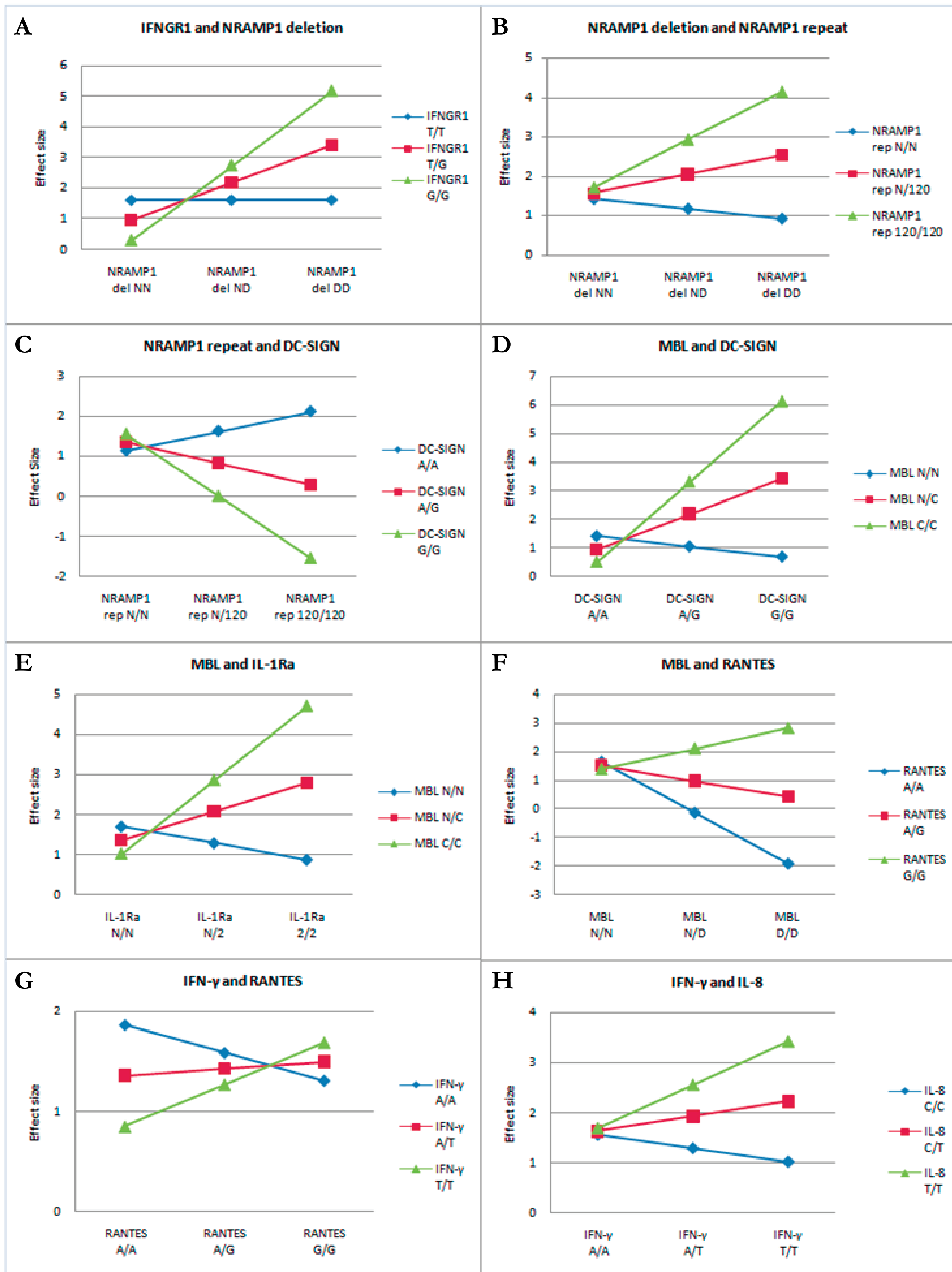


Fig. 4.2 Graphical representation of modelled interactions

The extent of the interaction between *IFN-γ* and *IL-8* on the risk of developing TB was of borderline significance. The absence of the IFN-γ T allele did not have an effect on the odds of getting TB. In the presence of a single *IFN-γ* T allele the risk of TB increased with each *IL-8* T allele, while the *IFN-γ* T/T homozygotes showed a dramatic increase in logit(TB) with each *IL-8* T allele (Fig. 2.4H).

ii) Combined set

Combining all the genetic information available in our database resulted in 106 variants scattered across many genes in 545 individuals. The optimal model which best describes and predicts TB case-control status could not be accurately constructed for the combined data set due to the labour intensive method used for the smaller data set. Computational methods suitable for large data sets could not be used as they impute missing genotypes which are not valid. First, optimal models were constructed separately, whereafter a combined candidate model was created from all the smaller optimal selections to create the final model. However, a number of statistical errors are possible in this kind of analysis (data not shown). Of the 318 statistically significant pair-wise interactions found, 67 were highly significant ($P < 0.01$) (Table 4.6) and 251 were significant ($P < 0.05$) (Fig. 4.3 Appendix A).

4.3.4 Host genotype and bacterial genotype analysis

The host genetic information (variant) with the bacterial strain information / type (trait) combined in a dichotomous genotype-trait association test yielded interesting results (Table 4.7 and Fig. 4.4). When comparing clustered vs non-clustered strains only the G allele of *DC-SIGN* -336 was associated with clustering ($P = 0.034$) (Fig. 4.4A). In the association test between big clusters (defined as five or more individuals harbouring the same strains) and small clusters (defined as less than five patients with the same strains) the D allele of *MBL* and G allele of *RANTES* were significantly associated with cluster size ($P = 0.040$ and $P = 0.0006$ respectively) (Fig. 4.4B). No significant association was found for any of the variants when segregating non-clusters vs big clusters. No significant association was found with either the Beijing family or the LCC family and any gene variant. The T allele of *IFN-γ* was shown to significantly decrease the probability of having a Haarlem strain ($P = 0.044$) and increase the odds of having a LAM strain ($P = 0.012$) (Fig. 4.4C & D). Both *IL-1Ra* allele 3 and *DC-SIGN* -366 allele G were associated with LAM ($P = 0.048$ and $P =$

0.035 respectively) (Fig. 4.4E & F). No statistically significant association with any gene variant was found when comparing the major Beijing strain (F29) (Streicher *et al.*, 2007) with the major LAM strain (F11). The only variant that was associated with the principal genetic groups was the T allele of *IL-8* with PGG 2, compared with PGG 1 and 3 (P = 0.034). Since PGG3 is a small group that consists of only a few individuals, it was decided to compare only PGG1 with PGG2 but no statistically significant association was found with any of the variants. Of all of the significant results obtained with gene-trait association tests, only the association of RANTES – 403 (P = 0.0006) with cluster size survived stringent correction of multiple testing.

Table 4.6. Pairs of variants showing highly significant interactions ($P < 0.01$) in models adjusted for age and gender. The gene name is given followed by the rs number for the SNP and the specific allele for variants with more than 2 alleles.

Gene name_rs number_allele	Gene name_rs number_allele	*P* value	Gene name_rs number_allele	Gene name_rs number_allele	*P* value	Gene name_rs number_allele	Gene name_rs number_allele	*P* value
NRAMP1TGTGdelaN	NRAMP1promCAa118	0.0077	TNFRSF1A_rs767455aT	TLR2_rs5743704aC	0.0077	X5q31_rs11739135aG	IL1RN_rs2419598aT	0.0095
IL.8_.100aC	IFNg874aA	0.0050	APG16L_rs2241880aT	TLR2_rs5743704aC	0.0052	RUNX1_rs2268277aC	CTSZ_rs448943aC	0.0055
CARD15_rs2066844aC	IFNGR1_AP3aT	0.0046	IL1RN_rs2419598aT	IL12B_rs730691aC	0.0038	RUNX1_rs2268277aC	TLR4_rs1927914aC	0.0017
CTSZ_rs13720aC	CARD15_rs2066844aC	0.0082	IL1RN_rs2419598aT	IL18_rs187238aG	0.0093	RUNX1_rs2268277aC	TNF_rs1800629aG	0.0083
IL4_rs2243248aT	IL1Raa1	0.0041	IL1RN_rs2419598aT	MHC2TA_rs3087456aG	0.0024	Wnt5a_rs472631aT	PTPN22_rs2476601aG	0.0090
IL18_rs189667aA	NRAMP1TGTGdelaN	0.0044	TNF_rs1799964aT	TEX264_rs4355273aG	0.0071	Wnt5a_rs815541aC	Socs3_rs4969168aC	0.0061
IL18_rs360718aA	MBPExon1aA	0.0050	CO2region_rs11718165aA	DCSIGN336aA	0.0078	Wnt5a_rs7624718aA	IL10_rs1800890aT	0.0034
IL18_rs360718aA	CTSZ_rs448943aC	0.0062	CO2region_rs11718165aA	SOCS3_rs4969169aC	0.0060	Wnt5a_rs7624718aA	TEX264_rs28994878aC	0.0070
IL18_rs187238aG	NRAMP1TGTGdelaN	0.0033	CO2region_rs11718165aA	TEX264_rs28994878aC	0.0022	Wnt5a_rs7624718aA	INSIG2_rs7566605aG	0.0005
Socs3_rs4969168aC	CTSZ_rs13720aC	0.0004	INSIG2_rs7566605aG	CARD15_rs2066844aC	0.0094	Fzd5_rs3731568aA	IL1Raa1	0.0056
TEX264_rs28994878aC	CTSZ_rs448943aC	0.0077	PADI4_rs2240340aA	SPD_11aT	0.0045	Fzd5_rs3731568aA	IL10_rs1800890aT	0.0075
TLR2_rs5743704aC	IL4_rs2243248aT	0.0009	PPARG_rs2067819aG	CARD15_rs2066847awt	0.0047	Fzd5_rs3731568aA	TNFRSF1A_rs767455aT	0.0094
TLR2_rs5743708aG	MBPExon1aA	0.0085	TNFR2_rs3397aC	MBPExon1aA	0.0053	Fzd5_rs6708488aA	MDR1_rs1045642aC	0.0002
TLR4_rs1927914aC	IFNg874aA	0.0077	TNFR2_rs3397aC	IL12B_rs3212227aA	0.0098	Fzd5_rs6708488aA	MDR1_rs2032582aG	0.0000
TLR4_rs1927914aC	IL10_rs1800890aT	0.0020	GP130_rs3730293aT	NRAMP1promCAa118	0.0014	Fzd5_rs6708488aA	Wnt5a_rs9311564aA	0.0008
TLR4_rs1927914aC	CCL2_rs3760399aT	0.0037	MS4A2_rs569108aT	MBPExon1aA	0.0053	Fzd5_rs718290aT	TNFRSF1A_rs767455aT	0.0098
TLR4_rs4986791aT	IFNGR1_AP3aT	0.0035	MS4A2_rs569108aT	NRAMP1TGTGdelaN	0.0012	TLR2_GTa18	MS4A2_rs569108aT	0.0095
TLR4_rs11536889aG	IL10_rs1800890aT	0.0098	MS4A2_rs569108aT	BTNL2_rs2076530aA	0.0042	D5S2941a8	CTSZ_rs448943aC	0.0090
TLR4_rs11536891aT	IL12RB2_rs11576006aT	0.0082	IL6ST_rs3729961aC	NRAMP1promCAa118	0.0027	sp11Exon11delaGAAG	IL10_rs1800890aT	0.0072
TLR4_rs11536891aT	IL10_rs1800890aT	0.0095	IL23R_rs11209026aG	BTNL2_rs3817974aA	0.0089	sp110_rs3948463aG	PTGER4_rs1992660aA	0.0010
TLR4_rs11536891aT	TLR2_rs3804100aT	0.0069	IL23R_rs11209026aG	CO2region_rs11718165aA	0.0077	sp110_rs3948463aG	IL1RN_rs2419598aT	0.0016
MDR1_rs3789243aC	IL12RB2_rs375947aA	0.0079	TNFSF15_rs7848647aC	MS4A2_rs569108aT	0.0014			
PTGER4_rs1992660aA	Socs3_rs4969168aC	0.0063	PTPN22_rs2476601aG	NELL1_rs1607616aT	0.0077			

Table 4.7. *P*-values for dichotomous trait-genotype (allele count) association tests, adjusted for age.

Variant	Clustered vs non clustered	Big clusters vs small clusters	Non clusters vs big clusters	Beijing vs non-Beijing	Haarlem vs non-Haarlem	LAM vs non-LAM	LCC vs non-LCC	Beijing (F29) vs LAM (F11)	PGG1 vs others	PGG2 vs others	PGG3 vs others	PGG1 vs PGG2
MBP/MBL Exon1nB	0.5081	0.2146	0.3129	0.3307	0.3077	0.7925	0.4976	0.6131	0.1694	0.2621	0.7210	0.1801
MBP/MBLExon1nC	0.3095	0.8950	0.4376	0.7859	0.4032	0.9856	0.6128	0.5007	0.6680	0.5692	0.7559	0.6315
MBP/MBLExon1nD	0.9877	**0.0401**	0.9882	0.4222	0.6199	0.4303	0.9733	0.9922	0.5342	0.7987	0.9926	0.5989
NRAMP1promCAn120	0.3055	0.7930	0.3753	0.7270	0.4692	0.7083	0.1807	0.9072	0.9681	0.9972	0.9631	0.9776
NRAMP1promCAn122	0.9866	0.9846	0.9908	0.0842	0.9917	0.9859	0.9866	0.9879	0.9840	0.9844	0.9919	0.9841
NRAMP1TGTGdelnD	0.8272	0.5757	0.9074	0.2713	0.1640	0.6899	0.8869	0.2766	0.3530	0.2723	0.6432	0.3087
IFNGR1_AP3nG	0.813	0.3380	0.7719	0.8800	0.6236	0.8365	0.5090	0.9820	0.7713	0.4583	0.0926	0.9817
IFN-γ874nT	0.7938	0.5465	0.9375	0.4498	**0.0443**	**0.0123**	0.1674	0.1642	0.1923	0.0657	0.2473	0.1347
IL-1Ran4	0.8754	0.4822	0.7983	0.3587	0.3848	0.1938	0.9886	0.1339	0.2696	0.1630	0.9931	0.2362
IL-1Ran3	0.7342	0.5332	0.5581	0.9857	0.9912	**0.0485**	0.6416	0.9889	0.5646	0.3605	0.9919	0.5032
IL-1Ran5	0.968	0.8534	0.9619	0.3540	0.9892	0.6929	0.9882	0.6474	0.4727	0.6936	0.9929	0.5281
SPD_11n2	0.3671	0.4695	0.3157	0.3765	0.3992	0.5388	0.9651	0.6637	0.7547	0.3381	0.2063	0.6178
IL.8_.100nT	0.7251	0.8817	0.7222	0.3146	0.5507	0.0945	0.3208	0.5937	0.1489	**0.0341**	0.1633	0.1011
RANTES_.403nG	0.3915	**0.0006**	0.3669	0.1172	0.5612	0.2906	0.7060	0.2129	0.2257	0.6555	0.2354	0.3016
DCSIGN871nG	0.1612	0.8227	0.2499	0.7452	0.3304	0.3904	0.6399	0.7327	0.8861	0.7638	0.7760	0.8455
DCSIGN336nG	**0.0347**	0.7906	0.1622	0.9219	0.6607	**0.0351**	0.0600	0.8950	0.9484	0.7686	0.6900	0.8955

Significant interactions are highlighted in green ($P < 0.05$) and in red ($P < 0.001$).

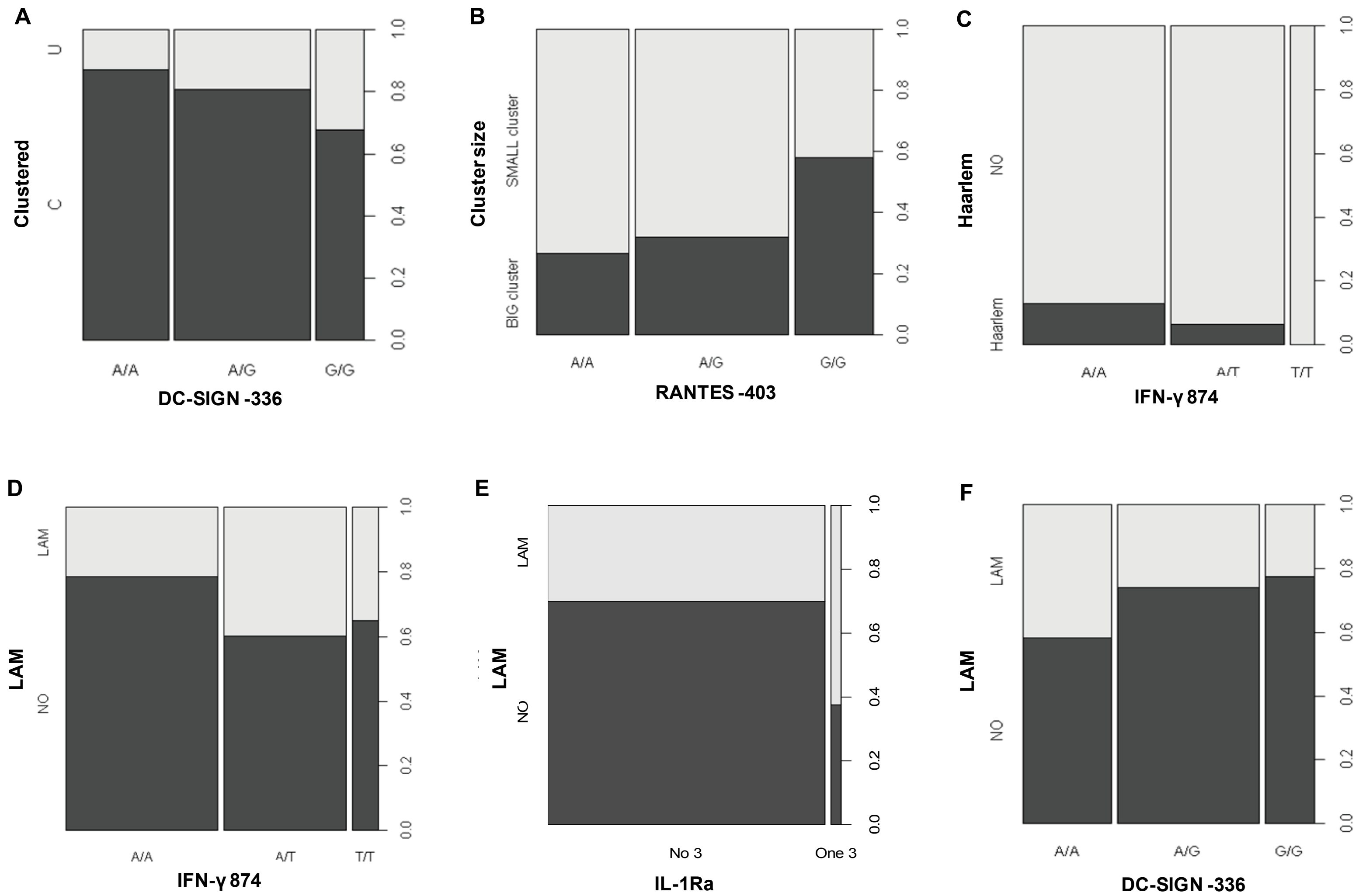


Fig. 4.4. Graphical representations of significant trait-genotype associations

4.4 DISCUSSION

The condensed group of 11 variants involved in TB susceptibility genotyped here enabled us to test the theory that disease outcome may be due to the interaction of several gene effects. With only four associations found between the individual variants and TB susceptibility, but eight instances of statistically significant gene-gene interactions, the importance of epistasis is clearly identifiable in this study.

One of the eight interactions were between two variants in the *NRAMP1* gene, namely the 5'$(GT)_n$ repeat and the 3' UTR (TGTG) deletion. Both of these have previously been associated with TB independently (Table 1) and were shown to have separate main effects at the 5' and the 3' together with dominance effects at the one end but not the other (Bellamy *et al.*, 1998). From this it can be deduced that there are separate functional polymorphisms in linkage disequilibrium (LD) with these two markers that play an important role in susceptibility to TB.

The interaction observed between *NRAMP1* and *IFNGR1* can be explained by their mutual involvement in macrophage activation. Functional IFNGR1 is essential for containing *M. tuberculosis* (Dorman *et al.*, 2004) and suggests a critical role for IFN-y and or its receptors. NRAMP1 is essential in macrophage activation as it activates microbicidal responses in the infected macrophage and therefore plays an important role in the early innate response to mycobacterial infection (Li *et al.*, 2006). NRAMP1 also enhances the pro-inflammatory responses which are important in the containment of mycobacterial infection. Macrophage activation is greatly enhanced by IFN-γ and IFN-γ response elements that are present in the human *NRAMP1* promoter region (Searle & Blackwell, 1999), providing further evidence for their interaction.

The interactions observed between *IFN-γ* and both *IL-8* and *RANTES* are likely to be due to their pro-inflammatory nature. IFN-γ is a Th1 type cytokine that plays a crucial role in the initial protective immune response against an infection, including *M. tuberculosis* (Vidyarani *et al.*, 2006). It activates macrophages and plays a vital role in antimicrobial protection, as demonstrated in a knockout mouse model where an increase in susceptibility to tuberculosis could be seen once the gene was knocked out (Cooper *et al.*, 1993). IL-8, a

neutrophil activating peptide, is an important chemokine in the human inflammatory process and functions as a potent chemoattractant for the recruitment of leukocytes to inflammatory sites (Ma *et al.*, 2003). IL-8 is a tissue-derived peptide that mediates the activation and migration of neutrophils and T lymphocytes into tissue from the peripheral blood. RANTES, also known as CCL5, is a ligand for chemokine receptor 5 and recruits several types of inflammatory cells including eosinophils, monocytes and T lymphocytes to sites of inflammation. The recruitment of cells is vital for granuloma formation and RANTES recruits infected macrophages more effectively than uninfected macrophages (Skwor *et al.*, 2006).

MBL and DC-SIGN are both C-type lectins and function in uptake and signalling in the pulmonary innate immune system (Ji *et al.*, 2005). Microbial carbohydrate structures are recognized by pathogen associated molecular pattern (PAMP) receptors such as MBL and DC-SIGN (Buzás *et al.*, 2006). MBL plays an important role in the modulation of inflammation as it binds to the pathogen to initiate the lectin pathway acting as the first line of defence against microbes (Holmskov *et al.,* 2003; Turner *et al.*, 2003). Phagocytes represent the first cellular defence in the alveoli, the surface of which is rich in C-type lectin pattern recognition receptors such as DC-SIGN. On the basis that DC-SIGN might mediate intracellular signalling events leading to cytokine secretion, it has previously been proposed that this C-type lectin could be used by pathogens like *M. tuberculosis* to their own advantage as part of an immune strategy (Geijtenbeek *et al.,* 2003; Tailleux *et al.*, 2003). Bearing in mind that these two C-type lectins, MBL and DC-SIGN, have complementary functions in the lectin pathway, a functional interaction seems likely.

The nature of two of the eight interactions (*MBL : ILRa* and *NRAMP1 : DC-SIGN*) for TB susceptibility is not clear and might have been detected by chance as it is difficult to find simple biological explanations for the interactions observed. However, MBL plays a very important role in the innate immune response by modulating inflammation and IL-1 and therefore IL-1Ra is responsible for a cascade of events leading to a potent pro-inflammatory immune response (Dinarello, 1988). The mechanism by which MBL modulates inflammation is not clear but it has been proposed that it exerts a complex effect on cytokine release by monocytes (Jack *et al.*, 2001). A previous study showed that monocyte release of certain cytokines, such as IL-1, was enhanced at MBL concentration

below 4µg/ml but reduced at higher concentrations, suggesting that both the *MBL* genotype and the presence or absence of an on-going acute phase response will help to determine the nature of such cytokine responses in individuals (Turner, 2003). Given the interrelationships and complexity of the innate immune system, and the fact that different pathways frequently form part of the same network (Delves & Roitt, 2000; Raza *et al.*, 2008) these interactions could well indicate functionality in the same network known as downstream interactions.

The combined data set consisting of 545 individuals and 106 variants yielded a vast number of pair-wise interactions. Of the 318 statistically significant interactions ($P < 0.05$) found, 67 were highly significant ($P < 0.001$). This large number of interactions found could be explained by the fact that all of the selected genes were well-characterised candidate genes for susceptibility to TB based on their biological function and proposed role in the multistep course of this complex disease. However, we have to consider the fact that some of these interactions might be spurious due to the large number of data points analysed. All of these interactions need to be thoroughly investigated to assess the plausibility and importance of each pair wise interaction separately. The analysis of this immense combined data set showed once again how complex and integrated human biological systems are, making it difficult at this point to be fully understood. This gene-gene pair-wise interaction analysis is a multi gene approach that may be the future for host genetic susceptibility studies of TB as the effect of one gene might not be revealed if the effect of another gene is not considered.

From our bacterial strain data, it can be seen that LAM is the most prevalent family of strains, however the frequency of occurrence of Beijing family of strains is on the increase. Individuals belonging to the Beijing family of strains can be grouped together in a small group based on their strain information and relative small strain differences whereas in the LAM family, individuals are much more distant from each other as it is a broader family. The LCC and Haarlem strain families are very small and therefore play a less important role in the TB epidemic.

It has been hypothesized that the interaction between the genotype of the human host, and the bacterial strain genotype, could influence the susceptibility of the host in terms of both

the progression to disease and perhaps the type of disease seen. These interactions were tested here by means of a dichotomous genotype-trait association test. Most of the interactions were slightly statistically significant except for *RANTES* -403 that was strongly associated with cluster size (*P* = 0.0006). The odds of being part of a small cluster decreases highly significantly with each G allele at this locus, i.e. each G allele halves the likelihood of an individual being more susceptible to a unique strain that will form a small cluster. It is difficult to draw any reliable conclusions from the weak associations found as most of them will probably disappear after multiple corrections. The low numbers in some of the bacterial traits as well as the human genotype categories such as the Haarlem strain (*n* = 21) and the *MBL* D allele (*n* =6) might have contributed to the weak associations found.

Since several polymorphisms have been associated with TB susceptibility, our main focus was to test how polymorphisms in multiple genes acting simultaneously, affect the immune function and therefore susceptibility to TB which is a result of multiple genetic and environmental factors. The effect of one gene might not be revealed if the effect of another gene is not considered and therefore, focusing on single polymorphisms at a time may generate false negatives. This is the first time, to our knowledge, that interactions between these genes have been shown. However there is a potential limitation to this study, namely spurious associations, particularly where the mechanism of interaction of the genes involved is not immediately clear or intuitive. The genes studied above were all well-known and well-characterised in the search for susceptibility genes to TB. Considering the amount of data rapidly becoming available from SNP chip studies, similar interaction studies between known susceptibility genes and others could well identify new candidates involved in TB susceptibility. Methods for studying gene-gene interactions are based on a multilocus and multi gene approach, consistent with the nature of complex-trait diseases, and may provide the paradigm for future genetic studies of TB.

4.5 ACKNOWLEDGEMENTS

I am grateful to Prof. Lize van der Merwe for her help with the statistical analysis.

CHAPTER 5

MEASUREMENT OF CYTOKINES BY MULTIPLEX ANALYSIS AMONG CHILDREN FROM A HIGH TB INCIDENCE AREA

5.1 INTRODUCTION

Due to the complexity of the immune system and the multifactorial nature of a complex disease such as TB, it has become increasingly evident that unravelling the underlying mechanisms of the genetic component may lead to new insights into the pathogenesis of TB and the development of new treatment strategies (Selvaraj *et al.*, 2008).

Cytokines (also known as immunomodulating agents) such as interferons, interleukins, tumor necrosis factor and lymphokines as well as chemokines are small regulatory proteins that are released by cells to regulate the immune response by triggering inflammation. The management of several components of the host immune system is required to respond successfully to an invading pathogen such as *M. tuberculosis* (Frankenstein *et al.*, 2006). Depending on the type and amount of cytokine and chemokine secretion, activated at the recognition of a pathogen, the host defensive mechanisms can be stimulated or over-stimulated leading to tissue injury, fever or cachexia (Zganiacz *et al.*, 2004). The interactions between lymphocytes, antigen-presenting cells and the cytokines secreted by them are therefore responsible for regulating the immune response to TB. A simplified version of some of the effects following *M. tuberculosis* infection can be seen in Fig. 5.1.

It is well known that owing to various polymorphisms in several cytokine genes, the protein levels of the main modulators of the immune system, cytokines and chemokines, are changed in several diseases such as cancer (O'Hayre *et al.*, 2008), cardiovascular diseases (Aukrust *et al.*, 2007), inflammatory diseases (Szekanecz & Koch, 2007) and infectious diseases (García-Zepeda *et al.,* 2007; Jankovic *et al.*, 2001) including susceptibility or resistance to TB (Cooke *et al.,* 2006; Henao *et al.*, 2006; Lopez-Maderuelo *et al.*, 2003; Ma *et al.*, 2003; Rossouw *et al.*, 2003; Vidyarani *et al.*, 2006). The functional polymorphisms or haplotype patterns in some of these cytokine genes might be vital for protective immune responses and may serve as biomarkers of protection or susceptibility to TB (Selvaraj *et al.*, 2008).

The discovery of biomarkers for use as surrogates for the response to TB treatment may play a crucial role in monitoring efficiency in both the current TB drugs as well as clinical trials for new TB drugs. There are three possible classes of biomarkers: (1) markers for relapse, that may reduce the duration of clinical trials, (2) markers to assess early treatment

that may replace sputum cultures and (3) markers for baseline differences that can lead to the identification of patients who require specialised treatment (Brahmbhatt *et al.,* 2006; Siawaya *et al.*, 2008). Cytokines represent once candidate class for promising biomarkers.

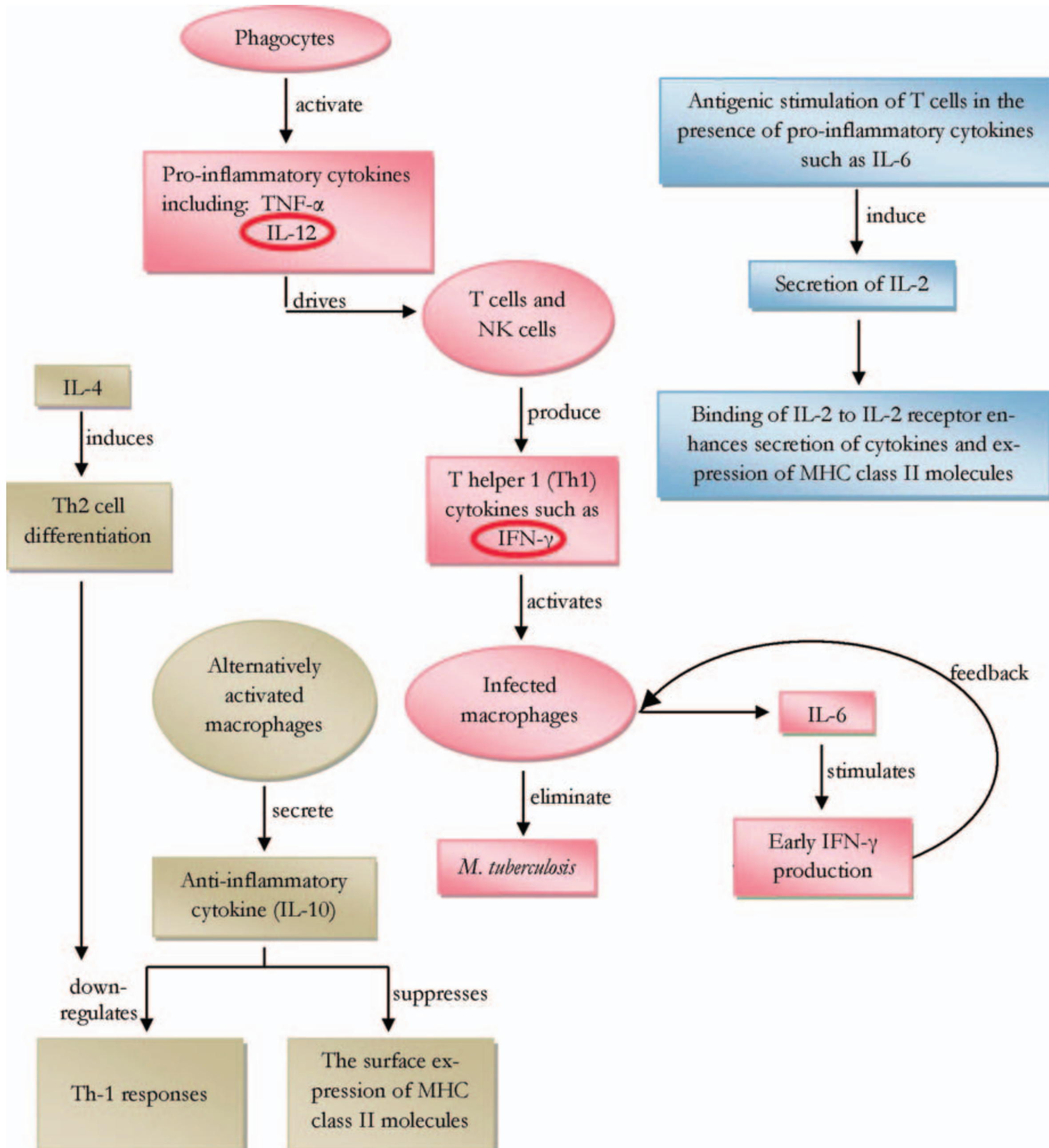


Fig. 5.1 A simplified version of some downstream effects after infection with *M. tuberculosis*

Therefore measurements of these modulators / cytokines can give new insights into the disease (more specifically the mechanisms underlying it) leading to enhanced diagnosis and treatment. (Colobran *et al.,* 2007; McInnes & Schett, 2007). To accurately predict which single modulator will give the best possible explanation as to what the underlying mechanism of the disease is, is near impossible, therefore, in my opinion, measuring multiple immunomodulators simultaneously is a way forward.

There are several techniques available to measure immunomodulators. The most widely used technique still remains the enzyme-linked immunosorbant assay (ELISA). However this technique is limited to measuring only one immunomodulator at a time, making it labour intensive and time consuming. Due to the complex network of immunomodulators and possible interactions that might exist, measuring only one modulator at a time will not be efficient in assessing its biological relevance. Therefore new multiplex technologies have been developed to measure a large number of cytokines simultaneously in one reaction. The two major advantages of this technique are: 1) a low volume of sample is required to simultaneously detect many immunomodulators of interest and 2) time is saved as many measurements are done at once.

It is now possible to look at the relationship between secreted cytokines / proteins and the genotype information of the individuals opening a new door to identification of response markers. The present study investigated 18 cytokines including pro-inflammatory, anti-inflammatory and chemokine factors in healthy (mantoux positive or negative) children using the well described and commonly used Linco-plex immunoassay (LINCO Research).

5.2 MATERIALS AND METHODS

5.2.1 Study cohort

All the subjects used in this study originated from the Ravensmead and Uitsig suburbs near Tygerberg Hospital, Parow, Western Cape, South Africa where the incidence of TB has been reported as ± 1000 per 100 000 population, but with a low HIV prevalence. A total of 340 healthy HIV negative children were collected from the Ravensmead-Uitsig community and used in the Lincoplex assay. All participants in this study or their legal guardians provided written informed consent for participation in the study. The Ethics Review Committee of the Faculty of Health Sciences, University of Stellenbosch, South

Africa approved this investigation (95/072). The Mantoux skin test was carried out on all participants using *M. tuberculosis* PPD. The Mantoux skin reactivity was read between 48-72 h after the test. Results were available for 309 (90.8%) of the individuals. The sex distribution was approximately equal, with 49.5% males and 50.5% females, with a median age of 14 years.

5.2.2 Sample collection and antigen stimulation

10 ml of blood was drawn from all the participants at the community centre and kept at 37°C in a portable incubator while it was transported to the lab in 1-2 hours. Heparinised (anti-coagulant factor) whole blood was diluted 1 in 10 with serum-free medium (RPMI, Gibco, Laboratory Scientific Services). Cells were stimulated in quadruplicate with two antigens, *M. bovis* bacillus Calmette-Guerin (BCG) and early secreted antigen -6 (ESAT-6). The negative control was without any antigen stimulation. Cell cultures were incubated at 37°C with 5% CO_2,whereafter the supernatants were harvested at day 3 and 7. Since a very high correlation was found between day 3 and day 7 with the ELISA results across both antigens (BCG and ESAT-6) and the percentage of individuals responding to antigen was higher at day 7, only day 7 samples were used for the Linco-plex immunoassay (see section 5.3.6).

5.2.3 Measurement of cytokines using the Linco-plex immunoassay

The human cytokine 18-plex Linco-plex assays were done according to manufacturer's instructions using a 96-well filter plate and a vacuum pump for the washing steps. All other reagents as well as the antibody-immobilized beads were supplied with the kits (LINCO Research). A standard curve was used in each assay ranging from 3.2 pg/ml to 10 000pg/ml. After blocking the filter plate with assay buffer, an appropriate matrix diluent, in this case RPMI with glutamine, was added to all the wells. The 18 antibody-immobilized beads were sonicated separately, whereafter they were mixed together according to manufacturer's instructions. This was followed by adding the samples in the appropriate wells as well as the mixed beads. After 1 hour incubation on a shaker, the fluid was removed by a vacuum pump followed by two wash steps with wash buffer. The biotinylated detection antibody cocktail was added into each well and left to incubate on a shaker for 30 min. at room temperature whereafter the streptavidin-phycoerythrin was added to each well. This was left to incubate on the shaker for 30 min. at room

temperature, whereafter the contents were gently removed by a vacuum pump. After another two wash steps sheath fluid was added to all the wells and put on a shaker for 5 min to allow the beads to resuspend. After this the plate was placed in the Luminex instrument where the beads were analyzed immediately on a Bio-plex array reader.

Quality controls were provided with each kit to measure the reliability and precision of the selected cytokines. All 18 cytokines measured fell into the manufacturer's expected range (Table 5.1). A stringent quality control, the supernatant of a 7 day whole blood assay (WBA) from a volunteer, was aliquoted out, stored at -80°C and used on every plate for each run.

Table. 5.1 Cytokines included in the 18-plex Lincoplex assay

Pro-inflammatory and Th1 type cytokines			Anti-inflammatory cytokines
IL-1α	TNF-α	IFN-γ	IL-4
IL-7	IL-17	IP-10	IL-10
IL-8	sCD40L	MCP-1	IL-13
IL-12p40	GM-CSF	MIP-1α	IL-1Ra
IL-12p70	RANTES		

5.2.4 Statistical analysis

Spearman correlation tests were used to correlate variables (Mantoux, age, BCG and ESAT-6) and cytokines. A one-way analysis of variants was used for the Mantoux reading 0 mm vs Mantoux reading > 0 mm. Parallel analysis was used to compute the factor loadings (O'Connor, 2000). The program STATISTICA version 8.0 (www.statsoft.com) was used for all the analyses.

5.3 RESULTS

5.3.1 Mantoux results

The distribution of the Mantoux readings (mm) can be seen in Fig. 5.2. In our data there were two major peaks, one at 0 mm and one at 15 mm. According to the international standards any TST measurement < 5 mm is scored as negative and between 5 mm and 10 mm is regarded as intermediate, indicating that individuals have been exposed to the pathogen. Any Mantoux reading ≥ 10 mm is regarded as a positive TST. Of the total study group, 123 had no measurable Mantoux reaction (0 mm) whereas 17 had a reaction

of ≥ 5 mm, 61 ≥ 10 mm and 106 ≥ 15 mm. It has been suggested that one would expect a peak between 5 mm and 10 mm as a result of the BCG vaccination (Mahomed *et al.,* 2006; Pai *et al.*, 2005). However we did not see this in our study group nor did we see a peak between 0 mm and 5 mm. Our two main peaks in this bimodal distribution were at 0 mm, indicating that they had no measurable anit-mycobacterial immune response, and at 15 mm, indicating that they have a significant immune response.

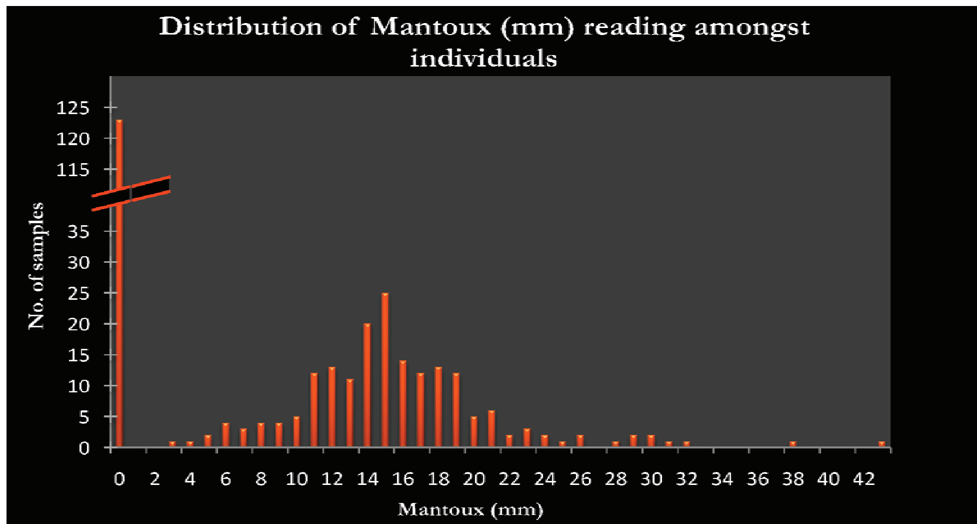


Fig. 5.2 The distribution of the Mantoux readings for all the study participants

A Spearman correlation test was performed between all the cytokine production levels [pg/ml] and the Mantoux readings > 0 mm. Three separate correlation tests for BCG stimulated, ESAT-6 stimulated and unstimulated samples were done for each cytokine. The results for the main Th1-type cytokine IFN-γ were as follows (Fig. 5.3), the BCG stimulated group (r = 0.18, *P* = 0.55), the ESAT-6 stimulated group (r = 0.28, *P* < 0.00) and the unstimulated group (r = -0.04, *P* = 0.61). All the other cytokines gave similar results. A correlation value of r ≥ 0.80 is considered to be a strong correlation; however some of the *P*-values were statistically significant in our study even though the correlation values differed only slightly from zero. Therefore the few significant *P* values obtained in our study do not indicate a good correlation, but merely a trend that might be caused by other factors. The *P*-values in these graphs are an indication of the degree of certainty that we have that the correlation coefficient (r) is as stated. No major correlations were found between any of the 18 cytokines and the Mantoux readings. From this it can be deduced

that the correlation between the production of these specific cytokines and the Mantoux readings are not well correlated.

The data was divided into two groups based on their Mantoux response, viz. 0 mm and > 0 mm. A correlation test was performed between all 18 cytokine production levels [pg/ml] and these two groups. Three separate correlation tests for BCG stimulated, ESAT-6 stimulated and unstimulated samples were done for each cytokine. Focusing on the Th1-type cytokines, we found increased cytokine production levels in the group with a Mantoux measurement of > 0 mm compared to the group with 0 mm. The response of two Th1-type cytokines, IFN-γ and IL-8, can be seen in Fig. 5.4A and B respectively. This difference was statistically significant in the BCG-stimulated ($P \leq 0.01$) and ESAT-6 stimulated groups ($P \leq 0.01$), but as expected not in the unstimulated group ($P = 0.33$).

The Th2-type cytokines showed a slight decrease in cytokine production levels in the group that had a Mantoux measurement of > 0 mm compared to the group that had a measurement of 0 mm. The response of the main Th2-type cytokine, IL-4, can be seen in Fig. 5.5. This difference was not statistically significant in the BCG-stimulated ($P = 0.55$), ESAT-6 stimulated ($P = 0.49$), or in the unstimulated group ($P = 0.44$). IL-4 was detected at very low levels throughout the Lincoplex assay and considerably lower compared to that of IFN-γ. The low detection level of IL-4 can be due to either low sensitivity of the assay to detect IL-4 or due to a low production of IL-4 amongst the participants. There was therefore no evidence for differences in IL-4 levels and a Mantoux positive or negative reaction.

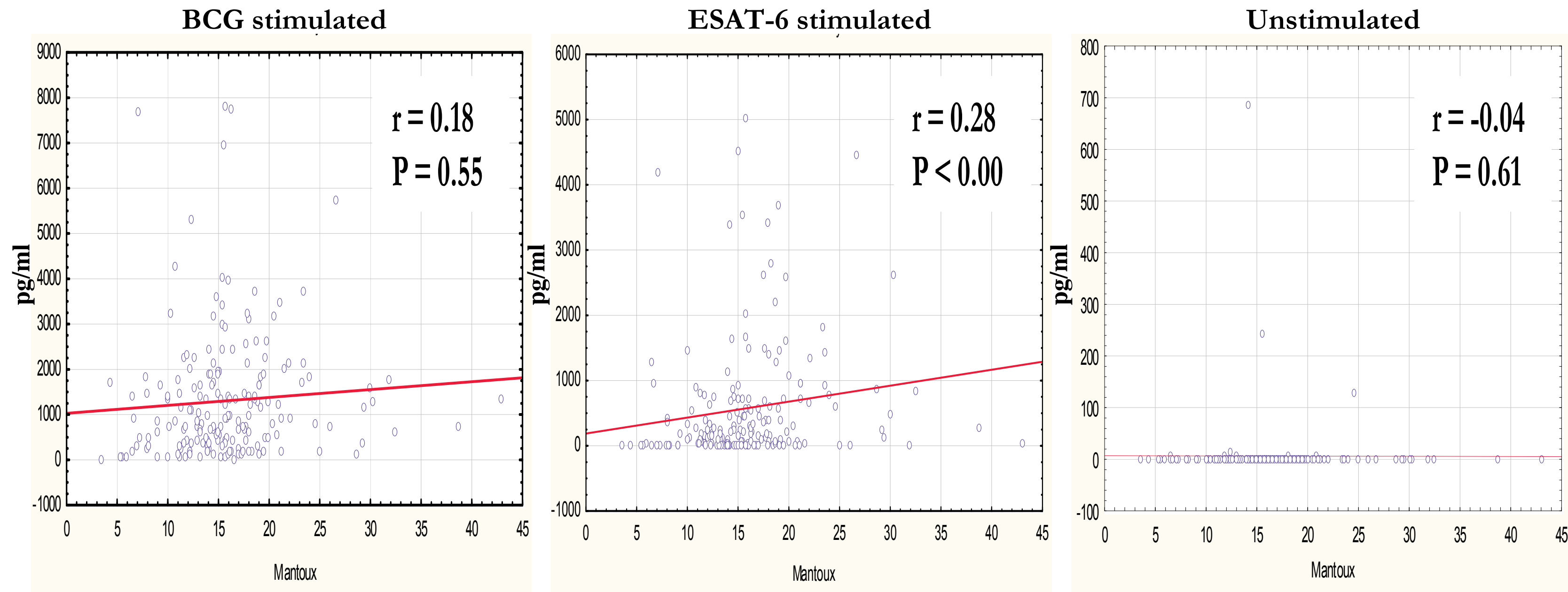


Fig. 5.3 Correlation between Mantoux readings > 0 mm and IFN-γ production [pg/ml] for BCG stimulated samples, ESAT-6 stimulated samples and unstimulated samples. r = Spearman correlation.

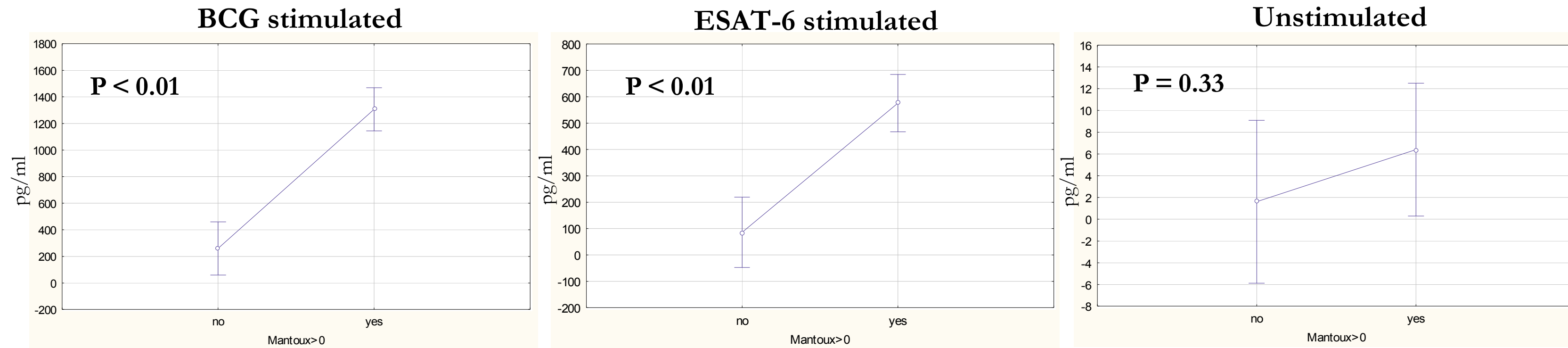


Fig. 5.4 A IFN-γ production [pg/ml] in individuals with Mantoux reading 0 mm vs Mantoux reading > 0 mm for BCG stimulated samples, ESAT-6 stimulated samples and unstimulated samples. The vertical bars denote 0.95 confidence intervals.

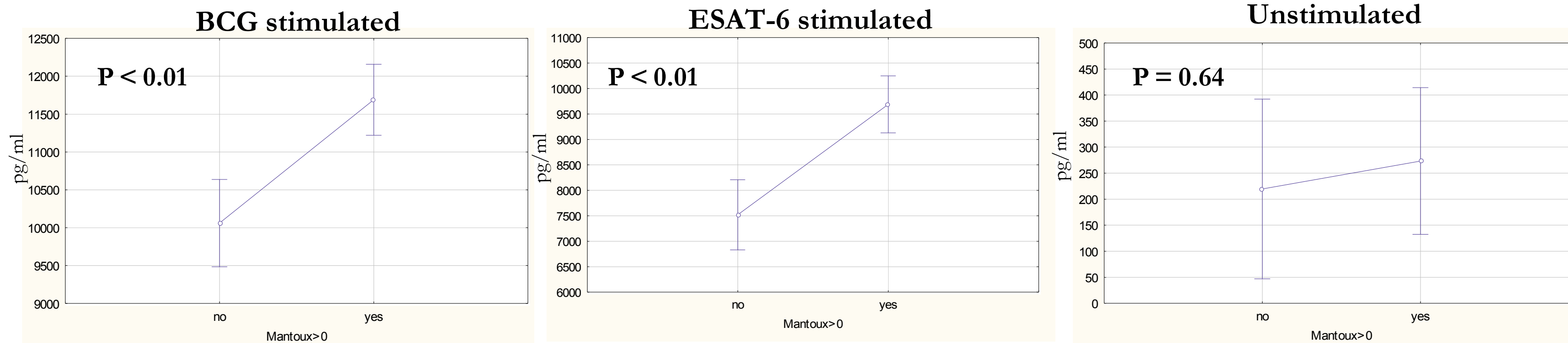


Fig. 5.4 B IL-8 production [pg/ml] in individuals with Mantoux reading 0 mm vs Mantoux reading > 0 mm for BCG stimulated samples, ESAT-6 stimulated samples and unstimulated samples. The vertical bars denote 0.95 confidence intervals.

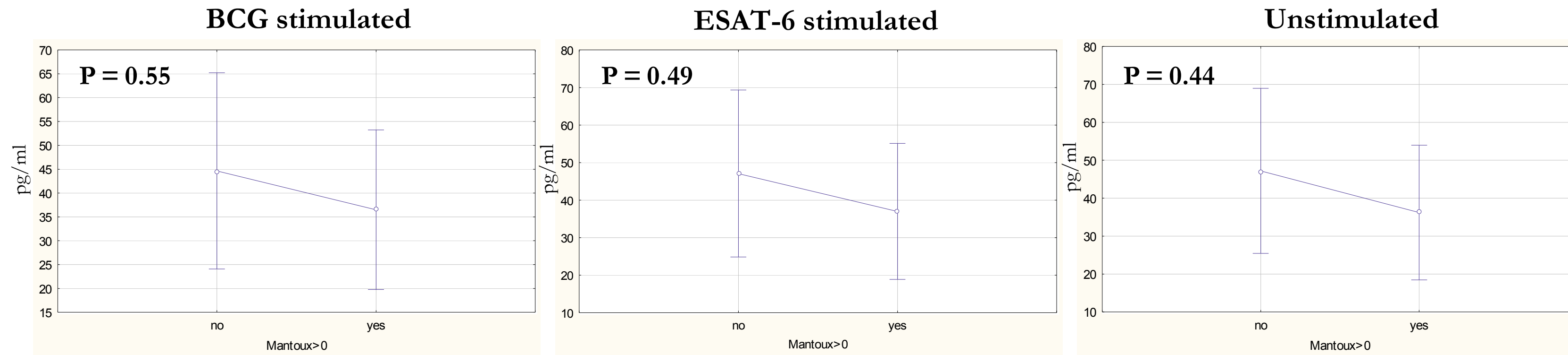


Fig. 5.5 IL-4 production [pg/ml] with Mantoux reading 0 mm vs Mantoux reading > 0 mm for BCG stimulated samples, ESAT-6 stimulated samples and unstimulated samples. The vertical bars denote 0.95 confidence intervals.

5.3.2 Age

The median age of the study group was 14 years, ranging from 2 months to 40 years. A Spearman correlation test was performed between all the cytokine production levels [pg/ml] and age. Three separate correlation tests for BCG stimulated, ESAT-6 stimulated and unstimulated samples were done for each cytokine. In this study, no major correlations were found between any of the 18 cytokines and age. As explained in section 5.3.1, some of the *P*- values were statistically significant even though the correlation values differed only slightly from zero. As the results were more or less the same for all 18 cytokines, I will discuss only the main Th1- and Th2-type cytokines i.e. IFN-γ and IL-4 respectively (Fig. 5.6 A & B).

The correlation between IFN-γ and age was statistically significant in the BCG-stimulated ($P < 0.00$) and ESAT-6 stimulated groups ($P < 0.00$), but not in the unstimulated group ($P = 0.49$) (Fig. 5.6A). Focusing on IL-4 we found that the correlation was statistically significant in the BCG- and ESAT-6-stimulated groups and in the unstimulated group (all three $P < 0.00$) (Fig. 5.6B). It is difficult to draw conclusions from the IL-4 cytokine production levels, as these were very low throughout the Lincoplex assay (see section 5.3.1). The low levels of IL-4 relative to IFN-γ are clearly evident in Fig. 5.6.

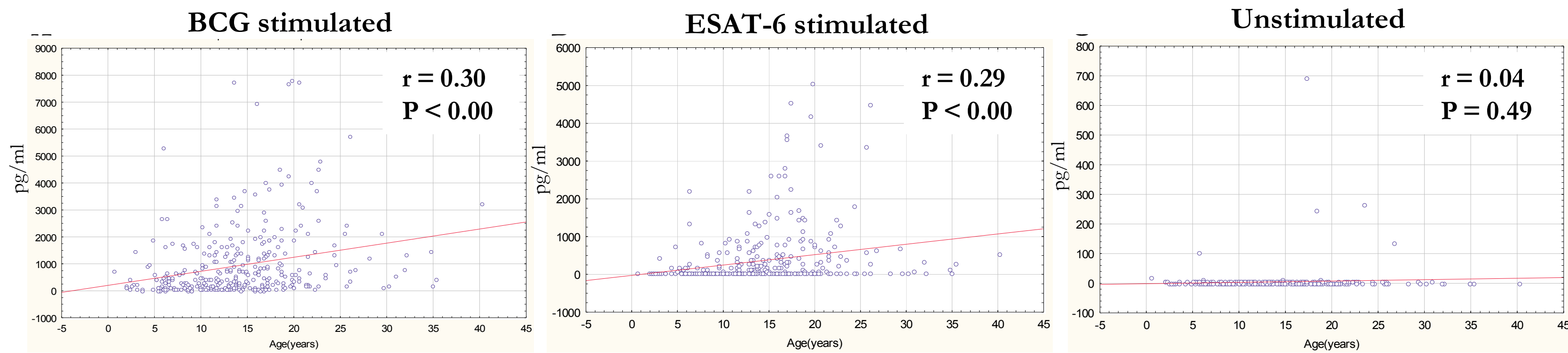


Fig. 5.6 A Correlation between IFN-γ production [pg/ml] and age for BCG stimulated samples, ESAT-6 stimulated samples and unstimulated samples.

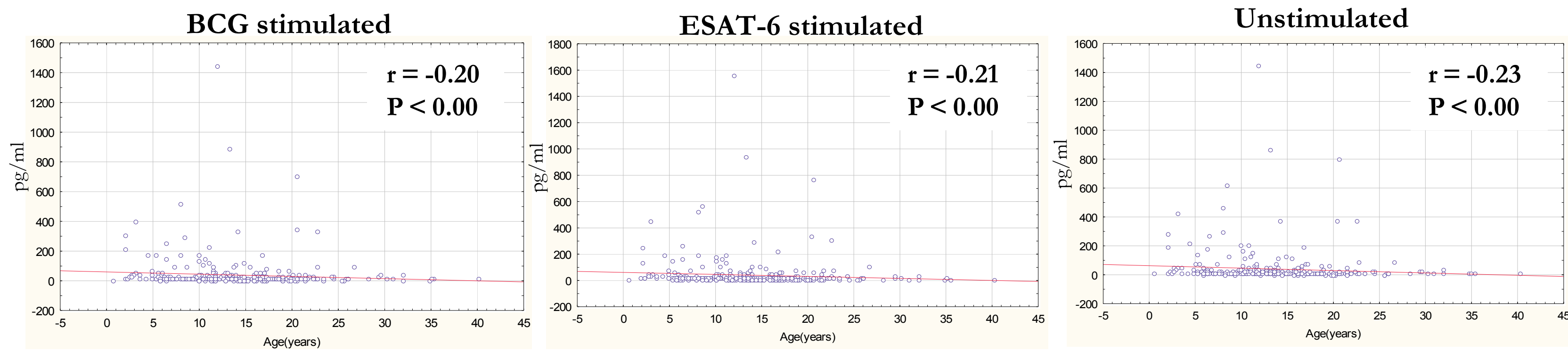


Fig. 5.6 B Correlation between IL-4 production [pg/ml] and age for BCG stimulated samples, ESAT-6 stimulated samples and unstimulated samples.

5.3.3 Correlation between BCG and ESAT-6 for all the cytokines

A Spearman correlation test was performed between the production levels [pg/ml] of all 18 cytokines for both the BCG stimulated and ESAT-6 stimulated samples. The correlations were statistically significant for all 18 cytokines. Some of these correlations can be seen in Fig. 5.7 [A] for IFN-γ, [B] for IL-8, [C] for IL-1Ra and [D] for IL-4. The IFN-γ response was higher in the BCG stimulated samples than the ESAT-6 stimulated samples. The IL-8 production was high in both BCG and ESAT-6 stimulated samples. A strong correlation was found between BCG and ESAT-6 stimulated samples for both IL-1Ra ($r = 0.83$) and IL-4 ($r = 0.90$). This indicates that BCG induces IL-1Ra and IL-4 production to the same level as ESAT-6.

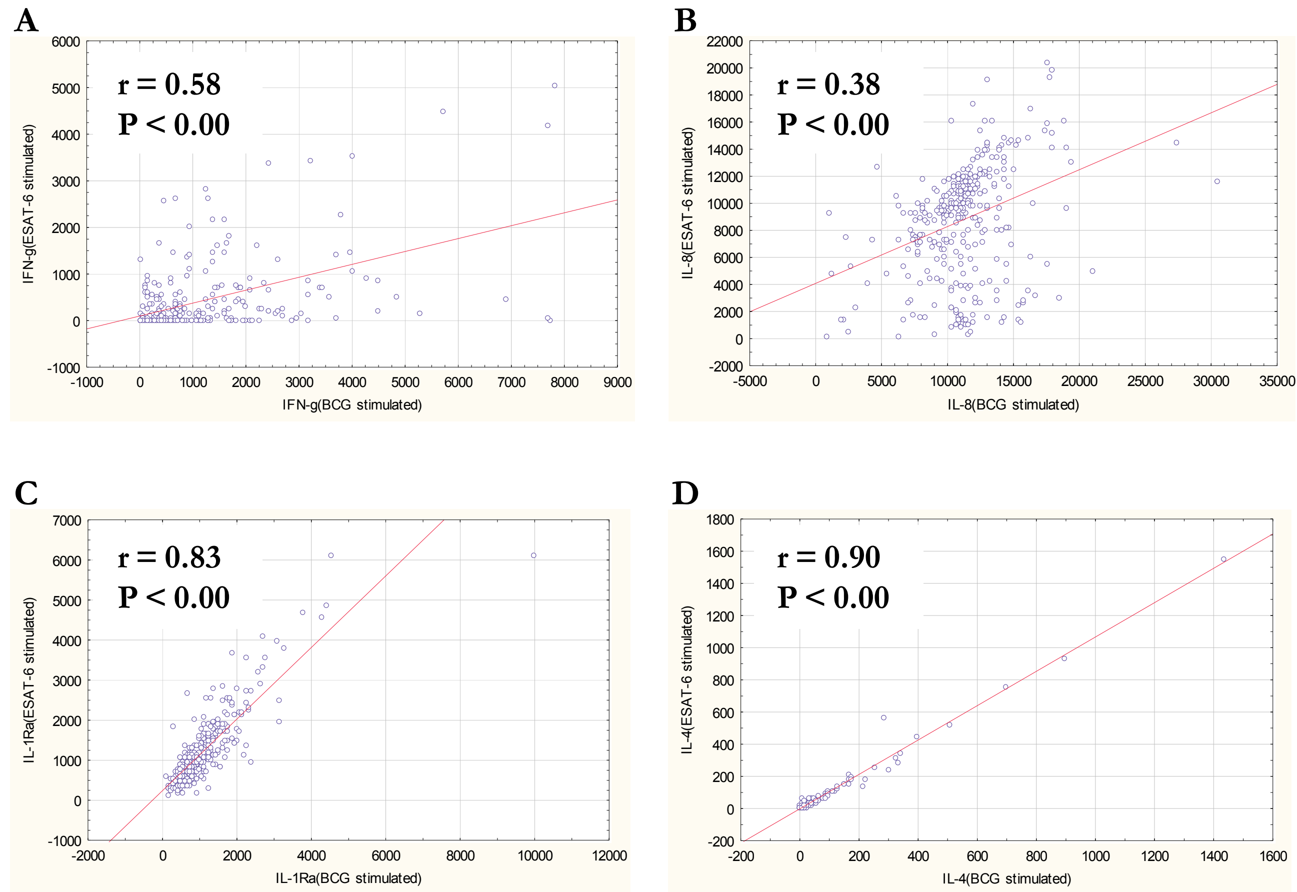


Fig. 5.7 Correlation between BCG and ESAT-6 stimulated responses for [A] IFN-γ, [B] IL-8, [C] IL-1Ra and [D] IL-4

5.3.4 Comparison of cytokine production levels

A correlation test was performed between antigen-specific cytokine production levels of all 18 cytokines. The highest statistically significant correlations that were found are summarised in Fig. 5.8. The full set of results is represented in a correlation matrix (Fig. 5.9). We found strong correlations between IFN-γ and the granulocyte-macrophage colony-stimulating factor (GM-CSF) for both BCG and ESAT-6 ($r = 0.89$ and $r = 0.92$ respectively).

	BCG				ESAT-6				
	GM-CSF	IFN-g	IL-13	MIP-1alpha	IFN-g	IL-13	IL-17	IP-10	SCD40L
GM-CSF	-	-	-	-	0.92	0.79	0.80	0.84	0.73
IFN-g	0.89	-	-	-	-	0.79	0.80	0.91	0.73
IL-13	0.74	0.72	-	-	-	-	0.76	0.75	-
IL-17	-	-	-	-	-	-	-	0.76	-
IP-10	-	0.75	-	-	-	-	-	-	0.72
SCD40L	0.76	0.76	0.70	-	-	-	-	-	-
TNF-alpha	0.71	-	-	0.71	-	-	-	-	-

Fig. 5.8 A summary of the Spearman correlation matrix, portraying the r values of major correlations found. The values in red text indicate statistically significant correlations with values ≥ 0.7 in yellow blocks and values ≥ 0.8 are shown in red blocks.

A strong positive correlation was observed between IP-10 and IFN-γ for both BCG and ESAT-6 stimulation ($r = 0.75$ and $r = 0.91$respectively). IP-10 production levels were strongly correlated with GM-CSF levels for ESAT-6 stimulation ($r = 0.84$). Production levels of IL-17 and GM-CSF, and IL-17 and IFN-γ were highly correlated when stimulated with ESAT-6 ($r = 0.80$ for both), but the correlation weakened when stimulated with BCG ($r = 0.68$ for both) (see Fig. 5.8). All of the above mentioned cytokines are pro-inflammatory cytokines and form part of the Th1 response. Other strong correlations after ESAT-6 stimulation were between IL-13 and both GM-CSF and IFN-γ ($r = 0.79$ in both cases).

	GM-CSF	IFN-g	IL-10	IL-12p40	IL-12p70	IL-13	IL-17	IL-1alpha	IL-1Ra	IL-4	IL-7	IL-8	IP-10	MCP-1	MIP-1alpha	RANTES	SCD40L	TNF-alpha
GM-CSF	1.00	0.92	0.32	0.13	0.05	0.79	0.80	0.42	0.20	0.09	0.08	0.52	0.84	0.10	0.52	0.15	0.73	0.51
IFN-g	0.89	1.00	0.29	0.18	0.05	0.79	0.80	0.39	0.20	0.06	0.04	0.48	0.91	0.12	0.43	0.09	0.73	0.45
IL-10	0.22	0.15	1.00	0.20	0.17	0.21	0.30	0.53	0.24	0.10	0.07	0.47	0.19	-0.06	0.46	0.33	0.12	0.64
IL-12p40	0.12	0.15	0.13	1.00	0.11	0.17	0.12	0.32	0.09	0.10	0.06	0.18	0.11	0.03	0.21	0.16	0.12	0.25
IL-12p70	-0.01	0.02	0.08	0.11	1.00	0.02	0.07	0.19	0.02	0.24	0.22	0.15	0.01	-0.04	0.12	0.24	0.10	0.14
IL-13	0.74	0.72	0.19	0.14	-0.01	1.00	0.76	0.26	0.06	0.09	0.14	0.37	0.75	0.16	0.32	0.06	0.69	0.35
IL-17	0.68	0.68	0.21	0.04	-0.01	0.64	1.00	0.38	0.20	0.06	0.06	0.47	0.76	0.13	0.39	0.12	0.66	0.37
IL-1alpha	0.44	0.37	0.32	0.25	0.11	0.18	0.15	1.00	0.31	0.19	0.05	0.59	0.23	-0.02	0.62	0.53	0.22	0.63
IL-1Ra	0.20	0.20	0.18	0.05	0.03	0.09	0.12	0.26	1.00	0.09	0.02	0.29	0.24	0.10	0.21	0.14	0.14	0.39
IL-4	0.14	0.08	0.11	0.15	0.24	0.12	0.10	0.27	0.10	1.00	0.10	0.02	0.07	-0.01	0.09	0.08	0.27	0.22
IL-7	0.05	0.01	0.02	0.03	0.20	0.03	0.00	0.15	0.02	0.05	1.00	0.07	0.03	0.06	0.04	0.04	0.10	0.06
IL-8	0.40	0.37	0.20	0.11	0.03	0.21	0.27	0.33	0.19	0.00	0.04	1.00	0.38	0.24	0.62	0.50	0.21	0.58
IP-10	0.58	0.75	0.13	0.06	-0.01	0.61	0.58	0.04	0.26	-0.04	-0.03	0.30	1.00	0.13	0.28	-0.08	0.72	0.35
MCP-1	0.11	0.09	0.06	0.05	-0.10	0.09	0.11	0.03	0.04	-0.05	0.02	0.44	0.09	1.00	-0.01	0.02	0.13	0.01
MIP-1alpha	0.63	0.50	0.37	0.19	0.10	0.44	0.35	0.59	0.16	0.14	0.12	0.37	0.24	0.04	1.00	0.66	0.22	0.59
RANTES	0.31	0.21	0.20	0.13	0.17	0.11	0.06	0.61	0.05	0.20	0.13	0.32	-0.15	0.11	0.61	1.00	0.03	0.38
SCD40L	0.76	0.76	0.16	0.13	0.05	0.70	0.67	0.18	0.08	0.29	-0.01	0.23	0.56	0.08	0.37	0.16	1.00	0.24
TNF-alpha	0.71	0.62	0.43	0.33	0.09	0.48	0.39	0.58	0.30	0.21	0.04	0.40	0.36	0.07	0.71	0.38	0.45	1.00

Fig. 5.9 A matrix of the Spearman correlation coefficients (r values) between antigen-specific cytokine production levels of all 18 cytokines. The upper quadrant portrays the result of stimulation by the ESAT-6 antigen (blue) and the lower quadrant by BCG antigen (green). The values in red text indicate statistically significant correlations with values ≥ 0.7 in yellow blocks and values ≥ 0.8 are shown in red blocks.

5.3.5 Factor analysis

When working with a large amount of data, factor analysis is a procedure used to explain the data in a logical and compressed way or, as in our case, according to biological relevance. Principal component analysis is used to group all the variables, in this case cytokines, into groups called factors. A significant proportion of the data can be explained by the derived factors. Parallel analysis was used in the program STATISTICA version 8.0 (www.statsoft.com) to compute the factor loadings.

The factor analysis starts off with a scree plot of all the multiple variables. Two plots were drawn for this dataset, one for BCG (Fig. 5.10 A) and one for ESAT-6 (Fig. 5.10 B). These plots are used to minimize the variables and group them into smaller groups called factors to examine their combined effect within the factor instead of separately. There are three lines on each plot, where the blue line represents the principal component of the actual data (raw data obtained in study, therefore not modelled data); the red line represents the principal component of the stimulated data and the green line represents the resampled data. Based on where the actual data (blue line) crosses the stimulated and resampled data (green and red line respectively) the number of factors involved can be deduced. In our study both datasets could be explained by four factors / eigenvalues.

The different factors obtained from the BCG and ESAT-6 analysis can be seen in Tables 5.2 A & B respectively. In both analyses the size of the first factor / eigenvalue was the largest, followed by the other factors contributing less and less. The size of an eigenvalue refers to the amount of variance (%) in the dataset that can be explained by that eigenvalue. The four major factors explained the data by a cumulative effect of 54.34% for BCG and 54.44% for ESAT-6 whereas the rest of the factors had no significant impact.

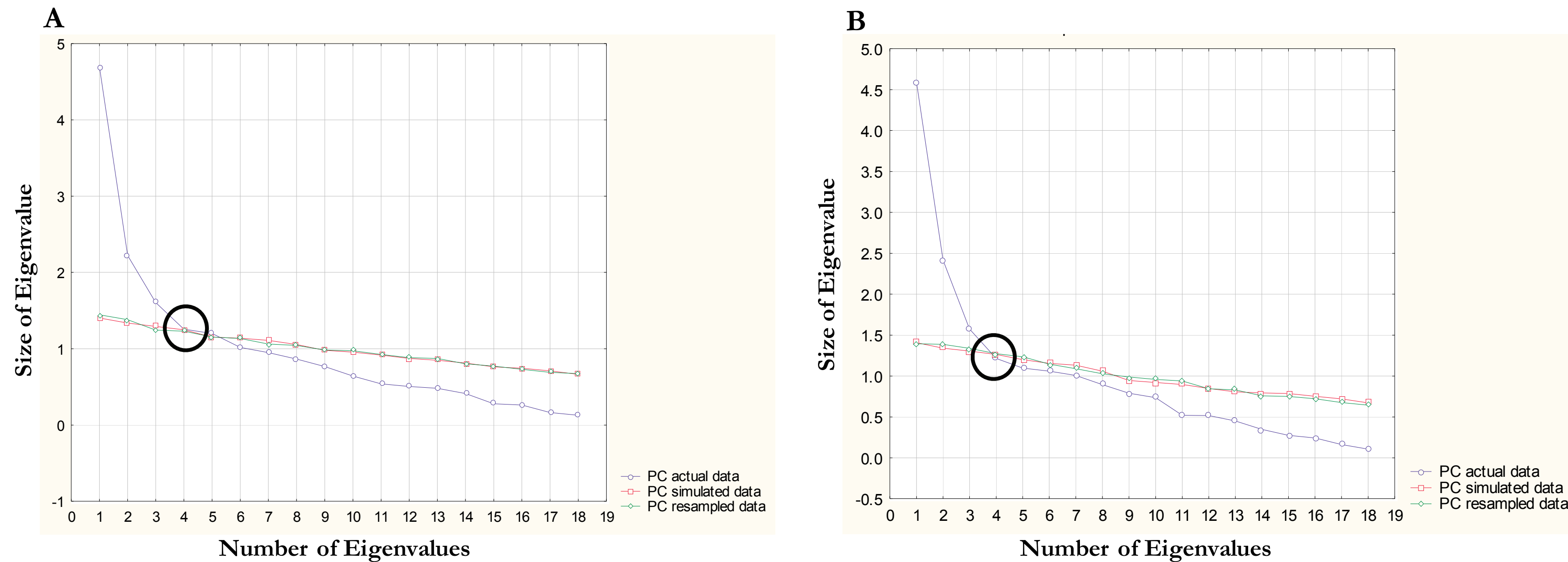


Fig. 5.10 Two scree plots of [A] BCG stimulated and [B] ESAT-6 stimulated samples of all the multiple variables. The blue line represents the principal component (PC) of the actual data; the red line represents the PC of the stimulated data and the green line represents the PC of the resampled data.

Table 5.2 Two line plots [A] BCG stimulated and [B] ESAT-6 stimulated samples of all the multiple variables.

A

Factor	Size of Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	4.69	26.08	4.69	26.08
2	2.22	12.35	6.92	38.43
3	1.61	8.93	8.52	47.36
4	1.26	6.98	9.78	54.34

B

Factor	Size of Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	4.59	25.52	4.59	25.52
2	2.42	13.43	7.01	38.94
3	1.57	8.74	8.58	47.68
4	1.22	6.76	9.80	54.44

A process called factor loading assigns certain loading parameters to each cytokine (Tables 5.3 and 5.4). A well-known standard rotation method called Varimax normalized is used to rotate the factor axis to determine which cytokine loads onto which factor. Based on this analysis it becomes evident which cytokines are the main contributors to each factor. Cytokines that have values ≥ 0.7 (highlighted in red) are considered to be major contributors to that specific factor. Where cytokines appear in more than one factor, they are counted only in the factor where they have the highest value. Focusing on the BCG factor analysis (Table 5.3), it can be seen that factor one which contains six cytokines accounted for about 26% of the variance, factor two which contains four cytokines for 12%, factor three which contains two cytokines for 9% and factor four which contains two cytokines for 7%. Both TNF-alpha and IL-1alpha (highlighted in orange) were present in two factors, but were counted once only where they had the highest value. Four of the cytokines were significantly associated with a certain factor but to a lesser extent (marked in yellow).

Table 5.3 Factor loadings for the BCG stimulated samples of which all were Varimax normalized. Cytokine loadings ≥ 0.7 are marked in red, > 0.5 in yellow and those present in two factors are marked in orange.

Variables (18 cytokines)	Factor 1	Factor 2	Factor 3	Factor 4
GM-CSF	**0.775**	0.463	-0.048	0.109
IFN-g	**0.716**	0.356	-0.053	0.120
IL-10	0.139	0.383	0.148	0.023
IL-12p40	0.065	0.194	0.439	0.057
IL-12p70	-0.041	-0.008	0.585	-0.060
IL-13	**0.719**	0.151	-0.036	-0.105
IL-17	0.676	-0.030	0.099	-0.143
IL-1alpha	0.070	0.699	0.504	0.110
IL-1Ra	0.100	0.056	0.374	0.393
IL-4	0.007	0.060	**0.809**	-0.070
IL-7	-0.124	0.050	0.028	0.433
IL-8	0.300	0.220	-0.075	0.676
IP-10	0.661	-0.139	-0.088	0.251
MCP-1	-0.016	0.018	-0.071	**0.766**
MIP-1alpha	0.098	**0.857**	-0.104	-0.012
RANTES	-0.065	**0.811**	0.105	0.150
SCD40L	**0.798**	0.068	0.234	-0.034
TNF-alpha	0.466	0.630	0.109	0.225

Focusing on the ESAT-6 factor analysis (Table 5.4), factor one which contains six cytokines accounted for about 26% of the variance, factor two which contains four cytokines for 13%, factor three which contains three cytokines for 9% and factor four which contains two cytokines for 7%. Both IL-10 and IL-1alpha (highlighted in orange)

were present in two factors, but were counted once only where they had the highest value. Cytokines marked in yellow were significantly associated with a certain factor but to a lesser extent. It was notable that one of the cytokine correlations, IL-7 in factor 4, was negative.

Table 5.4 Factor loadings for the ESAT-6 stimulated samples of which all were Varimax normalized. Cytokine loadings ≥ 0.7 are highlighted in red, > 0.5 in yellow and those present in two factors are highlighted in orange.

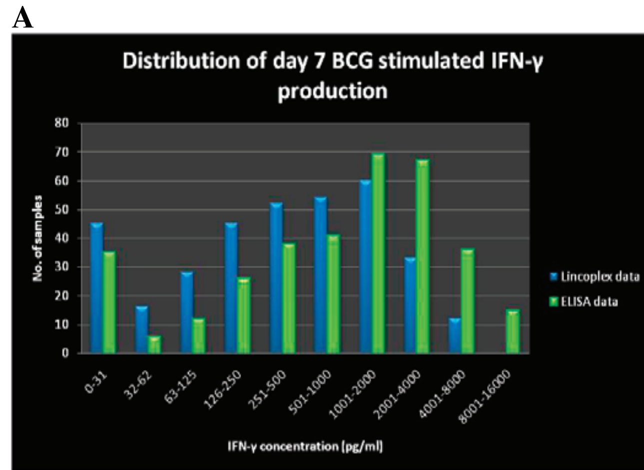
Variables (18 cytokines)	Factor 1	Factor 2	Factor 3	Factor 4
GM-CSF	**0.814**	0.302	0.049	-0.207
IFN-g	**0.861**	0.127	0.040	-0.053
IL-10	0.189	0.354	0.027	0.551
IL-12p40	0.047	0.243	0.266	0.359
IL-12p70	0.069	0.078	**0.765**	-0.132
IL-13	0.567	0.108	-0.048	0.185
IL-17	**0.717**	0.078	0.052	0.211
IL-1alpha	0.222	0.560	0.563	0.217
IL-1Ra	0.211	0.327	0.012	0.006
IL-4	-0.035	-0.033	**0.852**	0.076
IL-7	0.023	0.148	0.234	-0.587
IL-8	0.253	0.675	-0.006	0.051
IP-10	**0.773**	-0.033	-0.070	-0.155
MCP-1	0.186	0.173	-0.109	-0.446
MIP-1alpha	-0.031	**0.788**	-0.031	-0.183
RANTES	-0.207	**0.786**	0.088	0.006
SCD40L	**0.823**	0.055	0.238	-0.093
TNF-alpha	0.278	0.570	0.171	0.036

5.3.6 Comparison between ELISA and Lincoplex data

The enzyme-linked immunosorbant assays (ELISAs) are still widely used, but the reduced sample volume and time-saving advantages of the Lincoplex system have made it a far more attractive method to use. While with an ELISA only one cytokine can be detected at a time, up to 42 cytokines can be detected simultaneously with the Lincoplex. All the enzyme-linked immunosorbant assays (ELISAs) were done by others in the department prior to the Lincoplex analysis. Cells were stimulated in quadruplicate with BCG and ESAT-6 and incubated at 37°C with 5% CO_2. The supernatants were harvested on days 3 and 7 whereafter the IFN-γ response was measured. With the day 7 data from both the ELISA and lincoplex analyses we could do a comparison between these two assays based on their ability to measure IFN-γ levels. From the comparison between the Lincoplex and ELISA day 7 BCG stimulated IFN-γ concentration levels [pg/ml] it is evident that the

lincoplex technique is more sensitive in detecting lower concentrations than ELISAs (Fig. 5.11A). However, the ELISA detected higher IFN-γ concentrations than the Lincoplex assay. The same trend was observed for the ESAT-6 stimulated IFN-γ production levels (Fig. 5.11B). The cut off value used for ELISA IFN-γ measurements was 4000pg/ml, therefore any values beyond 4000pg/ml were extrapolated and might not be as accurate. The BCG stimulated IFN-γ levels were much higher than the ESAT-6 stimulated values (Fig. 5.11). A Spearman correlation test was performed between the ELISA and Lincoplex day 7 BCG and ESAT-6 stimulated IFN-γ production (Fig. 5.12A & B respectively). The IFN-γ results from these two techniques were highly correlated for both BCG ($r = 0.86$) and ESAT-6 ($r = 0.88$) antigens (Fig. 5.12).

A



B

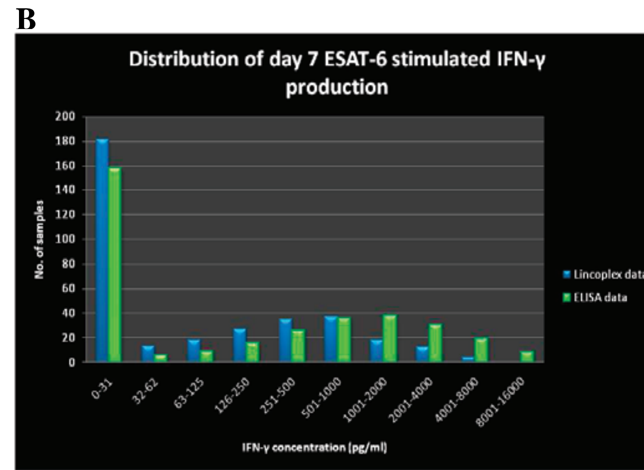
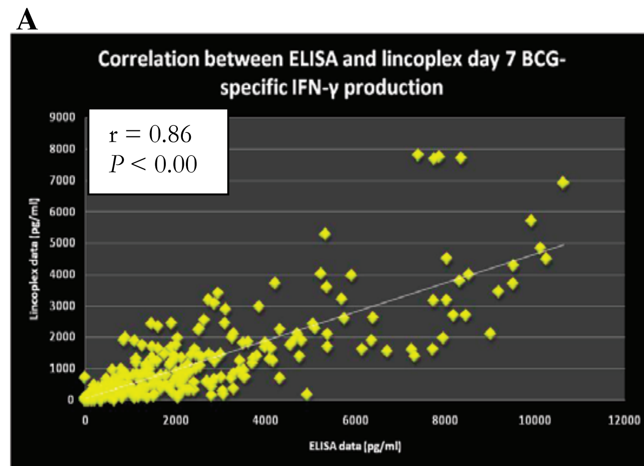


Fig. 5.11 Histograms of [A] BCG stimulated and [B] ESAT-6 stimulated samples for IFN-γ levels. The blue bars represent the Lincoplex data and the green bars ELISA data.

A



B

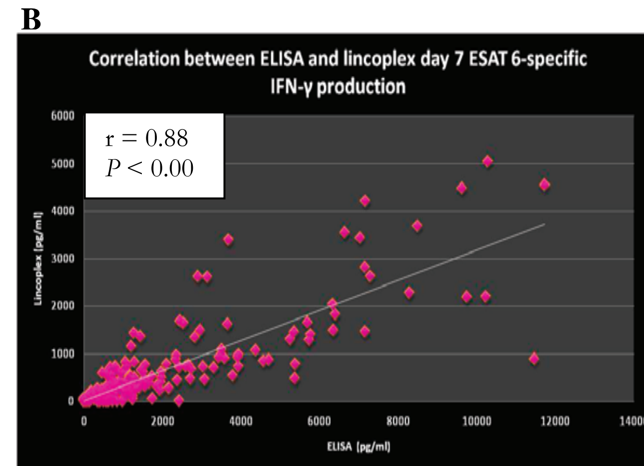


Fig. 5.12 Correlation between [A] BCG stimulated and [B] ESAT-6 stimulated samples of IFN-γ.

5.4 DISCUSSION

In general we found no major correlation between the Mantoux readings and any of the 18 cytokines measured. Focusing on the correlation test between BCG stimulated samples and Mantoux readings we found no correlation (as expected), perhaps since BCG used for the WBA, and PPD, used for the skin test, originates from two different Mycobacterium species. PPD is homogenous purified protein derivatives from *Mycobacterium tuberculosis* whereas BCG is derived from *Mycobacterium bovis*. Furthermore the skin test is a delayed response as the measurements are taken only 48 -72h after administration. Another reason why a good correlation would not necessarily be expected is because one is comparing delayed results with a 7 day BCG stimulated assay of peripheral blood.

The Mantoux skin test can be confounded by cross-reactivity with non-*M. tuberculosis* antigens resulting in false positives (Farhat *et al.,* 2006; Wang *et al.*, 2002) and is also prone to an increased rate of false negative results among immune-compromised individuals (Madhi *et al.*, 2000). To overcome the problem of limited Mantoux skin test specificity, alternative assays have been developed that utilize antigens such as ESAT-6 with increased specificity for *M. tuberculosis* (Menzies *et al.*, 2007).

Focusing on the correlation test between ESAT-6 stimulated samples and Mantoux readings, we found very weak correlations ($r = 0.28$). ESAT-6 is a small protein consisting of 95 amino acids and is secreted by the *Mycobacterium tuberculosis* complex (Sørensen *et al.*, 1995). Due to the extremely small size of the protein, a very sensitive assay is needed to detect ESAT-6 responses. This is true even for the Lincoplex analysis, as we had many ESAT-6 non-responders in our data. One would expect a correlation between ESAT-6 stimulated samples and Mantoux readings as both ESAT-6 and PPD test exposure to *M. tuberculosis*. The PPD skin test, assesses whether one has been exposed to *M. tuberculosis* and one will only have an ESAT-6 response once one has been exposed to *M. tuberculosis*, as it is a protein secreted by the *M. tuberculosis* complex. Therefore anyone that has a positive skin test should theoretically have an ESAT-6 response as they have been exposed to the *M. tuberculosis* complex.

Once the data was divided into two groups, viz. >0 mm and 0 mm (no response), we found very strong associations for both the BCG and ESAT-6 stimulated samples ($P <$

0.01) with the Th1-type cytokines such as IFN-γ. Neither the BCG results nor the ESAT-6 results were unexpected, since it is known that BCG induces the Th1 response (Luo *et al.*, 2003) and ESAT-6 is a protein secreted by the *M. tuberculosis* complex. A response to ESAT-6 is interpreted to imply that a person has been exposed to *M. tuberculosis*. Accordingly, we found a strong association between the ability to mount a positive response to BCG and ESAT-6. No significant correlation was observed for any of the Th2–type cytokines.

No major correlations were found between any of the 18 cytokines and age. The theory behind testing whether a correlation exists between age and certain cytokine production levels lies in the fact that the older one becomes the more chance one has had of being exposed to TB. Therefore one would expect to see an increase in ESAT-6 response as an individual gets older. On the other hand it is hypothesised that the BCG response would decrease with age, since all babies are vaccinated with BCG at birth in South Africa, (since 1975). This BCG immunity tends to fade over time (Wang *et al.*, 2002). We found a weak correlation ($r = 0.29$) with ESAT-6 and age but as mentioned before there were many non-responders. This could be due to the low detection levels of ESAT-6 response, poor antigenic properties of ESAT-6 or may indicate that only some of the participants had been exposed to *M. tuberculosis*.

The concentration of all 18 cytokines showed a significant correlation between BCG and ESAT-6 stimulated samples. IFN-γ production was significantly higher in the BCG stimulated samples than the ESAT-6 samples, however this might once again be due to the low detection of ESAT-6 responders. IL-8 production was high in both BCG and ESAT-6 stimulated samples. IL-8 is a pro-inflammatory cytokine that responds well to any antigen stimulation as it might not be T-cell receptor specific and therefore the response might be nonspecific. Interestingly, an extremely strong correlation ($r = 0.83$) was found between BCG and ESAT-6 stimulated samples for IL-1Ra secretion. According to these results, ESAT-6 is far more likely to induce IL-1Ra than IFN-γ. Due to the strong correlation observed between BCG and ESAT-6 stimulated samples this study might indicate that IL-1Ra is a better cytokine to measure exposure to *M. tuberculosis* than IFN-γ. Therefore the question arises, is IFN-γ really giving the best read out of an ESAT-6 stimulation assay or

will other cytokines such as IL-1Ra be a better indicator of *M. tuberculosis* exposure? This needs to be investigated further.

An extremely strong correlation ($r = 0.90$) was found between BCG and ESAT-6 stimulated samples with IL-4, but this might be due to the low levels of IL-4, since the majority of the measurements lay between 0 and 200 pg/ml. This might be explained by the low detection levels, which is a common problem with Lincoplex measurements of IL-4 (Siawaya *et al.*, 2008) or low IL-4 production by individuals. Therefore, it is difficult to draw any conclusions from the IL-4 data.

Strong correlations were observed between antigen-specific (BCG and ESAT-6) cytokine production levels of some of the 18 cytokines. The most statistically significant correlations were between IFN-γ and GM-CSF for both BCG and ESAT-6 ($r = 0.89$ and $r = 0.92$ respectively). For ESAT-6 stimulation, the most significant correlation were between IP-10 and GM-CSF ($r = 0.84$), IP-10 and IFN-γ ($r = 0.91$), IL-17 and GM-CSF ($r = 0.80$) and, IL-17 and IFN-γ ($r = 0.80$).

All four of these cytokines (IFN-γ, GM-CSF, IP-10 and IL-17) are pro-inflammatory cytokines that form part of the Th1 response. IFN-γ, secreted by dentritic cells, NK cells and CD8+ cytotoxic T cells, is known as the major Th1-type cytokine and suppresses Th2 cell activity (Schoenborn & Wilson, 2007). Therefore the production of this cytokine indicates a T cell-mediated response to an intracellular pathogen such as *M. tuberculosis*, resulting in macrophage activation (Kaufmann *et al.*, 2006). GM-CSF, secreted by T cells stimulates monocytes that mature into macrophages which play a crucial role in the immune/inflammatory cascade. IP-10, also known as a chemokine, is induced in several cell types including monocytes in response to IFN-γ (Sauty *et al.*, 1999). IP-10 acts as a chemoattractin for macrophages, T cells and NK cells. IL-17 forms part of a group of cytokines called the IL-17 family. It is involved in inducing and mediating pro-inflammatory responses as well as the production of other cytokines and chemokines such as IFN-γ, GM-CSF and IP-10 (Fig. 5.13).

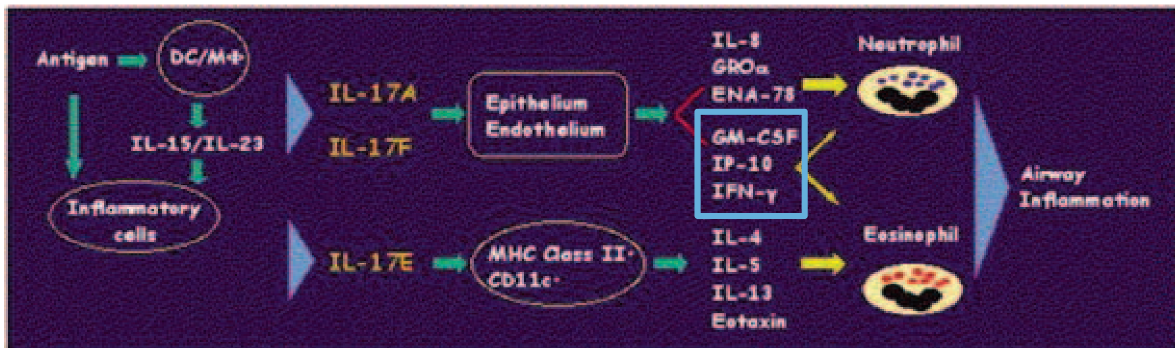


Fig. 5.13 The role of IL-17A, IL-17E and IL-17F in the expression of airway inflammation. Dentritic cell (DC), macrophage (MΦ), Growth-related oncogene α (GROα), Epithelial cell-derived neutrophil activating protein 78 (ENA-78) (Kawaguchi *et al.*, 2004)

These four pro-inflammatory cytokines have a strong mutual correlation after ESAT-6 stimulation which may indicate concerted secretion as a response to exposure to *M. tuberculosis*.

We found that the variance in the data could be explained by four major factors. The cumulative effect of these factors was the same for both the BCG and ESAT-6 models (54%). The first three factors were the same for both BCG and ESAT-6 models. Based on this analysis it became apparent which cytokines are the main contributors to each factor. Factor one, which contains GM-CSF, IFN-γ, IL-13, IL-17, IP-10 and the soluble CD40 ligand (SCD40L), accounted for about 26% of the variance. All these cytokines with the exception of IL-13 are pro-inflammatory and are Th1 biased cytokines / chemokines that are induced by stimulation with BCG and ESAT-6. The production of all these cytokines except IL-13, indicate a T cell-mediated response to an intracellular pathogen such as *M. tuberculosis* playing a crucial role in the immune/inflammatory cascade as it results in macrophage activation. IL-13 is one of the major Th2-type cytokines that plays an important role in mediating allergic inflammation. Since the Th2 immune response is suppressed by the Th1 immune response it was surprising to find IL-13 grouped with other Th1-type cytokines. However, this might be due to very low expression levels of IL-13 throughout the Lincoplex analysis resulting in a false signal.

Factor two which contains IL-1alpha, MIP-1alpha, RANTES and TNF-alpha accounted for about 12% of the variance. All these cytokines are pro-inflammatory and Th1 biased

cytokines / chemokines that are expressed at increasing levels after stimulation with BCG and ESAT-6. All these cytokines are involved in macrophage activation.

Factor three which contains IL-12p70 and IL-4 accounted for about 9% of the variance. These two cytokines are very different from each other, as IL-12p70 is a Th1-type pro-inflammatory cytokine and IL-4 is a Th2-type anti-inflammatory cytokine. IL-12 is a crucial regulator of cell-mediated immune responses and it is secreted by activated B-lymphocytes and macrophages as a 70-kD glycoprotein (IL-12p70). The development of pro-inflammatory Th1-type $CD8^+$ T cells and NK cells are promoted by IL-12 production. It has previously been shown that IL-4 is a powerful enhancer of the production of bioactive IL-12p70 in dentritic cells (DC) and macrophages. Human Th2 cells producing IL-4 therefore efficiently induce IL-12p70. It acts as a Th2 specific co-inducer of the production of bioactive IL-12p70, resulting in an inbuilt tendency of human Th2 cells to regress to the IFN-γ-producing Th0 / Th1 phenotype upon interaction with DC (Kaliński *et al.*, 2000). DC isolated from the airways and from the gut-associated lymphoid tissues has been shown to be IL-12 deficient (Iwasaki & Kelsall, 1999; Stumbles *et al.*, 1998). The grouping of these two cytokines together is therefore concordant with their known function. However, this might also be an over-interpretation of what is happening, since the IL-12p70 values were very low throughout the Lincoplex analysis.

Different cytokines shaped the fourth factor of the BCG and ESAT-6 models. Factor four contained IL-8 and MCP-1 in the BCG model and accounted for about 7% of the variance. Both these cytokines are pro-inflammatory and Th2-type cytokines. IL-8 forms part of the C-X-C chemokine family and acts as a potent activator and chemoattractant of neutrophils (Mukaida *et al.*, 1994). In contrast, MCP-1 forms part of the C-C chemokine family and functions as a chemoattractant and activator of lymphocytes and monocytes leading to monocyte / macrophage infiltration into tissues (Taub *et al.*, 1995). Both of these cytokines therefore act as chemoattractants that contribute to the influx of monocytes and neutrophils after infection / inflammation which can occur during both Th1 and Th2-biased infections. This might be why these two cytokines were expressed at noticeably higher levels (2-fold) compared to other cytokines detected in the Lincoplex assay. IL-8 and MCP-1have interactive effects and are strongly complementary *in vivo,* which may be a reason for these two cytokines appearing together in one factor. In the ESAT-6 model,

factor four contains IL-10 and IL-7 and accounted for about 7% of the variance. These two cytokines are both B cell-regulating cytokines but different in that IL-10 is an anti-inflammatory cytokine and IL-7 a pro-inflammatory cytokine. Based on the factor loading results these two cytokines have opposite effects as IL-10 has a factor loading of +0.55 and IL-7 has a factor loading of -0.58. This result might be explained by the opposite functions of the two cytokines, or the extremely low expression of IL-7 observed in the Lincoplex assay, and cannot be regarded as a reliable result.

In our comparative assessment of IFN-γ in the ELISA and from the Lincoplex assay, we found the results from these two analyses to be highly correlated with each other. This was true for both BCG and ESAT-6 stimulated IFN-γ levels. This comparative study between the ELISA and Lincoplex IFN-γ levels showed that, although ELISA's are still being widely used, the multiplex assay is highly accurate and comparative with ELISA results with added advantages like reduced sample volume and time-saving benefits. The Lincoplex assay was also more sensitive at detecting lower IFN-γ concentrations. Strong correlations have previously been reported between ELISA and Lincoplex IFN-γ levels measured (Siawaya *et al.,* 2008; dupont *et al.*, 2005).

In conclusion, the multiplex analysis is rapidly becoming the most widely used screening tool for the selection of candidate biomarkers associated with specific diseases. The yield of information with the Lincoplex assay compared to the ELISA assay is immense. We were able to simultaneously study 18 cytokines in 340 individuals, leading to a vast amount of phenotypic information. Inclusions of other disciplines, or more specifically immunogenetic approaches, play a crucial role in unraveling the, yet unknown about complex diseases such as TB. Therefore, since cytokines play a major role in immunity to tuberculosis (phenotypic information) the functional polymorphisms or haplotype patterns (genotypic information) in some of these cytokine genes might be vital for protective immune responses and may serve as biomarkers of protection or susceptibility to TB (Selvaraj *et al.*, 2008).

In future, multiplex methods could be tailor-made to test specific cytokines relevant to the disease of interest, in this case TB. This might lead to a better understanding of the pathophysiology of TB and serve as an objective tool to evaluate disease severity and

progression. Host susceptibility to *M. tuberculosis* is multifactorial and complex as it involves many components of the immune system (Raja, 2004). It is possible that manipulation of these specific cytokines may lead to new and more specific therapeutic alternatives (Jacob *et al.*, 2003).

5.5 ACKNOWLEDGEMENTS

I would like to thank Kim Stanley (US) for the ELISA analysis and Dr. Chantal Babb (US) for help with the Lincoplex analysis. I am grateful to Caroline Gallant and Leah Simkin (McGill University, Canada) for advice on database management and for compiling the Linco-plex database. I would also like to thank Prof. Martin Kidd (US) for help with the statistical analysis.

CHAPTER 6

CONCLUDING REMARKS

This study was undertaken to investigate susceptibility to tuberculosis in a South African Coloured (SAC) population, by consolidating a number of candidate gene studies and extending these results to include gene-gene interaction analysis and mycobacterial genotypes.

South Africa has one of the highest incidences of TB in the world, reflected by the situation in communities such as Ravensmead and Uitsig in the Western Cape, where a large number of individuals have had TB episodes on two or more occasions. It is not known whether this is due to the higher susceptibility of these patients, factors in their environment, or whether it is due to the virulence of particular strains of *M. tuberculosis* or all of these.

The SAC population is a unique admixed population with a strong African admixture component. Therefore, the SAC population might have more genetic variability than other more homogeneous population groups. It is known that an admixed population has a stronger linkage disequilibrium (LD), making the SAC population an ideal population to study disease susceptibility, since functional markers can be identified.

In this study we were able to replicate many previously described associations found in the SAC population (see chapter 3). Although we were unable to reproduce some of the previous associations found, the results are still of value. A result that is not statistically significant is often seen as a negative result and therefore not published. However, both positive and negative results from association studies are important.

Our case-control association studies in a SAC population contribute significantly to the TB host genetics field, since nine polymorphisms in eight candidate genes (Table 6.1), previously indicated to play an important role in TB susceptibility, were investigated in a sample size large enough to have sufficient power to detect whether there was an association with tuberculosis susceptibility.

Table 6.1 Summary of all the polymorphisms investigated in the case-control association study in the South African Coloured population

Gene	Variant (rs number)	Significant *P*-values		Results in Table:
		Genotype	Allele	
SP-D	Met11Thr (rs721917)			3.2
MBL	Codon 52 (rs5030737)			3.3
	Codon 54 (rs1800450)			
	Codon 57 (rs1800451)			
IFN-γ	+874 A/T (rs2430561)			3.4
IFNGR1	1050 T/G (rs11914)		**0.022**	3.5
IL-8	+100 C/T (rs2227538)		**0.019**	3.6
RANTES	-403 A/G (rs2107538)			3.7
NRAMP1	5' $(GT)^n$ repeat		allele 118: **0.011**	3.8
			allele 120: **0.011**	
	3' UTR (TGTG) deletion			3.9
IL-1Ra	86bp VNTR	**0.037**	allele 1: **0.019**	3.11
			allele 2: **0.025**	

This is the first time, to our knowledge, that gene-gene interaction analysis was done in the context of TB host susceptibility. In this case, a condensed set of 11 variants in these eight genes was used (*SP-D, MBL*, *IFN-γ*, *IFNGR1*, *IL-8*, *RANTES, NRAMP1* (*SLC11A1)*, and *DC-SIGN*). Of the 103 possible interacting pairs, we detected significant interactions between eight pairs of variants. The combined data set consisting of 106 variants yielded a very large number of pair-wise interactions. Of the 318 statistically significant interactions ($P < 0.05$) found, 67 were highly significant ($P < 0.01$). Considering the amount of data rapidly becoming available from SNP chip studies, similar interaction studies between known susceptibility genes and others could well identify new candidates involved in TB susceptibility. This is a multilocus and multi gene approach, consistent with the nature of complex-trait diseases, and may provide the paradigm for future genetic studies of TB. In addition to the host genotype conferring susceptibility, the interaction between the genotype of the human host and the bacterial strain genotype was also investigated in a subset of 229 TB patients, with no significant interactions detected.

This admixed population was also assessed for population substructure using 209 unlinked randomly selected microsatellites, since substructure may have a confounding effect on genetic association studies. Our results finding no substructure in the population were congruent with those with Barreiro *et al*., (2006), who found no population stratification.

However, we now have a better understanding of population stratification and advances in bio-informatics and genotyping methods show that more than 25 SNPs are required to reliably test for population stratification (Hao *et al.,* 2004; Turakulov & Easteal, 2003). In this case, 209 microsatellites were used to strengthen the evidence that no population substructure exists in the SAC population

The population structure or ancestral contributions to this admixed population were investigated using a large dataset consisting of 959 individuals from the SAC group genotyped for nearly 75,000 markers. This was the first investigation at this high level of genetic resolution of a South African population. Our study showed that the genetic background agrees with the historical record and also reveals more interesting detailed insights into the events that resulted in the present-day population. We showed that the SAC population group has four major ancestral components: the Khoesan, European, African and Asian (Indian), in approximate proportions of 42%, 28%, 21% and 9% respectively. Knowledge of the population structure will be important for association studies. One of the current genome-wide methods to find genetic risk factors for disease is that of admixture mapping (see section 6.2). The SAC would appear to be ideally suited to this type of investigation, having ancestral populations with apparently widely different rates of TB infection and disease. Given the high number of private alleles found in this population (Tishkoff *et al.*, 2009), novel susceptibility alleles could be identified.

Owing to the complexity of the immune system and the multifactorial nature of a complex disease such as TB, it has become increasingly evident that unravelling the underlying mechanisms indicated by the genetic component may lead to new insights into the pathogenesis of TB and the development of new treatment strategies. Therefore, since cytokines play a major role in immunity to tuberculosis (phenotypic information) the functional polymorphisms or haplotype patterns (genotypic information) in some of these cytokine genes might be vital for protective immune responses and may serve as biomarkers of protection or susceptibility to TB (Selvaraj *et al.*, 2008). We simultaneously investigated 18 cytokines in 340 individuals using multiplex analysis, leading to a vast amount of phenotypic information. Host susceptibility to *M. tuberculosis* is multifactorial and complex as it involves many components of the immune system (Raja, 2004). It is

possible that manipulation of these specific cytokines and their balance may lead to new and more specific therapeutic alternatives (Jacob *et al.*, 2003).

6.1 Potential limitations of the study

Owing to the high incidence of latent TB infection in the Ravensmead-Uitsig community, many of the control samples used in this study were probably TST skin test positive, therefore latently infected. Consequently, our case-control association study investigated host genetic factors that may predispose individuals to progression to disease. The possibility exists that a separate set of genes are involved in infection, compared to progression to disease. However, the genes studied here might be involved in primary infection.

Some schools of thought suggest that statistical genetics should include corrections for multiple testing, however, no corrections were done in the association studies presented here. It is debatable how relevant multiple corrections are, as some argue that these tests are too stringent, resulting in the exclusion of true susceptibility variants. This is arguably particularly relevant in a complex disease without a single major susceptibility locus, which is further complicated by the extraneous factor of an infectious agent. Too stringent requirements may well result in the spurious exclusion of associated loci. All the variants studied here had a prior probability of being associated with TB susceptibility based on their function or previous analyses. Therefore, it is argued that variants studied here were not randomly selected and therefore did not require corrections for multiple testing.

Another possible statistical limitation is spurious associations in the combined gene-gene interaction dataset consisting of 106 variants and in the gene-trait interaction analysis. Since multiple pair-wise interaction tests were done between the different variants it may be prone to false positive associations, because a percentage of multiple tests with the same dataset will have positive P values just by chance.

The TBM sample size ($n = 50$) might not have been sufficient or have had enough power to detect an association between TBM susceptibility and the genes investigated. Although a stronger association might have been expected with TBM than pulmonary TB as it is an extreme phenotype, our sample size was too small to detect any association.

It has been hypothesised that many other inherited DNA features and not only gene polymorphisms may also influence susceptibility to a disease. Epigenetic mechanisms, such as DNA methylation and histone acetylation changes, regulate the transcription rate and/or tissue-specific expression of certain genes without altering the primary nucleotide sequence of the DNA. *M. tuberculosis* might interfere with epigenetic regulation to add to the complexity of the disease process.

6.2 Future directions

The significant associations found in this study should be validated in other population groups to assess the importance of these polymorphisms in susceptibility to TB. All pair-wise gene-gene and gene-trait interactions should also be validated and improved computational methods need to be developed to handle large data sets when investigating gene-gene interactions. Other future directions could include:

i) Genome-wide association studies

Genome-wide association studies (GWAS) allow the genotyping of the most frequent genetic polymorphisms in the genome without making assumptions about the genomic location of the causal variants. Since most of the genome is surveyed, it eliminates the disadvantages of the single polymorphism or candidate gene approach where only a few polymorphisms are investigated (Hirschhorn & Daly, 2005). The completion of the human genome sequence, the deposition of SNPs into public databases, the rapid improvements in SNP genotyping methods (such as the development of microarray platforms) and the International HapMap Project have allowed the genetic association field to progress to this study design. Previous investigations (Daly *et al.,* 2001; Gabriel *et al.*, 2002; Patil *et al.*, 2001) and the International HapMap project have shown that most common variation in the genome can be represented by approximately 300 000 SNPs (Balding, 2006) in white populations. African and other populations with greater variation and less LD need more SNPs (Hirschhorn & Daly, 2005) to ensure coverage of the entire genome. When this study design was first suggested, one of the major theoretical stumbling blocks was the problem of statistical correction for multiple questions.

GWAS have the potential to identify many false associations and therefore replication or validation in independent populations is essential (Burgner *et al.*, 2006). However, these

studies are very expensive. In addition, the best approach to adjust for multiple comparisons in these studies has not been determined. A possible strategy for the elimination of false-positive results is to use a two-step study design and to adopt strict rules for declaring significant associations. A few individuals are genotyped genome-wide in the first stage of the study. In the second stage, promising SNPs are genotyped in the remainder of the study population (Balding, 2006). In cases where a significant association is determined, the results are verified in another population, preferably using another genotyping method to exclude technical artefacts. These studies need a large sample size to ensure adequate power. An additional challenge for GWA studies is that it is currently based on the "common disease, common variant" (CDCV) hypothesis, which means that rare susceptibility variants may not be detected.

There are currently no published GWAS in TB, but susceptibility to two other infectious diseases, namely Kawasaki disease (Burgner *et al.*, 2009) and chronic hepatitis B (Kamatani *et al.*, 2009), has been investigated using GWAS. The Wellcome Trust Case-Control Consortium has initiated GWAS that will consider up to 2000 cases and 2000 controls for TB and malaria (Burgner *et al.,* 2006; Wellcome Trust Case Control Consortium, 2007).

ii) Admixture mapping

Admixture mapping, a novel approach for disease gene discovery, is an additional methodology for which the resolution is higher than that of linkage analysis, but lower than that of association studies (Chakraborty & Weiss, 1988; Stephens *et al.*, 1994). This technique involves using a population which has arisen from two or more genetically different parent populations where the frequency of the disease, and therefore presumably of the underlying risk variants as well, is different in the founding populations. The aim is to localise the segments of the genome inherited from a specific ancestral population in the cases in order to identify the locus responsible for the phenotype of interest. The primary tools are the genetic markers that occur with significantly different allele frequencies in different population groups. When risk alleles vary across populations, genetically mixed individuals with the disease under investigation are likely to have a higher probability of having inherited the loci near the disease loci from the population at higher risk of the disease. The study design has recently been used successfully in a variety of complex diseases, e.g. hypertension (Zhu *et al.*, 2005), multiple sclerosis (Reich *et al.*, 2005) and

prostate cancer (Freedman *et al.*, 2006) in African Americans. Preparation for studies in admixed populations from Mexico (Bonilla *et al.,* 2005; Martinez-Marignac *et al.*, 2007) and Hispanic/Latino populations (Mao *et al.,* 2007; Price *et al.*, 2007) are also underway, but none of these studies have involved TB.

There is an apparent disparity in the rates that European and black individuals are infected with TB and progress to disease (Stead et al., 1990). This is not merely a reflection of socio-economic circumstances, but it has been speculated that it is the result of centuries of exposure to TB in Europe, which may have resulted in a degree of selection for a more resistant population, whereas Sub-Saharan Africa was exposed to the disease only relatively recently. Although environmental and social factors are difficult to control for, it does appear that the underlying susceptibility is different and thus an approach using an admixed populations from these parent populations could yield useful research.

Admixture mapping may serve as an excellent approach for an initial genome scan owing to several advantages when compared to linkage or association studies (Darvasi & Shifman, 2005). This technique has a higher statistical power to detect genes of modest effect than linkage analysis, and if risk alleles are differentially distributed between ancestral populations, and if the frequencies differ greatly across populations, the power approaches that of association mapping (Montana & Pritchard, 2004). In addition, admixture mapping studies are not as affected by allelic heterogeneity as are other study designs (Martinez-Marignac *et al.*, 2007). This approach could yield novel susceptibility genes for TB. An admixture mapping study with respect to TB is planned for the admixed population studied in this thesis.

iii) Epigenetics

A hypothesis that genetic polymorphisms are not the only inherited features of DNA which may influence disease susceptibility has been considered for several years (Richards, 2006). Epigenetic mechanisms, such as different changes in DNA methylation and histone acetylation, regulate the transcription rate and/or tissue-specific expression of certain genes without altering the primary nucleotide sequence of the DNA.

DNA methylation plays an important role in the arrangement of some key biological activities such as imprinting and silencing of chromosomal domains (Reik & Walter, 2001; Wang & Leung, 2004). It usually occurs at the cytosine residue in the context of a CpG dinucleotide and promoter regions, where it can act as an important modifier of transcription. The relationship between the methylation of the CpG islands and gene expression is very complex. Some studies have reported that the change of methylation intensity of the promoter CpG islands is negatively correlated with gene expression levels (Futscher *et al.,* 2002; Song *et al.*, 2005), while others have observed no correlation (Illingworth *et al.,* 2008; Weber *et al.*, 2007).

Histone acetylation is crucial for gene transcription as histone acetylation sites are required for chromatin folding and gene activity. Histone acetyltransferases form part of a family of enzymes that are involved in several biological processes. It has also been shown that histone deacetylation plays an important role in transcription (Shahbazian & Grunstein, 2007).

The above DNA features have been shown to be transmitted to subsequent generations, but they may also be influenced by the environment (Richards, 2006). However, this inheritance is not as stable as DNA-based inheritance and it is not completely elucidated. It has been shown that *M. tuberculosis* may interfere with epigenetic regulation (Pennini *et al.*, 2006). Therefore, once the inheritance is understood, it is possible that these epigenetic mechanisms may be found to play a role in host susceptibility to TB.

iv) Copy number and rare variants

The majority of the methods above are based on the CDCV hypothesis which means that rare susceptibility variants may not be detected, and are therefore designed to detect common genetic markers. However, a number of sceptics have postulated that this hypothesis is not true, and will therefore not lead us to some of the more meaningful susceptibility-associated genes (Goldstein, 2009).

DNA copy number variations (CNVs), that can be defined as 1 kb and larger stretches of DNA that display copy number differences compared to the normal population, have generated remarkable enthusiasm in the science community due to their role in functional

variation (Iafrate *et al.,* 2004; Scherer *et al.*, 2007). It was observed that CNV hotspots exist (Lee *et al.*, 2008) and that CNV segments were considerably enhanced among sequences with low or moderate SNP content (Cutler *et al.*, 2007). CNVs in individuals have been associated with disease or susceptibility to several diseases (Baldini, 2004; Friedman *et al.*, 2006; Marshall *et al.*, 2008; Stefansson *et al.*, 2008). It has also been shown that the copy number amplification of regions containing the *MBL*, *pfmdr1* and *gch1* genes influenced the susceptibility to bacterial infection in the zebrafish (Jackson *et al.*, 2007). Individuals with the complex disorder of schizophrenia were recently found to owe their predisposition to many individually rare mutations, including rare mutations and duplications (Walsh *et al.*, 2008). If this scenario is also true for TB, it will require a more intense search for the rare variants governing susceptibility.

6.3 Conclusion

TB is not only influenced by the bacterium as stated several years ago (Dubos & Dubos, 1952), but it is a complex disease influenced by immunological, environmental, socio-economic and host genetic susceptibility factors. Looking at the history of TB, medical solutions alone were not enough to prevent or cure TB, therefore other fields such as human genetic susceptibility to TB need to be further investigated.

Even though this research is years from being clinically applied, the era of personalised medicine is already upon us, and it may not be long before the admittedly controversial benefits of personalised genetics can be extended from conditions such as cardiovascular disease, to infectious diseases such as TB. If individuals can be identified as potentially more vulnerable, they may require different vaccination strategies, a higher index of suspicion if exposed to TB, and prophylactic treatment.

Complex diseases may in all likelihood continue to earn their name and susceptibility could prove to be due to the interaction between both common and rare genetic variants in the host, variation in the bacterium, and the environment, requiring highly sophisticated algorithms to predict susceptibility.

REFERENCES

Abe T, Iinuma Y, Ando M, Yokoyama T, Yamamoto T et al. (2003) NRAMP1 polymorphisms, susceptibility and clinical features of tuberculosis. J.Infect 46: 215-220.

Abel L, Dessein A (1997) The impact of host genetics on susceptibility to human infectious diseases. Curr.Opin.Immunol 9: 509-516.

Abula H, Abulb A, Dashtic H, Behbehanic A, Hussaind T et al. (2008) Comparison of Local and Systemic Production of lnterleukin-8 in Smoking and Non-Smoking Patients with Infectious Lung Disease. Med Principles Pract 5: 19-24.

Adhikari M (2005) Not white enough, not black enough: racial identity in the south african coloured community. Ohio University Press. p.

Akahoshi M, Ishihara M, Remus N, Uno K, Miyake K et al. (2004) Association between IFNA genotype and the risk of sarcoidosis. Hum.Genet 114: 503-509.

Akahoshi M, Nakashima H, Miyake K, Inoue Y, Shimizu S et al. (2003) Influence of interleukin-12 receptor beta1 polymorphisms on tuberculosis. Hum.Genet 112: 237-243.

Al Arif L, Goldstein R, Affronti L, Janicki B (1979) HLA-Bw15 and tuberculosis in a North American black population. Am.Rev.Respir.Dis 120: 1275-1278.

Alcaïs A, Abel L (2004) Application of genetic epidemiology to dissecting host susceptibility/resistance to infection illustrated with the study of common mycobacterial infections. In: Bellamy R, editor. Susceptibility to infectious diseases: the importance of host genetics. Cambridge University Press. pp. 7-33.

Allen P (1924) Report of Tuberculosis Survey of the Union of South Africa. Cape Town: cape Times : .

Amirzargar A, Yalda A, Hajabolbaghi M, Khosravi F, Jabbari H et al. (2004) The association of HLA-DRB, DQA1, DQB1 alleles and haplotype frequency in Iranian patients with pulmonary tuberculosis. Int.J.Tuberc.Lung Dis 8: 1017-1021.

Anh DD, Borgdorff MW, Van LN, Lan NT, van Gorkom T et al. (2000) Mycobacterium tuberculosis Beijing genotype emerging in Vietnam. Emerg Infect Dis 6: 302-305.

Aukrust P, Yndestad A, Smith C, Ueland T, Gullestad L et al. (2007) Chemokines in cardiovascular risk prediction. Thromb Haemost 97: 748-754.

Awomoyi A, Charurat M, Marchant A, Miller E, Blackwell J et al. (2005) Polymorphism in IL1B: IL1B-511 association with tuberculosis and decreased lipopolysaccharide-induced IL-1beta in IFN-gamma primed ex-vivo whole blood assay. J.Endotoxin.Res 11: 281-286.

Awomoyi A, Marchant A, Howson J, McAdam K, Blackwell J et al. (2002) Interleukin-10, polymorphism in SLC11A1 (formerly NRAMP1), and susceptibility to tuberculosis. J.Infect.Dis 186: 1808-1814.

Awomoyi A, Nejentsev S, Richardson A, Hull J, Koch O et al. (2004) No association between interferon-gamma receptor-1 gene polymorphism and pulmonary tuberculosis in a Gambian population sample. Thorax 59: 291-294.

Awomoyi A, Sirugo G, Newport M, Tishkoff S (2006) Global distribution of a novel trinucleotide microsatellite polymorphism (ATA)n in intron 8 of the SLC11A1 gene and susceptibility to pulmonary tuberculosis. Int.J.Immunogenet 33: 11-15.

Azzawi M, Pravica V, Hasleton PS, Hutchinson IV (2001) A single nucleotide polymorphism at position --109 of the RANTES proximal promoter. Eur J Immunogenet 28: 95-96.

Babb C, van der Merwe L, Beyers N, Pheiffer C, Walzl G et al. (2007) Vitamin D receptor gene polymorphisms and sputum conversion time in pulmonary tuberculosis patients. Tuberculosis 87: 295-302.

Bacanu S, Devlin B, Roeder K (2000) The power of genomic control. Am.J.Hum.Genet 66: 1933-1944.

Baghdadi J, Orlova M, Alter A, Ranque B, Chentoufi M et al. (2006) An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. J.Exp.Med 203: 1679-1684.

Balamurugan A, Sharma S, Mehra N (2004) Human leukocyte antigen class I supertypes influence susceptibility and severity of tuberculosis. J.Infect.Dis 189: 805-811.

Balding D (2006) A tutorial on statistical methods for population association studies. Nat.Rev.Genet 7: 781-791.

Baldini A (2004) DiGeorge syndrome: an update. Curr Opin Cardiol 19: 201-204.

Bamshad M, Wooding S, Watkins W, Ostler C, Batzer M et al. (2003) Human population genetic structure and inference of group membership. Am.J.Hum.Genet 72: 578-589.

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 Medicine 3: e20.

Bellamy R (2006) Genome-wide approaches to identifying genetic factors in host susceptibility to tuberculosis. Microbes Infect 8: 1119-1123.

Bellamy R, Beyers N, McAdam K, Ruwende C, Gie R et al. (2000) Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. Proc.Natl.Acad.Sci.U.S.A 97: 8005-8009.

Bellamy R, Kwiatkowski D, Hill A (1998) Absence of an association between intercellular adhesion molecule 1, complement receptor 1 and interleukin 1 receptor antagonist gene polymorphisms and severe malaria in a West African population. Trans.R.Soc.Trop.Med.Hyg 92: 312-316.

Bellamy R, Ruwende C, Corrah T, McAdam K, Thursz M et al. (1999) Tuberculosis and chronic hepatitis B virus infection in Africans and variation in the vitamin D receptor gene. J.Infect.Dis 179: 721-724.

Bellamy R, Ruwende C, Corrah T, McAdam K, Whittle H et al. (1998) Assessment of the interleukin 1 gene cluster and other candidate gene polymorphisms in host susceptibility to tuberculosis. Tuber.Lung Dis 79: 83-89.

Bellamy R, Ruwende C, Corrah T, McAdam K, Whittle H et al. (1998) Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. N.Engl.J.Med 338: 640-644.

Bellamy R, Ruwende C, McAdam K, Thursz M, Sumiya M et al. (1998) Mannose binding protein deficiency is not associated with malaria, hepatitis B carriage nor tuberculosis in Africans. QJM 91: 13-18.

Ben Ali M, Barbouche M, Bousnina S, Chabbou A, Dellagi K (2004) Toll-like receptor 2 Arg677Trp polymorphism is associated with susceptibility to tuberculosis in Tunisian patients. Clin.Diagn.Lab Immunol 11: 625-626.

Bergholdt R, Taxvig C, Eising S, Nerup J, Pociot F (2005) CBLB variants in type 1 diabetes and their genetic interaction with CTLA4. J.Leukoc.Biol 77: 579-585.

Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN (2002) Global dissemination of the Mycobacterium tuberculosis W-Beijing family strains. Trends Microbiol 10: 45-52.

Blackwell J (2001) Genetics and genomics in infectious disease susceptibility. Trends Mol.Med 7: 521-526.

Blakemore AI, Watson PF, Weetman AP, Duff GW (1995) Association of Graves' disease with an allele of the interleukin-1 receptor antagonist gene. J Clin Endocrinol Metab 80: 111-115.

Bloom B, Murray C (1992) Tuberculosis: commentary on a reemergent killer. Science 257: 1055-1064.

Bloom B, Small P (1998) The evolving relation between humans and Mycobacterium tuberculosis. N.Engl.J.Med 338: 677-678.

Blum K, Noble E, Sheridan P, Montgomery A, Ritchie T et al. (1990) Allelic association of human dopamine D2 receptor gene in alcoholism. JAMA 263: 2055-2060.

Boas F (1942) Race, language, and culture. University of Chicago Press : .

Bonilla C, Gutierrez G, Parra E, Kline C, Shriver M (2005) Admixture analysis of a rural population of the state of Guerrero, Mexico. Am.J.Phys.Anthropol 128: 861-869.

Boonzaaier E, Malherbe C, Smith A, Berens P (1996) The cape herders: a history of the khoikhoi of southern africa. David Philip publishers, Cape Town.

Bornman L, Campbell S, Fielding K, Bah B, Sillah J et al. (2004) Vitamin D receptor polymorphisms and susceptibility to tuberculosis in West Africa: a case-control and family study. J.Infect.Dis 190: 1631-1641.

Botha MC (1972) Blood group gene frequencies. An indication of the genetic constitution of population samples in Cape Town. Am J Roentgenol Radium Ther Nucl Med 115: Suppl:1-27.

Bothamley G, Beck J, Schreuder G, D'Amaro J, de Vries R et al. (1989) Association of tuberculosis and M. tuberculosis-specific antibody levels with HLA. J.Infect.Dis 159: 549-555.

Brahmajothi V, Pitchappan R, Kakkanaiah V, Sashidhar M, Rajaram K et al. (1991) Association of pulmonary tuberculosis and HLA in south India. Tubercle 72: 123-132.

Brahmbhatt S, Black G, Carroll N, Beyers N, Salker F et al. (2006) Immune markers measured before treatment predict outcome of intensive phase tuberculosis therapy. Clin.Exp.Immunol 146: 243-252.

Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C et al. (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 99: 3684-3689.

Bryder L (1988) Below the magic mountain. a social history of tuberculosis in twentieth century. Claredon Press.

Bulat-Kardum L, Etokebe G, Knezevic J, Balen S, Matakovic-Mileusnic N et al. (2006) Interferon-gamma receptor-1 gene promoter polymorphisms (G-611A; T-56C) and susceptibility to tuberculosis. Scand.J.Immunol 63: 142-150.

Burgner D, Davila S, Breunis WB, Ng SB, Li Y et al. (2009) A genome-wide association study identifies novel and functionally related susceptibility Loci for Kawasaki disease. PLoS Genet 5: e1000319.

Burgner D, Jamieson S, Blackwell J (2006) Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better?. Lancet Infect.Dis 6: 653-663.

Butcher EC, Picker LJ (1996) Lymphocyte homing and homeostasis. Science 272: 60-66.

Buzás EI, György B, Pásztói M, Jelinek I, Falus A et al. (2006) Carbohydrate recognition systems in autoimmunity. Autoimmunity 39: 691-704.

Calafell F, Shuster A, Speed W, Kidd J, Kidd K (1998) Short tandem repeat polymorphism evolution in humans. Eur.J.Hum.Genet 6: 38-49.

Campbell S, Sabeti P, Fielding K, Sillah J, Bah B et al. (2003) Variants of the CD40 ligand gene are not associated with increased susceptibility to tuberculosis in West Africa. Immunogenetics 55: 502-507.

Cann HM, de Toma C, Cazes L, Legrand M, Morel V et al. (2002) A human genome diversity cell line panel. Science 296: 261-262.

Cao K, Moormann A, Lyke K, Masaberg C, Sumba O et al. (2004) Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. Tissue Antigens 63: 293-325.

Cardon L, Palmer L (2003) Population stratification and spurious allelic association. Lancet 361: 598-604.

Cartwright F (1977) A Social History of Medicine. New York: Longman

Casanova J, Abel L (2002) Genetic Dissection of Immunity to Mycobacteria: The Human Model. Annu.Rev.Immunol 20: 581-620.

Casanova J, Schurr E, Abel L, Skamene E (2002) Forward genetics of infectious diseases: immunological impact. Trends Immunol 23: 469-472.

Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NTN et al. (2008) The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. PLoS Pathog 4: e1000034.

Cervino A, Lakiss S, Sow O, Bellamy R, Beyers N et al. (2002) Fine mapping of a putative tuberculosis-susceptibility locus on chromosome 15q11-13 in African families. Hum.Mol.Genet 11: 1599-1603.

Cervino A, Lakiss S, Sow O, Hill A (2000) Allelic association between the NRAMP1 gene and susceptibility to tuberculosis in Guinea-Conakry. Ann.Hum.Genet 64: 507-512.

Chakraborty R, Smouse P (1988) Recombination of haplotypes leads to biased estimates of admixture proportions in human populations. Proc.Natl.Acad.Sci.U.S.A 85: 3071-3074.

Chakraborty R, Weiss K (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc.Natl.Acad.Sci.U.S.A 85: 9119-9123.

Chan M, Borgdorff M, Van L (2001) Seventy percent of the Mycobacterium tuberculosis isolates in Hong Kong represent the Beijing genotype. Epidemiol. Infect. 127: 169-171.

Chandanayingyong D, Maranetra N, Bovornkitti S (1988) HLA antigen profiles in Thai tuberculosis patients. Asian Pac.J.Allergy Immunol 6: 77-80.

Chu S, Tam CM, Wong HS, Kam KM, Lau YL et al. (2007) Association between RANTES functional polymorphisms and tuberculosis in Hong Kong Chinese. Genes Immun 8: 475-479.

Cilliers S (1963) The coloureds of south africa; a factual survey. Banier Publishers (Pty) Ltd, Cape Town.

Clay FE, Cork MJ, Tarlow JK, Blakemore AI, Harrington CI et al. (1994) Interleukin 1 receptor antagonist gene polymorphism association with lichen sclerosus. Hum Genet 94: 407-410.

Cohen T, Murray M (2004) Modeling epidemics of multidrug-resistant M. tuberculosis of heterogeneous fitness. Nat.Med 10: 1117-1121.

Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E et al. (1994) Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. JAMA 271: 698-702.

Cole S, Brosch R, Parkhill J, Garnier T, Churcher C et al. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393: 537-544.

Colobran R, Pujol-Borrell R, Armengol MP, Juan M (2007) The chemokine network. II. On how polymorphisms and alternative splicing increase the number of molecular species and configure intricate patterns of disease susceptibility. Clin Exp Immunol 150: 1-12.

Comstock G (1978) Tuberculosis in twins: a re-analysis of the Prophit survey. Am.Rev.Respir.Dis 117: 621-624.

Connell TG, Rangaka MX, Curtis N, Wilkinson RJ (2006) QuantiFERON-TB Gold: state of the art for the diagnosis of tuberculosis infection?. Expert Rev Mol Diagn 6: 663-677.

Cooke G, Campbell S, Fielding K, Sillah J, Manneh K et al. (2004) Interleukin-8 polymorphism is not associated with pulmonary tuberculosis in the Gambia. J.Infect.Dis 189: 1545-1546.

Cooke G, Campbell S, Sillah J, Gustafson P, Bah B et al. (2006) Polymorphism within the interferon-gamma/receptor complex is associated with pulmonary tuberculosis. Am.J.Respir.Crit Care Med 174: 339-343.

Cooke GS, Campbell SJ, Bennett S, Lienhardt C, McAdam KP et al. (2008) Mapping of a Novel Susceptibility Locus Suggests a Role for MC3R and CTSZ in Human Tuberculosis. Am J Respir Crit Care Med 178: 203-207.

Cooper A, Dalton D, Stewart T, Griffin J, Russell D et al. (1993) Disseminated tuberculosis in interferon gamma gene-disrupted mice. J.Exp.Med 178: 2243-2247.

Coovadia H, Benatar S (1991) A century of tuberculosis - south african perspectives. Oxford University Press, Cape Town.

Corbett EL, Marston B, Churchyard GJ, De Cock KM (2006) Tuberculosis in sub-Saharan Africa: opportunities, challenges, and change in the era of antiretroviral treatment. Lancet 367: 926-937.

Correa P, Gomez L, Cadena J, Anaya J (2005) Autoimmunity and tuberculosis. Opposite association with TNF polymorphism. J.Rheumatol 32: 219-224.

Cosar H, Ozkinay F, Onay H, Bayram N, Bakiler AR et al. (2008) Low levels of mannose-binding lectin confers protection against tuberculosis in Turkish children. Eur J Clin Microbiol Infect Dis 27: 1165-1169.

Cox R, Arnold D, Cook D, Lundberg D (1982) HLA phenotypes in Mexican Americans with tuberculosis. Am.Rev.Respir.Dis 126: 653-655.

Cox R, Downs M, Neimes R, Ognibene A, Yamashita T et al. (1988) Immunogenetic analysis of human tuberculosis. J.Infect.Dis 158: 1302-1308.

Crouch E, Wright JR (2001) Surfactant proteins a and d and pulmonary host defense. Annu Rev Physiol 63: 521-554.

Cutler G, Marshall LA, Chin N, Baribault H, Kassner PD (2007) Significant gene content variation characterizes the genomes of inbred mouse strains. Genome Res 17: 1743-1754.

Dale JW, Bothamley GH, Drobniewski F, Gillespie SH, McHugh TD et al. (2005) Origins and properties of Mycobacterium tuberculosis isolates in London. J Med Microbiol 54: 575-582.

Daly M, Rioux J, Schaffner S, Hudson T, Lander E (2001) High-resolution haplotype structure in the human genome. Nat.Genet 29: 229-232.

Daniel T (1997) Captain of Death: the story of Tuberculosis. University of Rochester Press. New York

Danis VA, Millington M, Hyland VJ, Grennan D (1995) Cytokine production by normal human monocytes: inter-subject variation and relationship to an IL-1 receptor antagonist (IL-1Ra) gene polymorphism. Clin Exp Immunol 99: 303-310.

Darvasi A, Shifman S (2005) The beauty of admixture. Nat Genet 37: 118-119.

Davey Smith G (2006) Randomised by (your) god: robust inference from an observational study design. J Epidemiol Community Health 60: 382-388.

Davey Smith G, Ebrahim S (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?. Int J Epidemiol 32: 1-22.

David HL (1971) Resistance to D-cycloserine in the tubercle bacilli: mutation rate and transport of alanine in parental cells and drug-resistant mutants. Appl Microbiol 21: 888-892.

David HL, Newman CM (1971) Some observations on the genetics of isoniazid resistance in the tubercle bacilli. Am Rev Respir Dis 104: 508-515.

Davies J (1998) Antibiotic resistance in mycobacteria. Novartis Found Symp 217: 195-205; discussion 205-8.

De la Rua-Domenech R (2006) Human Mycobacterium bovis infection in the United Kingdom: Incidence, risks, control measures and review of the zoonotic aspects of bovine tuberculosis. Tuberculosis (Edinb) 86: 77-109.

Delgado J, Baena A, Thim S, Goldfeld A (2002) Ethnic-specific genetic associations with pulmonary tuberculosis. J.Infect.Dis 186: 1463-1468.

Delgado J, Baena A, Thim S, Goldfeld A (2006) Aspartic acid homozygosity at codon 57 of HLA-DQ beta is associated with susceptibility to pulmonary tuberculosis in Cambodia. J.Immunol 176: 1090-1097.

Delves PJ, Roitt IM (2000) The immune system. Second of two parts. N Engl J Med 343: 108-117.

Deng H (2001) Population admixture may appear to mask, change or reverse genetic effects of genes underlying complex traits. Genetics 159: 1319-1323.

Deng Y, Newman B, Dunne M, Silburn P, Mellick G (2004) Case-only study of interactions between genetic polymorphisms of GSTM1, P1, T1 and Z1 and smoking in Parkinson's disease. Neurosci.Lett 19;366: 326-331.

Denis M (1991) Killing of Mycobacterium tuberculosis within human monocytes: activation by cytokines and calcitriol. Clin.Exp.Immunol 84: 200-206.

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997-1004.

Devlin B, Roeder K, Bacanu S (2001) Unbiased methods for population-based association studies. Genet.Epidemiol 21: 273-284.

Diangelo S, Lin Z, Wang G, Phillips S, Ramet M et al. (1999) Novel, non-radioactive, simple and multiplex PCR-cRFLP methods for genotyping human SP-A and SP-D marker alleles. Dis.Markers 15: 269-281.

Dinarello CA (1988) Biology of interleukin 1. FASEB J 2: 108-115.

Dlugovitzky D, Rateni L, Torres-Morales A, Ruiz-Silva J, Piñesky R et al. (1997) Levels of interleukin-8 in tuberculous pleurisy and the profile of immunocompetent cells in pleural and peripheral compartments. Immunol Lett 55: 35-39.

Dorman SE, Picard C, Lammas D, Heyne K, van Dissel JT et al. (2004) Clinical features of dominant and recessive interferon gamma receptor 1 deficiencies. Lancet 364: 2113-2121.

Dormer B, Friedlander J, Wiles F (1943) A South African team looks at tuberculosis. Reprint. Proceedings of the Transvaal Mine Medical Officers Association 23: 557.

Drobniewski F, Balabanova Y, Ruddy M, Weldon L, Jeltkova K et al. (2002) Rifampin- and multidrug-resistant tuberculosis in Russian civilians and prison inmates: dominance of the beijing strain family. Emerg Infect Dis 8: 1320-1326.

Druszczynska M, Strapagiel D, Kwiatkowska S, Kowalewicz-Kulbat M, Rozalska B et al. (2006) Tuberculosis bacilli still posing a threat. Polymorphism of genes regulating anti-mycobacterial properties of macrophages. Pol.J.Microbiol 55: 7-12.

Duan H, Zhou X, Ma Y, Li C, Chen X et al. (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.

Dubaniewicz A, Jamieson S, Dubaniewicz-Wybieralska M, Fakiola M, Nancy ME et al. (2005) Association between SLC11A1 (formerly NRAMP1) and the risk of sarcoidosis in Poland. Eur.J.Hum.Genet 13: 829-834.

Dubaniewicz A, Lewko B, Moszkowska G, Zamorska B, Stepinski J (2000) Molecular subtypes of the HLA-DR antigens in pulmonary tuberculosis. Int J.Infect.Dis 4: 129-133.

Dubaniewicz A, Moszkowska G, Szczerkowska Z, Hoppe A (2003) Analysis of DQB1 allele frequencies in pulmonary tuberculosis: preliminary report. Thorax 58: 890-891.

Dubos R, Dubos J (1952) The white plague: tuberculosis, man and society. Little, Brown & Co, Boston.

Ducati R, Ruffino-Netto A, Basso L, Santos D (2006) The resumption of consumption-- a review on tuberculosis. Mem.Inst.Oswaldo Cruz 101: 697-714.

Dupont NC, Wang K, Wadhwa PD, Culhane JF, Nelson EL (2005) Validation and comparison of luminex multiplex cytokine analysis kits with ELISA: determinations of a panel of nine cytokines in clinical sample culture supernatants. J Reprod Immunol 66: 175-191.

Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC (1999) Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. JAMA 282: 677-686.

East African/British Medical Research Council (1972) Controlled clinical trial of short-course (6 months) regimens of chemotherapy for treatment of pulmonary tuberculosis.. Lancet i: 1079-1085.

Editorial (2001) Genes, drugs and race. Nature Genetics 29: 239-240.

El Baghdadi J, Remus N, Benslimane A, El Annaz H, Chentoufi M et al. (2003) Variants of the human NRAMP1 gene and susceptibility to tuberculosis in Morocco. Int.J.Tuberc.Lung Dis 7: 599-602.

El Sahly H, Reich R, Dou S, Musser J, Graviss E (2004) The effect of mannose binding lectin gene polymorphisms on susceptibility to tuberculosis in different ethnic groups. Scand.J.Infect.Dis 36: 106-108.

Elphick R (1985) Khoikhoi and the founding of white south africa. Ravan Press, Johannesburg.

Etokebe G, Bulat-Kardum L, Johansen M, Knezevic J, Balen S et al. (2006) Interferon-gamma gene (T874A and G2109A) polymorphisms are associated with microscopy-positive tuberculosis. Scand.J.Immunol 63: 136-141.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Farhat M, Greenaway C, Pai M, Menzies D (2006) False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria?. Int J Tuberc Lung Dis 10: 1192-1204.

Ferguson J, Voelker D, McCormack F, Schlesinger L (1999) Surfactant protein D binds to Mycobacterium tuberculosis bacilli and lipoarabinomannan via carbohydrate-lectin interactions resulting in reduced phagocytosis of the bacteria by macrophages. J.Immunol 163: 312-321.

Ferguson JS, Voelker DR, Ufnar JA, Dawson AJ, Schlesinger LS (2002) Surfactant protein D inhibition of human macrophage uptake of Mycobacterium tuberculosis is independent of bacterial agglutination. J Immunol 168: 1309-1314.

Fine PE (1981) Immunogenetics of susceptibility to leprosy, tuberculosis, and leishmaniasis. An epidemiological perspective. Int J Lepr Other Mycobact Dis 49: 437-454.

Fishberg M (1922) Pulmonary tuberculosis. 3rd ed.

Fitness J, Floyd S, Warndorff D, Sichali L, Malema S et al. (2004) Large-scale candidate gene study of tuberculosis susceptibility in the Karonga district of northern Malawi. Am.J.Trop.Med.Hyg 71: 341-349.

Fitness J, Floyd S, Warndorff DK, Sichali L, Malema S et al. (2004) Large-scale candidate gene study of tuberculosis susceptibility in the Karonga district of northern Malawi. Am J Trop Med Hyg 71: 341-349.

Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O et al. (2002) Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. J Bacteriol 184: 5479-5490.

Flores-Villanueva P, Ruiz-Morales J, Song C, Flores L, Jo E et al. (2005) A functional promoter polymorphism in monocyte chemoattractant protein-1 is associated with increased susceptibility to pulmonary tuberculosis. J.Exp.Med 202: 1649-1658.

Floros J, Lin H, Garcia A, Salazar M, Guo X et al. (2000) Surfactant protein genetic marker alleles identify a subgroup of tuberculosis in a Mexican population. J.Infect.Dis 182: 1473-1478.

Foster M, Sharp R (2002) Race, ethnicity, and genomics: social classifications as proxies of biological heterogeneity. Genome Res 12: 844-850.

Frankenstein Z, Alon U, Cohen IR (2006) The immune-body cytokine network defines a social architecture of cell interactions. Biol Direct 1: 32.

Fraser D, Bulat-Kardum L, Knezevic J, Babarovic P, Matakovic-Mileusnic N et al. (2003) Interferon-gamma receptor-1 gene polymorphism in tuberculosis patients from Croatia. Scandinavian Journal of Immunology 57: 480-484.

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851-861.

Freedman M, Haiman C, Patterson N, McDonald G, Tandon A et al. (2006) Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc.Natl.Acad.Sci.U.S.A 103: 14068-14073.

Friedland JS, Hartley JC, Hartley CG, Shattock RJ, Griffin GE (1995) Inhibition of ex vivo proinflammatory cytokine secretion in fatal Mycobacterium tuberculosis infection. Clin Exp Immunol 100: 233-238.

Friedman JM, Baross A, Delaney AD, Ally A, Arbour L et al. (2006) Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. Am J Hum Genet 79: 500-513.

Futscher BW, Oshiro MM, Wozniak RJ, Holtan N, Hanigan CL et al. (2002) Role for DNA methylation in the control of cell type specific maspin expression. Nat Genet 31: 175-179.

Gabriel S, Schaffner S, Nguyen H, Moore J, Roy J et al. (2002) The structure of haplotype blocks in the human genome. Science 296: 2225-2229.

Gadzhiev B, Sergeev A, Movsum-zade K (1987) [Association of mutant alleles of serum cholinesterase with various multi-factorial and infectious diseases]. Genetika 23: 559-564.

Gagneux S, Deriemer K, Van T, Kato-Maeda M, de Jong B et al. (2006) Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc.Natl.Acad.Sci.U.S.A 103: 2869-2873.

Gandhi N, Moll A, Sturm A, Pawinski R, Govender T et al. (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.

Gao P, Fujishima S, Mao X, Remus N, Kanda M et al. (2000) Genetic variants of NRAMP1 and active tuberculosis in Japanese populations. International Tuberculosis Genetics Team. Clin.Genet 58: 74-76.

Garcia-Laorden MI, Pena MJ, Caminero JA, Garcia-Saavedra A, Campos-Herrero MI et al. (2006) Influence of mannose-binding lectin on HIV infection and tuberculosis in a Western-European population. Mol Immunol 43: 2143-2150.

García-Zepeda EA, Rojas-López A, Esquivel-Velázquez M, Ostoa-Saloma P (2007) Regulation of the inflammatory immune response by the cytokine/chemokine network in amoebiasis. Parasite Immunol 29: 679-684.

Gardai SJ, Xiao Y, Dickinson M, Nick JA, Voelker DR et al. (2003) By binding SIRPalpha or calreticulin/CD91, lung collectins act as dual function surveillance molecules to suppress or enhance inflammation. Cell 115: 13-23.

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.

Glynn J, Kremer K, Borgdorff M, Rodriquez M, van Sooligen D (2006) Beijing/W Genotype Mycobacterium tuberculosis and Drug Resistance. Emerg Infect Dis 12: 736-743.

Glynn JR, Crampin AC, Yates MD, Traore H, Mwaungulu FD et al. (2005) The importance of recent infection with Mycobacterium tuberculosis in an area with high HIV prevalence: a long-term molecular epidemiological study in Northern Malawi. J Infect Dis 192: 480-487.

Goldfeld A, Delgado J, Thim S, Bozon M, Uglialoro A et al. (1998) Association of an HLA-DQ allele with clinical tuberculosis. JAMA 279: 226-228.

Goldstein D, Hirschhorn J (2004) In genetic control of disease, does 'race' matter?. Nat.Genet 36: 1243-1244.

Goldstein DB (2009) Common genetic variation and human traits. N Engl J Med 360: 1696-1698.

Gomez L, Camargo J, Castiblanco J, Ruiz-Narvaez E, Cadena J et al. (2006) Analysis of IL1B, TAP1, TAP2 and IKBL polymorphisms on susceptibility to tuberculosis. Tissue Antigens 67: 290-296.

Gonzalez E, Dhanda R, Bamshad M, Mummidi S, Geevarghese R et al. (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.

Grange JM, Stanford JL, Rook GA (1995) Tuberculosis and cancer: parallels in host responses and therapeutic approaches?. Lancet 345: 1350-1352.

Gray R, Wheatley K (1991) How to avoid bias when comparing bone marrow transplantation with chemotherapy. Bone Marrow Transplant 7 Suppl 3: 9-12.

Greeff JM (2007) Deconstructing Jaco: genetic heritage of an Afrikaner. Ann Hum Genet 71: 674-688.

Greenwood C, Fujiwara T, Boothroyd L, Miller M, Frappier D et al. (2000) Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. Am.J.Hum.Genet 67: 405-416.

Greenwood D, Slack R, Peutherer J (1997) Medical microbiology - a guide to microbial infections: pathogenesis, immunity, laboratory diagnosis and control. Churchill Livingstone, London. 200 p.

Gruenheid S, Pinner E, Desjardins M, Gros P (1997) Natural resistance to infection with intracellular pathogens: the Nramp1 protein is recruited to the membrane of the phagosome. J.Exp.Med 185: 717-730.

Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN et al. (2006) Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. J Infect Dis 193: 121-128.

Hao K, Li C, Rosenow C, Wang WH (2004) Detect and adjust for population stratification in population-based association study using genomic control markers: an application of Affymetrix GeneChip Human Mapping 10K array. EJHG 12: 1001-1006.

Hafez M, el Salab S, el Shennawy F, Bassiony M (1985) HLA-antigens and tuberculosis in the Egyptian population. Tubercle 66: 35-40.

Hammer M, Karafet T, Redd A, Jarjanazi H, Santachiara-Benerecetti S et al. (2001) Hierarchical patterns of global human Y-chromosome diversity. Mol.Biol.Evol 18: 1189-1203.

Hanekom M, van der Spuy GD, Gey van Pittius NC, McEvoy CRE, Ndabambi SL et al. (2007) Evidence that the spread of Mycobacterium tuberculosis strains with the Beijing genotype is human population dependent. J Clin Microbiol 45: 2263-2266.

Hanis C, Hewett-Emmett D, Bertin T, Schull W (1991) Origins of US Hispanics. Implications for diabetes. Diabetes Care 14: 618-627.

Hawkins B, Higgins D, Chan S, Lowrie D, MITCHISON D et al. (1988) HLA typing in the Hong Kong Chest Service/British Medical Research Council study of factors associated with the breakdown to active tuberculosis of inactive pulmonary lesions. Am.Rev.Respir.Dis 138: 1616-1621.

Heese J (1971) Die herkoms van die afrikaner, 1657-1867. Cape Town: A Balkema.

Henao M, Montes C, Paris S, Garcia L (2006) Cytokine gene polymorphisms in Colombian patients with different clinical presentations of tuberculosis. Tuberculosis (Edinb) 86: 11-19.

Henao M, Montes C, Paris S, Garcia L (2006) Cytokine gene polymorphisms in Colombian patients with different clinical presentations of tuberculosis. Tuberculosis.(Edinb.) 86: 11-19.

Henderson RA, Watkins SC, Flynn JL (1997) Activation of human dendritic cells following infection with Mycobacterium tuberculosis. J Immunol 159: 635-643.

Hess J, Kaufmann SH (1999) Development of novel tuberculosis vaccines. C R Acad Sci III 322: 953-958.

Hill A, Ruwende C, McGuire W, Bellamy R, Coleman E et al. (1996) Association of the TNF -238 promoter polymorphism with susceptibility to tuberculosis and malaria in Africa. Human Immunology 47: 118.

Hirschhorn J, Daly M (2005) Genome-wide association studies for common diseases and complex traits. Nat.Rev.Genet 6: 95-108.

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 U S A 101: 4871-4876.

Hoal E, Lewis L, Jamieson S, Tanzer F, Rossouw M 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.

Hoal-van Helden E, Epstein J, Victor T, Hon D, Lewis L et al. (1999) Mannose-binding protein B allele confers protection against tuberculous meningitis. Pediatr.Res 45: 459-464.

Holmskov U, Thiel S, Jensenius JC (2003) Collections and ficolins: humoral lectins of the innate immune defense. Annu Rev Immunol 21: 547-578.

Hull J, Thomson A, Kwiatkowski D (2000) Association of respiratory syncytial virus bronchiolitis with the interleukin 8 gene region in UK families. Thorax 55: 1023-1027.

Huxley J (1951) Genetics, evolution and human destiny. In: Macmillan, New York. pp. 591-621.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK et al. (2004) Detection of large-scale variation in the human genome. Nat Genet 36: 949-951.

Illingworth R, Kerr A, Desousa D, Jørgensen H, Ellis P et al. (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol 6: e22.

Iwasaki A, Kelsall BL (1999) Freshly isolated Peyer's patch, but not spleen, dendritic cells produce interleukin 10 and induce the differentiation of T helper type 2 cells. J Exp Med 190: 229-239.

Jack DL, Read RC, Tenner AJ, Frosch M, Turner MW et al. (2001) Mannose-binding lectin regulates the inflammatory response of human professional phagocytes to Neisseria meningitidis serogroup B. J Infect Dis 184: 1152-1162.

Jackson AN, McLure CA, Dawkins RL, Keating PJ (2007) Mannose binding lectin (MBL) copy number polymorphism in Zebrafish (D. rerio) and identification of haplotypes resistant to L. anguillarum. Immunogenetics 59: 861-872.

Jacob SE, Nassiri M, Kerdel FA, Vincek V (2003) Simultaneous measurement of multiple Th1 and Th2 serum cytokines in psoriasis and correlation with disease severity. Mediators Inflamm 12: 309-313.

Jamieson S, Miller E, Black G, Peacock C, Cordell H 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.

Janeway C, Travers P, Walport M, Shlomchik M (2001) Immunobiology: the immune system in health and disease. Garland Science, New York.

Jankovic D, Liu Z, Gause WC (2001) Th1- and Th2-cell commitment during infectious disease: asymmetry in divergent pathways. Trends Immunol 22: 450-457.

Ji X, Gewurz H, Spear GT (2005) Mannose binding lectin (MBL) and HIV. Mol Immunol 42: 145-152.

Jindani A, Nunn AJ, Enarson DA (2004) Two 8-month regimens of chemotherapy for treatment of newly diagnosed pulmonary tuberculosis: international multicentre randomised trial. Lancet 364: 1244-1251.

Jorde L, Watkins W, Bamshad M, Dixon M, Ricker C et al. (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am.J.Hum.Genet 66: 979-988.

Kaliński P, Smits HH, Schuitemaker JH, Vieira PL, van Eijk M et al. (2000) IL-4 is a mediator of IL-12p70 induction by human Th2 cells: reversal of polarized Th2 phenotype by dendritic cells. J Immunol 165: 1877-1881.

Kamatani Y, Wattanapokayakit S, Ochi H, Kawaguchi T, Takahashi A et al. (2009) A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. Nat Genet 41: 591-595.

Kapur V, Whittam TS, Musser JM (1994) Is Mycobacterium tuberculosis 15,000 years old?. J Infect Dis 170: 1348-1349.

Kasvosve I, Gomo Z, Mvundura E, Moyo V, Saungweme T et al. (2000) Haptoglobin polymorphism and mortality in patients with tuberculosis. Int.J.Tuberc.Lung Dis 4: 771-775.

Kaufmann S, Baumann S, Nasser EA (2006) Exploiting immunology and molecular genetics for rational vaccine design against tuberculosis. Int.J.Tuberc.Lung Dis 10: 1068-1079.

Kawaguchi M, Adachi M, Oda N, Kokubu F, Huang S (2004) IL-17 cytokine family. J Allergy Clin Immunol 114: 1265-73; quiz 1274.

Keegan T (1996) Colonial south africa and the origins of the racial order. David Philip publishers, Claremont, South Afica.

Kharakter Z, Mazhak K, Pavlenko A (1990) [The role of genetically determined haptoglobin phenotypes in patients with destructive pulmonary tuberculosis]. Probl.Tuberk : 50-52.

Khomenko A, Litvinov V, Chukanova V, Pospelov L (1990) Tuberculosis in patients with various HLA phenotypes. Tubercle 71: 187-192.

Kim H, Lee Y, Lee S, Jung J, Jin H et al. (2006) Gene-gene interaction between IL-13 and IL-13Ralpha1 is associated with total IgE in Korean children with atopic asthma. J.Hum.Genet 51: 1055-1062.

Kim H, Park M, Song E, Park H, Kwon S et al. (2005a) Association of HLA-DR and HLA-DQ Genes With Susceptibility to Pulmonary Tuberculosis in Koreans: Preliminary Evidence of Associations With Drug Resistance, Disease Severity, and Disease Recurrence. Hum Immunol 66: 1074-1081.

Kim J, Verdu P, Pakstis A, Speed W, Kidd J et al. (2005b) Use of autosomal loci for clustering individuals and populations of East Asian origin. Hum.Genet 117: 511-519.

Kim L, Fijal B, Witte J (2001) Hierarchical modeling of the relation between sequence variants and a quantitative trait: addressing multiple comparison and population stratification issues. Genet.Epidemiol 21 Suppl 1: S668-S673.

Kittles R, Chen W, Panguluri R, Ahaghotu C, Jackson A et al. (2002) CYP3A4-V and prostate cancer in African Americans: causal or confounding association because of population stratification?. Hum.Genet 110: 553-560.

Knowler W, Williams R, Pettitt D, Steinberg A (1988) Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am.J.Hum.Genet 43: 520-526.

Koch R (1882) Die Aetiologie der Tuberculose. Berl Klinische Wochenschr 19: 221-230.

Koeleman B, Lie B, Undlien D, Dudbridge F, Thorsby E et al. (2004) Genotype effects and epistasis in type 1 diabetes and HLA-DQ trans dimer associations with disease. Genes Immun 5: 381-388.

Koh W, Kwon O, Kim E, Lee K, Ki C et al. (2005) NRAMP1 gene polymorphism and susceptibility to nontuberculous mycobacterial lung diseases. Chest 128: 94-101.

Kritzinger FE, den Boon S, Verver S, Enarson DA, Lombard CJ et al. (2009) No decrease in annual risk of tuberculosis infection in endemic area in Cape Town, South Africa. Trop Med Int Health 14: 136-142.

Labidi AH, Estes RC, David HL, Bollon AP (2001) Mycobacterium recombinant vaccines. Tunis Med 79: 65-81.

Lander E, Schork N (1994) Genetic dissection of complex traits. Science 265: 2037-2048.

Larsen CG, Thomsen MK, Gesser B, Thomsen PD, Deleuran BW et al. (1995) The delayed-type hypersensitivity reaction is dependent on IL-8. Inhibition of a tuberculin skin reaction by an anti-IL-8 monoclonal antibody. J Immunol 155: 2151-2157.

Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G (2008) Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 27: 1133-1163.

Lee AS, Gutiérrez-Arcelus M, Perry GH, Vallender EJ, Johnson WE et al. (2008) Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies. Hum Mol Genet 17: 1127-1136.

Lee H, Lee H, Kim D, Ko D, Han S et al. (2005) Lack of an association between interleukin-12 receptor beta1 polymorphisms and tuberculosis in Koreans. Respiration 72: 365-368.

Lee S, Kim B, Kim J, Lee S, Choi S et al. (2004) Gene-gene interaction between interleukin-4 and interleukin-4 receptor alpha in Korean children with asthma. Clin.Exp.Allergy 34: 1202-1208.

Leff A, Lester TW, Addington WW (1979) Tuberculosis. A chemotherapeutic triumph but a persistent socioeconomic problem. Arch Intern Med 139: 1375-1377.

Lennard A, Gorman P, Carrier M, Griffiths S, Scotney H et al. (1992) Cloning and chromosome mapping of the human interleukin-1 receptor antagonist gene. Cytokine 4: 83-89.

Leth-Larsen R, Garred P, Jensenius H, Meschi J, Hartshorn K et al. (2005) A common polymorphism in the SFTPD gene influences assembly, function, and concentration of surfactant protein D. J Immunol 174: 1532-1538.

Lewis S, Baker I, Davey SG (2005) Meta-analysis of vitamin D receptor polymorphisms and pulmonary tuberculosis risk. Int.J.Tuberc.Lung Dis 9: 1174-1177.

Lewontin R (1972) The apportionment of human diversity. Evol Biol 6: 381-398.

Li C, Campbell S, Kumararatne D, Bellamy R, Ruwende C et al. (2002) Association of a polymorphism in the P2X7 gene with tuberculosis in a Gambian population. J.Infect.Dis 186: 1458-1462.

Li H, Zhang T, Zhou Y, Huang Q, Huang J (2006) SLC11A1 (formerly NRAMP1) gene polymorphisms and tuberculosis susceptibility: a meta-analysis. Int.J.Tuberc.Lung Dis 10: 3-12.

Liaw Y, Tsai-Wu J, Wu C, Hung C, Lee C et al. (2002) Variations in the NRAMP1 gene and susceptibility of tuberculosis in Taiwanese. Int.J.Tuberc.Lung Dis 6: 454-460.

Lio D, Marino V, Serauto A, Gioia V, Scola L et al. (2002) Genotype frequencies of the +874T-->A single nucleotide polymorphism in the first intron of the interferon-gamma gene in a sample of Sicilian patients affected by tuberculosis. Eur.J.Immunogenet 29: 371-374.

Lipscombe R, Sumiya M, Hill A, Lau Y, Levinsky R et al. (1992) High frequencies in African and non-African populations of independent mutations in the mannose binding protein gene. Hum.Mol.Genet 1: 709-715.

Liu H, Chao D, Nakayama E, Taguchi H, Goto M et al. (1999) Polymorphism in RANTES chemokine promoter affects HIV-1 disease progression. Proc.Natl.Acad.Sci.U.S.A 96: 4581-4585.

Liu W, Cao W, Zhang C, Tian L, Wu X et al. (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.

Liu W, Zhang C, Tian L, Li C, Wu X et al. (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.

Liu W, Zhang C, Wu X, Tian L, Li C et al. (2003) [A case-control study on the vitamin D receptor gene polymorphisms and susceptibility to pulmonary tuberculosis]. Zhonghua Liu Xing.Bing.Xue.Za Zhi 24: 389-392.

Lombard Z, Brune A, Hoal E, Babb C, van Helden P et al. (2006) HLA class II disease associations in southern Africa. Tissue Antigens 67: 97-110.

Lopez-Maderuelo D, Arnalich F, Serantes R, Gonzalez A, Codoceo R et al. (2003) Interferon-gamma and interleukin-10 gene polymorphisms in pulmonary tuberculosis. Am.J.Respir.Crit Care Med 167: 970-975.

Luo Y, Chen X, O'Donnell MA (2003) Role of Th1 and Th2 cytokines in BCG-induced IFN-gamma production: cytokine promotion and simulation of BCG effect. Cytokine 21: 17-26.

Ma P, Chen D, Pan J, Du B (2002) Genomic polymorphism within interleukin-1 family cytokines influences the outcome of septic patients. Crit Care Med 30: 1046-1050.

Ma X, Dou S, Wright J, Reich R, Teeter L et al. (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.

Ma X, Reich R, Gonzalez O, Pan X, Fothergill A et al. (2003) No evidence for association between the polymorphism in the 3 ' untranslated region of interleukin-12B and human susceptibility to tuberculosis. Journal of Infectious Diseases 188: 1116-1118.

Ma X, Reich R, Wright J, Tooker H, Teeter L et al. (2003) Association between interleukin-8 gene alleles and human susceptibility to tuberculosis disease. Journal of Infectious Diseases 188: 349-355.

Ma X, Wright J, Dou S, Olsen P, Teeter L et al. (2002) Ethnic divergence and linkage disequilibrium of novel SNPs in the human NLI-IF gene: evidence of human origin and lack of association with tuberculosis susceptibility. J.Hum.Genet 47: 140-145.

Maartens G, Wilkinson RJ (2007) Tuberculosis. Lancet 370: 2030-2043.

Madan T, Saxena S, Murthy K, Muralidhar K, Sarma P (2002) Association of polymorphisms in the collagen region of human SP-A1 and SP-A2 genes with pulmonary tuberculosis in Indian population. Clin.Chem.Lab Med 40: 1002-1008.

Madhi SA, Huebner RE, Doedens L, Aduc T, Wesley D et al. (2000) HIV-1 co-infection in children hospitalised with tuberculosis in South Africa. Int J Tuberc Lung Dis 4: 448-454.

Madsen HO, Garred P, Kurtzhals JA, Lamm LU, Ryder LP et al. (1994) A new frequent allele is the missing link in the structural polymorphism of the human mannan-binding protein. Immunogenetics 40: 37-44.

Madsen HO, Garred P, Thiel S, Kurtzhals JA, Lamm LU et al. (1995) Interplay between promoter and structural gene variants control basal serum level of mannan-binding protein. J Immunol 155: 3013-3020.

Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK (1996) Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. J Bacteriol 178: 1274-1282.

Mahasirimongkol S, Yanai H, Nishida N, Ridruechai C, Matsushita I et al. (2009) Genome-wide SNP-based linkage analysis of tuberculosis in Thais. Genes Immun 10: 77-83.

Mahmoudzadeh-Niknam H, Khalili G, Fadavi P (2003) Allelic distribution of human leukocyte antigen in Iranian patients with pulmonary tuberculosis. Hum.Immunol 64: 124-129.

Mahomed H, Hughes EJ, Hawkridge T, Minnies D, Simon E et al. (2006) Comparison of mantoux skin test with three generations of a whole blood IFN-gamma assay for tuberculosis infection. Int J Tuberc Lung Dis 10: 310-316.

Malik A, Godfrey-Faussett P (2005) Effects of genetic variability of Mycobacterium tuberculosis strains on the presentation of disease. Lancet Infect.Dis 5: 174-183.

Malik S, Abel L, Tooker H, Poon A, Simkin L et al. (2005) Alleles of the NRAMP1 gene are risk factors for pediatric tuberculosis disease. Proc.Natl.Acad.Sci.U.S.A 102: 12183-12188.

Malik S, Greenwood C, Eguale T, Kifle A, Beyene J et al. (2006) Variants of the SFTPA1 and SFTPA2 genes and susceptibility to tuberculosis in Ethiopia. Hum.Genet 118: 752-759.

Manchester K (1984) Tuberculosis and leprosy in antiquity: an interpretation. Med Hist 28: 162-173.

Mao X, Bigham A, Mei R, Gutierrez G, Weiss K et al. (2007) A genomewide admixture mapping panel for Hispanic/Latino populations. Am.J.Hum.Genet 80: 1171-1178.

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L et al. (2008) Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 82: 477-488.

Martinez A, Sanchez E, Valdivia A, Orozco G, Lopez-Nevot M et al. (2006) Epistatic interaction between FCRL3 and NFkappaB1 genes in Spanish patients with rheumatoid arthritis. Ann.Rheum.Dis 65: 1188-1191.

Martinez-Marignac V, Valladares A, Cameron E, Chan A, Perera A et al. (2007) Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. Hum.Genet 120: 807-819.

Mastroianni CM, Paoletti F, Rivosecchi RM, Lancella L, Ticca F et al. (1994) Cerebrospinal fluid interleukin 8 in children with purulent bacterial and tuberculous meningitis. Pediatr Infect Dis J 13: 1008-1010.

Mathema B, Kurepina N, Bifani P, Kreiswirth B (2006) Molecular epidemiology of tuberculosis: current insights. Clin.Microbiol.Rev 19: 658-685.

Matrakshin A, Tsoi K, Pospelov L, Kapina M, Kholod O et al. (1993) [A genotypic study of children ill with tuberculosis and of healthy BCG-revaccinated ones of Tuvinian nationality]. Probl.Tuberk : 25-27.

McGarry F, Neilly J, Anderson N, Sturrock R, Field M (2001) A polymorphism within the interleukin 1 receptor antagonist (IL-1Ra) gene is associated with ankylosing spondylitis. Rheumatology (Oxford) 40: 1359-1364.

McInnes IB, Schett G (2007) Cytokines in the pathogenesis of rheumatoid arthritis. Nat Rev Immunol 7: 429-442.

McIntyre KW, Stepan GJ, Kolinsky KD, Benjamin WR, Plocinski JM et al. (1991) Inhibition of interleukin 1 (IL-1) binding and bioactivity in vitro and modulation of acute inflammation in vivo by IL-1 receptor antagonist and anti-IL-1 receptor monoclonal antibody. J Exp Med 173: 931-939.

McKeigue PM (1997) Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. Am J Hum Genet 60: 188-196.

McVicar N (1932) The prevelance of certain diseases amongst the natives of the Ciskei. South African Medical Journal 6: 721.

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

Metcalf C (1991) A history of tuberculsosis. In: Coovadia HM & Benatar SR, editor. A century of tuberculosis. South African perspectives. Oxford University Press Cape Town. pp. 1-31.

Meya DB, McAdam KPWJ (2007) The TB pandemic: an old problem seeking new solutions. J Intern Med 261: 309-329.

Meyer CG, May J, Stark K (1998) Human leukocyte antigens in tuberculosis and leprosy. Trends Microbiol 6: 148-154.

Miller E, Jamieson S, Joberty C, Fakiola M, Hudson D et al. (2004) Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians. Genes Immun 5: 63-67.

Mirsaeidi S, Houshmand M, Tabarsi P, Banoei M, Zargari L et al. (2006) Lack of association between interferon-gamma receptor-1 polymorphism and pulmonary TB in Iranian population sample. J.Infect 52: 374-377.

Modi WS, Dean M, Seuanez HN, Mukaida N, Matsushima K et al. (1990) Monocyte-derived neutrophil chemotactic factor (MDNCF/IL-8) resides in a gene cluster along with several other members of the platelet factor 4 gene superfamily. Hum Genet 84: 185-187.

Mombo L, Lu C, Ossari S, Bedjabaga I, Sica L et al. (2003) Mannose-binding lectin alleles in sub-Saharan Africans and relation with susceptibility to infections. Genes Immun 4: 362-367.

Montana G, Pritchard J (2004) Statistical tests for admixture mapping with case-control and cases-only data. Am.J.Hum.Genet 75: 771-789.

Morgan M, Kalantri S, Flores L, Pai M (2005) A commercial line probe assay for the rapid detection of rifampicin resistance in Mycobacterium tuberculosis: a systematic review and meta-analysis. BMC Infect Dis 5: 62.

Morse D (1967) Tuberculosis: disease in antiquity. Springfield. 249 p.

Mountain A (2003) The first people of the cape. David Philips Publishers, Claremont, South Africa.

Mountain A (2004) An unsung heritage. David Philip Publishers, Caremont, South Africa.

Mukaida N, Okamoto S, Ishikawa Y, Matsushima K (1994) Molecular mechanism of interleukin-8 gene expression. J Leukoc Biol 56: 554-558.

Mullighan CG, Marshall SE, Welsh KI (2000) Mannose binding lectin polymorphisms are associated with early age of disease onset and autoimmunity in common variable immunodeficiency. Scand J Immunol 51: 111-122.

Munch Z, Van Lill S, Booysen C, Zietsman H, Enarson D et al. (2003) Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. International Journal of Tuberculosis and Lung Disease 7: 271-277.

Murphy P (1993) Molecular mimicry and the generation of host defense protein diversity. Cell 72: 823-826.

Murray C, Styblo K, Rouillon A (1990) Tuberculosis in developing countries: burden, intervention and cost. Bull.Int.Union Tuberc.Lung Dis 65: 6-24.

Murray CJ, Salomon JA (1998) Modeling the impact of global tuberculosis control strategies. Proc Natl Acad Sci U S A 95: 13881-13886.

Murray P, Barron E, Pfaller M, Tenover F, Yolken R (1999) Manual of clinical microbiology. American Society of Microbiology, Washington.

Murtha AP, Nieves A, Hauser ER, Swamy GK, Yonish BA et al. (2006) Association of maternal IL-1 receptor antagonist intron 2 gene polymorphism and preterm birth. Am J Obstet Gynecol 195: 1249-1253.

Mwantembe O, Gaillard MC, Barkhuizen M, Pillay V, Berry SD et al. (2001) Ethnic differences in allelic associations of the interleukin-1 gene cluster in South African patients with inflammatory bowel disease (IBD) and in control individuals. Immunogenetics 52: 249-254.

Nagel R (2005) Epistasis and the genetics of human diseases. C.R.Biol 328: 606-615.

Newport M, Allen A, Awomoyi A, Dunstan S, McKinney E et al. (2004) The toll-like receptor 4 Asp299Gly variant: no influence on LPS responsiveness or susceptibility to pulmonary tuberculosis in The Gambia. Tuberculosis.(Edinb.) 84: 347-352.

Newton C, Graham A, Heptinstall L, Powell S, Summers C et al. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res 17: 2503-2516.

Nickel R, Casolaro V, Wahn U, Beyer K, Barnes K et al. (2000) Atopic dermatitis is associated with a functional mutation in the promoter of the C-C chemokine RANTES. J.Immunol 164: 1612-1616.

Niimi T, Sato S, Sugiura Y, Yoshinouchi T, Akita K et al. (2002) Transforming growth factor-beta gene polymorphism in sarcoidosis and tuberculosis patients. Int.J.Tuberc.Lung Dis 6: 510-515.

Ninomiya S, Niimi T, Shimizu S, Sato S, Achiwa H et al. (2004) Matrix metalloproteinase-1 polymorphism of promoter region in sarcoidosis and tuberculosis patients. Sarcoidosis.Vasc.Diffuse.Lung Dis 21: 19-24.

Nolan CM (1997) Nosocomial multidrug-resistant tuberculosis--global spread of the third epidemic. J Infect Dis 176: 748-751.

Nurse G, Weiner J, Jenkins T (1985) The peoples of southern africa and their affinities. Clarendon Press, Oxford. 79.

O'Connor B (2000) SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behavior Research Methods, Instruments, & Computers 32: 396-402.

O'Hayre M, Salanga CL, Handel TM, Allen SJ (2008) Chemokines and cancer: migration, intracellular signalling and intercellular communication in the microenvironment. Biochem J 409: 635-649.

Ogus A, Yoldas B, Ozdemir T, Uguz A, Olcen S et al. (2004) The Arg753GLn polymorphism of the human toll-like receptor 2 gene in tuberculosis disease. Eur.Respir.J 23: 219-223.

Orme IM (2001) The search for new vaccines against tuberculosis. J Leukoc Biol 70: 1-10.

Ottenhoff TH, Kumararatne D, Casanova JL (1998) Novel human immunodeficiencies reveal the essential role of type-I cytokines in immunity to intracellular bacteria. Immunol Today 19: 491-494.

Pace E, Gjomarkaj M, Melis M, Profita M, Spatafora M et al. (1999) Interleukin-8 induces lymphocyte chemotaxis into the pleural space. Role of pleural macrophages. Am J Respir Crit Care Med 159: 1592-1599.

Pacheco AG, Cardoso CC, Moraes MO (2008) IFNG +874T/A, IL10 -1082G/A and TNF -308G/A polymorphisms in association with tuberculosis susceptibility: a meta-analysis study. Hum Genet 123: 477-484.

Pacheco E, Fonseca C, Montes C, Zabaleta J, Garcia L et al. (2004) CD14 gene promoter polymorphism in different clinical forms of tuberculosis. FEMS Immunol.Med.Microbiol 40: 207-213.

Packard R (1989) White plague, black labour:tuberculosis and the political economy of health and disease in South Africa. Berkeley: University of California Press.

Pai M, Gokhale K, Joshi R, Dogra S, Kalantri S et al. (2005) Mycobacterium tuberculosis infection in health care workers in rural India: comparison of a whole-blood interferon gamma assay with tuberculin skin testing. JAMA 293: 2746-2755.

Pankratz N, Nichols W, Uniacke S, Halter C, Murrell J et al. (2003) Genome-wide linkage analysis and evidence of gene-by-gene interactions in a sample of 362 multiplex Parkinson disease families. Hum.Mol.Genet 12: 2599-2608.

Papanicolaou GJ, Parsa NZ, Meltzer PS, Trent JM (1997) Assignment of interferon gamma receptor (IFNGR1) to human chromosome bands 6q24.1-->q24.2 by in situ hybridization. Cytogenet Cell Genet 76: 181-182.

Papiha S, Agarwal S, White I (1983) Association between phosphoglucomutase (PGM1) and group-specific component (Gc) subtypes and tuberculosis. J.Med.Genet 20: 220-222.

Papiha S, Singh B, Lanchbury J, Roberts D, Parsad C et al. (1987) Association of HLA and other genetic markers in South Indian patients with pulmonary tuberculosis. Tubercle 68: 159-167.

Parker H, Kim L, Sutter N, Carlson S, Lorentzen T et al. (2004) Genetic structure of the purebred domestic dog. Science 304: 1160-1164.

Pasqualoto KF, Ferreira EI (2001) An approach for the rational design of new antituberculosis agents. Curr Drug Targets 2: 427-437.

Patil N, Berno A, Hinds D, Barrett W, Doshi J et al. (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294: 1719-1723.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2: e190.

Pennini M, Pai R, Schultz D, Boom W, Harding C (2006) Mycobacterium tuberculosis 19-kDa lipoprotein inhibits IFN-gamma-induced chromatin remodeling of MHC2TA by TLR2 and MAPK signaling. J.Immunol 176: 4323-4330.

Pfaff C, Parra E, Bonilla C, Hiester K, McKeigue P et al. (2001) Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am.J.Hum.Genet 68: 198-207.

Pospelov L, Matrakshin A, Chernousova L, Tsoi K, Afanasjev K et al. (1996) Association of various genetic markers with tuberculosis and other lung diseases in Tuvinian children. Tuber.Lung Dis 77: 77-80.

Pospelova L, Matrashkin A, Larionova E, Eremeev V, Mes'ko E (2005) [The association of tuberculosis with the specificities of the HLA gene DRB1 in different regions of Tuva]. Probl.Tuberk.Bolezn.Legk : 23-25.

Pravica V, Perrey C, Stevens A, Lee J, Hutchinson I (2000) A single nucleotide polymorphism in the first intron of the human IFN-gamma gene: absolute correlation with a polymorphic CA microsatellite marker of high IFN-gamma production. Hum.Immunol 61: 863-866.

Presanis JS, Kojima M, Sim RB (2003) Biochemistry and genetics of mannan-binding lectin (MBL). Biochem Soc Trans 31: 748-752.

Price AL, Patterson N, Yu F, Cox DR, Waliszewska A et al. (2007) A genomewide admixture map for Latino populations. Am J Hum Genet 80: 1024-1036.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904-909.

Pritchard J, Donnelly P (2001) Case-control studies of association in structured or admixed populations. Theor.Popul.Biol 60: 227-237.

Pritchard J, Rosenberg N (1999) Use of unlinked genetic markers to detect population stratification in association studies. Am.J.Hum.Genet 65: 220-228.

Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Prodinger WM, Bunyaratvej P, Prachaktam R, Pavlic M (2001) Mycobacterium tuberculosis isolates of Beijing genotype in Thailand. Emerg Infect Dis 7: 483-484.

Puzyrev V, Freidin M, Rudko A, Strelis A, Kolokolova O (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.

Race ethnicity and genetics working group (2005) The use of racial, ethnic, and ancestral categories in human genetics research. Am J Hum Genet 77: 519-532.

Rad ME, Bifani P, Martin C, Kremer K, Samper S et al. (2003) Mutations in putative mutator genes of Mycobacterium tuberculosis strains of the W-Beijing family. Emerg Infect Dis 9: 838-845.

Raja A (2004) Immunology of tuberculosis. Indian J Med Res 120: 213-232.

Rajalingam R, Mehra N, Jain R, Myneedu V, Pande J (1996) Polymerase chain reaction--based sequence-specific oligonucleotide hybridization analysis of HLA class II antigens in pulmonary tuberculosis: relevance to chemotherapy and disease severity. J.Infect.Dis 173: 669-676.

Rajalingam R, Mehra N, Singal D (2000) Polymorphism in heat-shock protein 70-1 (HSP70-1) gene promoter region and susceptibility to tuberculoid leprosy and pulmonary tuberculosis in Asian Indians. Indian J.Exp.Biol 38: 658-662.

Rajalingam R, Singal D, Mehra N (1997) Transporter associated with antigen-processing (TAP) genes and susceptibility to tuberculoid leprosy and pulmonary tuberculosis. Tissue Antigens 49: 168-172.

Ravikumar M, Dheenadhayalan V, Rajaram K, Lakshmi S, Kumaran P et al. (1999) Associations of HLA-DRB1, DQB1 and DPB1 alleles with pulmonary tuberculosis in south India. Tuber.Lung Dis 79: 309-317.

Raza S, Robertson K, Lacaze P, Page D, Enright A et al. (2008) A logic-based diagram of signalling pathways central to macrophage activation. BMC Systems Biology 2: doi:10.1186/1752-0509-2-36.

Reich D, Goldstein D (2001) Detecting association in a case-control study while correcting for population stratification. Genet.Epidemiol 20: 4-16.

Reich D, Patterson N, De Jager PL, McDonald GJ, Waliszewska A et al. (2005) A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. Nat Genet 37: 1113-1118.

Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nat Rev Genet 2: 21-32.

Relethford J, Jorde L (1999) Genetic evidence for larger African population size during recent human evolution. Am.J.Phys.Anthropol 108: 251-260.

Richards E (2006) Inherited epigenetic variation--revisiting soft inheritance. Nat.Rev.Genet 7: 395-401.

Rider LG, Artlett CM, Foster CB, Ahmed A, Neeman T et al. (2000) Polymorphisms in the IL-1 receptor antagonist gene VNTR are possible risk factors for juvenile idiopathic inflammatory myopathies. Clin Exp Immunol 121: 47-52.

Riley LW (1993) Drug-resistant tuberculosis. Clin Infect Dis 17 Suppl 2: S442-6.

Risch N, Burchard E, Ziv E, Tang H (2002) Categorization of humans in biomedical research: genes, race and disease. Genome Biol 3: comment2007.

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273: 1516-1517.

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD et al. (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69: 138-147.

Roach DR, Bean AGD, Demangel C, France MP, Briscoe H et al. (2002) TNF regulates chemokine induction essential for cell recruitment, granuloma formation, and clearance of mycobacterial infection. J Immunol 168: 4620-4627.

Roberts M, Mock B, Blackwell J (1993) Mapping of genes controlling Leishmania major infection in CXS recombinant inbred mice. Eur.J.Immunogenet 20: 349-362.

Rockett K, Brookes R, Udalova I, Vidal V, Hill A et al. (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.

Roelsgaard E, Iversen E, Blocher C (1964) Tuberculosis in Tropical Africa.. Bull World Health Organ 30: 459-518.

Roh E, Park M, Yoon J, Kim M, Kwon S et al. (2004) Association of TAP1 and TAP2 gene polymorphisms with tuberculosis in Koreans. Human Immunology 65: S35.

Romualdi C, Balding D, Nasidze I, Risch G, Robichaux M et al. (2002) Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. Genome Res 12: 602-612.

Rook G, Steele J, Fraher L, Barker S, Karmali R et al. (1986) Vitamin D3, gamma interferon, and control of proliferation of Mycobacterium tuberculosis by human monocytes. Immunology 57: 159-163.

Rosenberg N (2004) DISTRUCT: a program fro the graphical display of population structure.. Molecular Ecology Notes 4: 137-138.

Rosenberg N, Pritchard J, Weber J, Cann H, Kidd K et al. (2002) Genetic structure of human populations. Science 298: 2381-2385.

Rosenberg NA, Li LM, Ward R, Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 73: 1402-1422.

Rosenzweig S, Schaffer A, Ding L, Sullivan R, Enyedi B et al. (2004) Interferon-gamma receptor 1 promoter polymorphisms: population distribution and functional implications. Clin.Immunol 112: 113-119.

Rossouw M, Nel H, Cooke G, van Helden P, Hoal E (2003) Association between tuberculosis and a polymorphic NFkappaB binding site in the interferon gamma gene. Lancet 361: 1871-1872.

Roth D, Soto G, Arenas F, Bautista C, Ortiz J et al. (2004) Association between Vitamin D Receptor Gene Polymorphisms and Response to Treatment of Pulmonary Tuberculosis. J.Infect.Dis 190: 920-927.

Ruggiero G, Cosentini E, Zanzi D, Sanna V, Terrazzano G et al. (2004) Allelic distribution of human leucocyte antigen in historical and recently diagnosed tuberculosis patients in Southern Italy. Immunology 111: 318-322.

Ryu S, Park Y, Bai G, Kim S, Park S et al. (2000) 3'UTR polymorphisms in the NRAMP1 gene are associated with susceptibility to tuberculosis in Koreans. Int.J.Tuberc.Lung Dis 4: 577-580.

Ryu Y, Kim E, Koh W, Kim H, Kwon O et al. (2006) Toll-like receptor 2 polymorphisms and nontuberculous mycobacterial lung diseases. Clin.Vaccine Immunol 13: 818-819.

Sadek M, Sada E, Toossi Z, Schwander S, Rich E (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.

Sanjeevi C, Narayanan P, Prabakar R, Charles N, Thomas B et al. (1992) No association or linkage with HLA-DR or -DQ genes in south Indians with pulmonary tuberculosis. Tuber.Lung Dis 73: 280-284.

Santtila S, Savinainen K, Hurme M (1998) Presence of the IL-1RA allele 2 (IL1RN*2) is associated with enhanced IL-1beta production in vitro. Scand J Immunol 47: 195-198.

Satten G, Flanders W, Yang Q (2001) Accounting for unmeasured population substructure in case-control studies of genetic association using a novel latent-class model. Am.J.Hum.Genet 68: 466-477.

Sauty A, Dziejman M, Taha RA, Iarossi AS, Neote K et al. (1999) The T cell-specific CXC chemokines IP-10, Mig, and I-TAC are expressed by activated human bronchial epithelial cells. J Immunol 162: 3549-3558.

Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE et al. (2007) Challenges and standards in integrating surveys of structural variation. Nat Genet 39: S7-15.

Schoenborn JR, Wilson CB (2007) Regulation of interferon-gamma during innate and adaptive immune responses. Adv Immunol 96: 41-101.

Schwartz RS (2001) Racial profiling in medical research. N Engl J Med 344: 1392-1393.

Scola L, Crivello A, Marino V, Gioia V, Serauto A et al. (2003) IL-10 and TNF-alpha polymorphisms in a sample of Sicilian patients affected by tuberculosis: implication for ageing and life span expectancy. Mech.Ageing Dev 124: 569-572.

Searle S, Blackwell J (1999) Evidence for a functional repeat polymorphism in the promoter of the human NRAMP1 gene that correlates with autoimmune versus infectious disease susceptibility. J.Med.Genet 36: 295-299.

Seldin MF (2007) Admixture mapping as a tool in gene discovery. Curr Opin Genet Dev 17: 177-181.

Selvaraj P, Alagarasu K, Harishankar M, Vidyarani M, Nisha Rajeswari D et al. (2008) Cytokine gene polymorphisms and cytokine levels in pulmonary tuberculosis. Cytokine 43: 26-33.

Selvaraj P, Kurian S, Chandra G, Reetha A, Charles N et al. (2004) Vitamin D receptor gene variants of BsmI, ApaI, TaqI, and FokI polymorphisms in spinal tuberculosis. Clin.Genet 65: 73-76.

Selvaraj P, Narayanan P, Reetha A (1999) Association of functional mutant homozygotes of the mannose binding protein gene with susceptibility to pulmonary tuberculosis in India. Tuber.Lung Dis 79: 221-227.

Selvaraj P, Narayanan P, Reetha A (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.

Selvaraj P, Sriram U, Mathan KS, Reetha A, Narayanan P (2001) Tumour necrosis factor alpha (-238 and -308) and beta gene polymorphisms in pulmonary tuberculosis: haplotype analysis with HLA-A, B and DR genes. Tuberculosis.(Edinb.) 81: 335-341.

Selvaraj P, Uma H, Reetha A, Kurian S, Xavier T et al. (1998) HLA antigen profile in pulmonary tuberculosis patients & their spouses. Indian J.Med.Res 107: 155-158.

Selvaraj P, Uma H, Reetha AM, Xavier T, Prabhakar R et al. (1998) Influence of HLA-DR2 phenotype on humoral immunity & lymphocyte response to Mycobacterium tuberculosis culture filtrate antigens in pulmonary tuberculosis. Indian J Med Res 107: 208-217.

Selwyn P, Hartel D, Lewis V, Schoenbaum E, Vermund S et al. (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.

Selwyn PA, Sckell BM, Alcabes P, Friedland GH, Klein RS et al. (1992) High risk of active tuberculosis in HIV-infected drug users with cutaneous anergy. JAMA 268: 504-509.

Shahbazian MD, Grunstein M (2007) Functions of site-specific histone acetylation and deacetylation. Annu Rev Biochem 76: 75-100.

Shaw M, Collins A, Peacock C, Miller E, Black G et al. (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.

Shell R (1994) Children of bondage. Witwatersrand University Press, Johannesburg. p.

Shibasaki M, Yagi T, Yatsuya H, Okamoto M, Nishikawa M et al. (2009) An influence of Interferon-gamma gene polymorphisms on treatment response to tuberculosis in Japanese population. J Infect 58: 467-469.

Siawaya J, Roberts T, Babb C, Black G, Golakai H et al. (2008) An evaluation of commerical fluorescent bead-based luminex cytokine assays. Plos one 3: e2535.

Siawaya JFD, Bapela NB, Ronacher K, Beyers N, van Helden P et al. (2008) Differential expression of interleukin-4 (IL-4) and IL-4 delta 2 mRNA, but not transforming growth factor beta (TGF-beta), TGF-beta RII, Foxp3, gamma interferon, T-bet, or GATA-3 mRNA, in patients with fast and slow responses to antituberculosis treatment. Clin Vaccine Immunol 15: 1165-1170.

Simeoni E, Winkelmann B, Hoffmann M, Fleury S, Ruiz J et al. (2004) Association of RANTES G-403A gene polymorphism with increased risk of coronary arteriosclerosis. Eur.Heart J 25: 1438-1446.

Singh J, Upshur R, Padayatchi N (2007) XDR-TB in South Africa: No Time for Denial or Complacency. PLoS.Med 4: e50.

Singh S, Mehra N, Dingley H, Pande J, Vaidya M (1983) Human leukocyte antigen (HLA)-linked control of susceptibility to pulmonary tuberculosis and association with HLA-DR types. J.Infect.Dis 148: 676-681.

Singh S, Mehra N, Dingley H, Pande J, Vaidya M (1984) HLA haplotype segregation study in multiple case families of pulmonary tuberculosis. Tissue Antigens 23: 84-86.

Skwor T, Sedberry AS, Mackie J, Russell K, Berghman L et al. (2006) BCG vaccination of guinea pigs modulates Mycobacterium tuberculosis-induced CCL5 (RANTES) production in vitro and in vivo. Tuberculosis.(Edinb.) 86: 419-429.

Smedley A (1999) Race in North America: origin and evolution of a worldview, 2nd ed.. Westview Press.

Soborg C, Andersen A, Madsen H, Kok-Jensen A, Skinhoj P et al. (2002) Natural resistance-associated macrophage protein 1 polymorphisms are associated with microscopy-positive tuberculosis. J.Infect.Dis 186: 517-521.

Soborg C, Madsen H, Andersen A, Lillebaek T, Kok-Jensen A et al. (2003) Mannose-binding lectin polymorphisms in clinical tuberculosis. J.Infect.Dis 188: 777-782.

Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T et al. (2005) Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc Natl Acad Sci U S A 102: 3336-3341.

Sonnenberg P, Glynn JR, Fielding K, Murray J, Godfrey-Faussett P et al. (2005) How soon after infection with HIV does the risk of tuberculosis start to increase? A retrospective cohort study in South African gold miners. J Infect Dis 191: 150-158.

Sorensen T, Nielsen G, Andersen P, Teasdale T (1988) Genetic and environmental influences on premature death in adult adoptees. N.Engl.J.Med 318: 727-732.

Spielman R, McGinnis R, Ewens W (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am.J.Hum.Genet 52: 506-516.

Springer TA (1995) Traffic signals on endothelium for lymphocyte recirculation and leukocyte emigration. Annu Rev Physiol 57: 827-872.

Sreevatsan S, Pan X, Stockbauer K, Connell N, Kreiswirth B et al. (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.

Sriram U, Selvaraj P, Kurian S, Reetha A, Narayanan P (2001) HLA-DR2 subtypes & immune responses in pulmonary tuberculosis. Indian J.Med.Res 113: 117-124.

Stead W, Senner J, Reddick W, Lofgren J (1990) Racial differences in susceptibility to infection by Mycobacterium tuberculosis. N.Engl.J.Med 322: 422-427.

Stefansson H, Rujescu D, Cichon S, Pietiläinen OPH, Ingason A et al. (2008) Large recurrent microdeletions associated with schizophrenia. Nature 455: 232-236.

Stein CM, Zalwango S, Malone LL, Won S, Mayanja-Kizza H et al. (2008) Genome scan of M. tuberculosis infection and disease in Ugandans. PLoS ONE 3: e4094.

Stephens J, Briscoe D, O'Brien S (1994) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am.J.Hum.Genet 55: 809-824.

Sterne J, Rodrigues L, Guedes I (1998) Does the efficacy of BCG decline with time since vaccination?. Int.J.Tuberc.Lung Dis 2: 200-207.

Stockton J, Howson J, Awomoyi A, McAdam K, Blackwell J et al. (2004) Polymorphism in NOD2, Crohn's disease, and susceptibility to pulmonary tuberculosis. FEMS Immunol.Med.Microbiol 41: 157-160.

Stoneking M, Fontius J, Clifford S, Soodyall H, Arcot S et al. (1997) Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7: 1061-1071.

Streicher EM, Victor TC, van der Spuy G, Sola C, Rastogi N et al. (2007) Spoligotype signatures in the Mycobacterium tuberculosis complex. J Clin Microbiol 45: 237-240.

Stumbles PA, Thomas JA, Pimm CL, Lee PT, Venaille TJ et al. (1998) Resting respiratory tract dendritic cells preferentially stimulate T helper cell type 2 (Th2) responses and require obligatory cytokine signals for induction of Th1 immunity. J Exp Med 188: 2019-2031.

Styblo K (1988) Overview and epidemiological assessment of the current global tuberculosis situation: with an emphasis on tuberculosis control in developing countries. Bull Int Union Tuberc Lung Dis 63: 39-44.

Sumiya M, Super M, Tabona P, Levinsky RJ, Arai T et al. (1991) Molecular basis of opsonic defect in immunodeficient children. Lancet 337: 1569-1570.

Summerfield JA, Sumiya M, Levin M, Turner MW (1997) Association of mutations in mannose binding protein gene with childhood infection in consecutive hospital series. BMJ 314: 1229-1232.

Swan P (1985) The romantic's death. Nursing Times 81: 47-49.

Szekanecz Z, Koch AE (2007) Mechanisms of Disease: angiogenesis in inflammatory diseases. Nat Clin Pract Rheumatol 3: 635-643.

Sørensen AL, Nagai S, Houen G, Andersen P, Andersen AB (1995) Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis. Infect Immun 63: 1710-1717.

Sørensen GL, Hjelmborg JVB, Kyvik KO, Fenger M, Høj A et al. (2006) Genetic and environmental influences of surfactant protein D serum levels. Am J Physiol Lung Cell Mol Physiol 290: L1010-7.

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.

Tarlow JK, Blakemore AI, Lennard A, Solari R, Hughes HN et al. (1993) Polymorphism in human IL-1 receptor antagonist gene intron 2 is caused by variable numbers of an 86-bp tandem repeat. Hum Genet 91: 403-404.

Taub DD, Proost P, Murphy WJ, Anver M, Longo DL et al. (1995) Monocyte chemotactic protein-1 (MCP-1), -2, and -3 are chemotactic for human T lymphocytes. J Clin Invest 95: 1370-1376.

Taype C, Castro J, Accinelli R, Herrera-Velit P, Shaw M et al. (2006) Association between SLC11A1 polymorphisms and susceptibility to different clinical forms of tuberculosis in the Peruvian population. Infect.Genet.Evol 6: 361-367.

Telenti A, Iseman M (2000) Drug-resistant tuberculosis: what do we do now?. Drugs 59: 171-179.

Teran-Escandon D, Teran-Ortiz L, Camarena-Olvera A, Gonzalez-Avila G, Vaca-Marin M et al. (1999) Human leukocyte antigen-associated susceptibility to pulmonary tuberculosis: molecular analysis of class II alleles by DNA amplification and oligonucleotide hybridization in Mexican patients. Chest 115: 428-433.

The International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437: 1299-1320.

Thomas D, Witte J (2002) Point: population stratification: a problem for case-control studies of candidate-gene associations?. Cancer Epidemiol.Biomarkers Prev 11: 505-512.

Thorp C (2000) Hunter-gatherers and farmers: an enduring frontier in the Caledon valley, South Africa. Publishers of British Archaeological Reports.

Tishkoff S, Pakstis A, Stoneking M, Kidd J, stro-Bisol G et al. (2000) Short tandem-repeat polymorphism/alu haplotype variation at the PLAT locus: implications for modern human origins. Am.J.Hum.Genet 67: 901-925.

Tishkoff S, Williams S (2002) Genetic analysis of African populations: human evolution and complex disease. Nat.Rev.Genet 3: 611-621.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A et al. (2009) The genetic structure and history of Africans and African Americans. Science 324: 1035-1044.

Tosh K, Campbell S, Fielding K, Sillah J, Bah B et al. (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.

Tsai C, Fallin D, Chiang F, Hwang J, Lai L et al. (2003) Angiotensinogen gene haplotype and hypertension: interaction with ACE gene I allele. Hypertension 41: 9-15.

Tsai F, Hsieh Y, Chang C, Lin C, Tsai C (2002) Polymorphisms for interleukin 1 beta exon 5 and interleukin 1 receptor antagonist in Taiwanese children with febrile convulsions. Arch Pediatr Adolesc Med 156: 545-548.

Tso H, Ip W, Chong W, Tam C, Chiang A et al. (2005) Association of interferon gamma and interleukin 10 genes with tuberculosis in Hong Kong Chinese. Genes Immun 6: 358-363.

Turakulov R and Easteal S (2003) Number of SNPs loci needed to detect population structure. Hum Hered 55: 37-45.

Turner M, Johnson M, Booth C, Klein N, Rolland J et al. (2003) Assays for human mannose-binding lectin. Journal of Immunological Methods 276: 147-149.

Turner MW (2003) The role of mannose-binding lectin in health and disease. Mol Immunol 40: 423-429.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazón Lahr M et al. (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65: 43-62.

Union of South Africa (1914) Report of the Tuberculosis Comission 1914. Cape Town: Cape Times.

Van der Ross R (1993) 100 questions about Coloured South Africans. UWC Printing Department, Cape Town.

Van der Spuy GD, Kremer K, Ndabambi SL, Beyers N, Dunbar R et al. (2009) Changing Mycobacterium tuberculosis population highlights clade-specific pathogenic characteristics. Tuberculosis (Edinb) 89: 120-125.

Van Helden PD, Donald PR, Victor TC, Schaaf HS, Hoal EG et al. (2006) Antimicrobial resistance in tuberculosis: an international perspective. Expert Rev Anti Infect Ther 4: 759-766.

Van Soolingen D, Qian L, de Haas P (1995) Predominance of a single genotype of Mycobacterium tuberculosis in countries of east Asia. J Clin Microbiol 33: 3234-3238.

Vejbaesya S, Chierakul N, Luangtrakool K, Srinak D, Stephens H (2002) Associations of HLA class II alleles with pulmonary tuberculosis in Thais. Eur.J.Immunogenet 29: 431-434.

Verver S, Warren R, Beyers N, Richardson M, van der Spuy G 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.

Victor T, Streicher E, Kewley C, Jordaan A, van der Spuy G et al. (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.

Vidyarani M, Selvaraj P, Prabhu AS, Jawahar M, Adhilakshmi A et al. (2006) Interferon gamma (IFNgamma) & interleukin-4 (IL-4) gene variants & cytokine levels in pulmonary tuberculosis. Indian J.Med.Res 124: 403-410.

Wacholder S, Rothman N, Caporaso N (2000) Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J.Natl.Cancer Inst 92: 1151-1158.

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.

Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB et al. (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320: 539-543.

Wang J, Song C, Wang S (2001) [Association of HLA-DRB1 genes with pulmonary tuberculosis]. Zhonghua Jie.He.He.Hu Xi.Za Zhi 24: 302-305.

Wang L, Turner MO, Elwood RK, Schulzer M, FitzGerald JM (2002) A meta-analysis of the effect of Bacille Calmette Guérin vaccination on tuberculin skin test measurements. Thorax 57: 804-809.

Wang Y, Leung FCC (2004) An evaluation of new criteria for CpG islands in the human genome as gene markers. Bioinformatics 20: 1170-1177.

Warren R, Victor T, Streicher E, Richardson M, Beyers N 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.

Warren RM, Streicher EM, Sampson SL, van der Spuy GD, Richardson M et al. (2002) Microevolution of the direct repeat region of Mycobacterium tuberculosis: implications for interpretation of spoligotyping data. J Clin Microbiol 40: 4457-4465.

Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S et al. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 39: 457-466.

Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678.

Weyer K, Groenewald P, Zwarenstein M, Lombard CJ (1995) Tuberculosis drug resistance in the Western Cape. S Afr Med J 85: 499-504.

Wilkinson KA, Wilkinson RJ, Pathan A, Ewer K, Prakash M et al. (2005) Ex vivo characterization of early secretory antigenic target 6-specific T cells at sites of active disease in pleural tuberculosis. Clin Infect Dis 40: 184-187.

Wilkinson R, Llewelyn M, Toossi Z, Patel P, Pasvol G 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.

Wilkinson R, Patel P, Llewelyn M, Hirsch C, Pasvol G et al. (1999) Influence of polymorphism in the genes for the interleukin (IL)-1 receptor antagonist and IL-1beta on tuberculosis. J.Exp.Med 189: 1863-1874.

Williams SM, Addy JH, Phillips JA3, Dai M, Kpodonu J et al. (2000) Combinations of variations in multiple genes are associated with hypertension. Hypertension 36: 2-6.

Wilson J, Weale M, Smith A, Gratrix F, Fletcher B et al. (2001) Population genetic structure of variable drug response. Nat.Genet 29: 265-269.

Witkin SS, Gerber S, Ledger WJ (2002) Influence of interleukin-1 receptor antagonist gene polymorphism on disease. Clin Infect Dis 34: 204-209.

Wood R, Maartens G, Lombard CJ (2000) Risk factors for developing tuberculosis in HIV-1-infected adults from communities with a low or very high incidence of tuberculosis. J Acquir Immune Defic Syndr 23: 75-80.

World Health Organization (2001) WHO report 2001:Guidelines for drug susceptibility testing for second-line anti-tuberculosis drugs for DOTS-plus. WHO/HTM/TB/2001.

World Health Organization (2006) WHO Report 2006: Global tuberculosis control - surveillance, planning, financing. WHO/HTM/TB/2006.362.

World Health Organization (2009) WHO report 2009: Global Tuberculosis Control 2009 - epidemiology, strategy, financing. WHO/HTM/TB/2009.411.

Wu H, Kuzmenko A, Wan S, Schaffer L, Weiss A et al. (2003) Surfactant proteins A and D inhibit the growth of Gram-negative bacteria by increasing membrane permeability. J Clin Invest 111: 1589-1602.

Wyllie S, Seu P, Goss JA (2002) The natural resistance-associated macrophage protein 1 Slc11a1 (formerly Nramp1) and iron metabolism in macrophages. Microbes Infect 4: 351-359.

Yew W, Leung C (2005) Are some people not safer after successful treatment of tuberculosis?. Am. J. Respir. Crit. Care Med 171: 1324-1325.

Young D, Dye C (2006) The development and impact of tuberculosis vaccines. Cell 124: 683-687.

Zganiacz A, Santosuosso M, Wang J, Yang T, Chen L et al. (2004) TNF-alpha is a critical negative regulator of type 1 immune activation during intracellular bacterial infection. J Clin Invest 113: 401-413.

Zhang W, Shao L, Weng X, Hu Z, Jin A et al. (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.

Zhu X, Luke A, Cooper R, Quertermous T, Hanis C et al. (2005) Admixture mapping for hypertension loci with genome-scan markers. Nat.Genet 37: 177-181.

Zhu X, Tang H, Risch N (2008) Admixture mapping and the role of population structure for localizing disease genes. Adv Genet 60: 547-569.

Zhu X, Zhang S, Tang H, Cooper R (2006) A classical likelihood based approach for admixture mapping using EM algorithm. Hum Genet 120: 431-445.

Zignol M, Hosseini MS, Wright A, Weezenbeek CL, Nunn P et al. (2006) Global incidence of multidrug-resistant tuberculosis. J Infect Dis 194: 479-485.

Zimonjic DB, Rezanka LJ, Evans CH, Polymeropoulos MH, Trent JM et al. (1995) Mapping of the immune interferon gamma gene (IFNG) to chromosome band 12q14 by fluorescence in situ hybridization. Cytogenet Cell Genet 71: 247-248.

Ziv E, Burchard E (2003) Human population structure and genetic association studies. Pharmacogenomics 4: 431-441.