

# Identification of novel candidate genes for susceptibility to tuberculosis by identifying disease-causing mutations in individuals with Primary Immunodeficiency Disorders

Nikola Schlechter

*Dissertation presented for the degree Doctor of Philosophy in Science (Human Genetics)  
in the Faculty of Medicine and Health Sciences at Stellenbosch University*



Supervisor: Dr. Craig Kinnear

Co-supervisor: Dr. Marlo Möller

Faculty of Medicine and Health Sciences

Division of Molecular Biology and Human Genetics

March 2017

## DECLARATION

By submitting this thesis/dissertation, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: .....March 2017.....

Signature: .....

## ABSTRACT

Mendelian susceptibility to mycobacterial disease (MSMD) is a rare primary immunodeficiency disorder believed to affect less than 1:1 000 000 individuals globally. It is characterized by increased susceptibility to weakly virulent mycobacterial infections, such as Bacille Calmette Guerin, and in some cases more virulent agents, such as *Mycobacterium tuberculosis*, possibly implying a monogenic predisposition to tuberculosis (TB). TB claimed the lives of ~1.5 million people worldwide in 2014, making it the leading cause of death due to a single infectious agent and a very high-priority health problem. Ten genes have been associated with MSMD, five of which are also involved in TB susceptibility. In half of all MSMD patients, however, no mutations are found in these genes, highlighting the need to identify yet undiscovered MSMD-causing genes.

The present study aimed to identify novel MSMD-causing mutations by using whole exome sequencing (WES), and to determine whether the genes containing these variants are associated with increased TB susceptibility. It also investigated a potential link between MSMD and tuberculosis meningitis (TBM) by investigating variants identified in known MSMD-causing genes in TBM cases. Three MSMD patients and five of their healthy parents, as well as 10 TBM patients and 10 controls, were recruited. WES was performed on the Illumina HiSeq and produced ~60 000 - 75 000 variants per individual. Numerous filtering and variant prioritization tools were used to identify three plausible MSMD-causing variants in the MSMD patients based on *in silico* predictions of their potential to be disease-causing and the function of the genes they are situated in. These include two variants in transporter associated with antigen processing 1 (*TAP1*) and one in mitogen-activated protein kinase kinase kinase 14 (*MAP3K14*). *In vitro* functional studies verified their involvement in disease. Case-control association studies were performed using known single nucleotide polymorphisms in these two genes in a cohort of TB cases and controls. Four variants in three MSMD-causing genes were also identified in four of the TBM patients, which were absent from controls, and predicted to cause disease.

One variant was identified in each MSMD patient. The heterozygous I296M variant in *TAP1* was absent from ethnically matched controls and has not been previously identified according to

ExAC Browser, the 1000 Genomes Project and ESP6500. The variant is predicted to be deleterious and may alter antigen presentation to molecules of the type I major histocompatibility complex. The homozygous V345M variant in *MAP3K14* was absent from the above-mentioned databases, and predicted to be deleterious. It can potentially inhibit effective nuclear factor-kappa B signaling and thus lead to aberrant lymphoid immunity. The P67S variant in *TAP1* has not previously been identified in homozygous form, although it has been recorded in the heterozygous state in three Asians and one European. This heterozygous variant is predicted to be a benign polymorphism. However, homozygous variants can have more severe effects on protein function, and could influence different ethnic populations dissimilarly.

In summary, these results identify three novel putative variants involved in MSMD and thus increased TB susceptibility. It also hints at a potential link between MSMD and TBM, which should be investigated further. Identification of MSMD-causing mutations can inform treatment strategies by aiding in the implementation of patient-specific vaccine strategies and treatment regimes. It can lead to identification of at-risk relatives, and also provide novel candidate genes to be evaluated for increased TB susceptibility in the general population.

## OPSOMMING

Mendeliese vatbaarheid vir mikobakteriële siekte (MVMS) is 'n baie seldsame primêre immuniteitsgebrek wat minder as 1 uit elke 1 000 000 individue wêreldwyd affekteer. Dit word gekenmerk deur verhoogde vatbaarheid vir mikobakteriële infeksies van lae virulensie, soos Bacille Calmette Guerin, en in sommige gevalle meer virulente agense soos *Mycobacterium tuberculosis*, wat moontlik op 'n monogeniese vatbaarheid vir tuberkulose (TB) kan dui. TB was die oorsaak van ongeveer 1.5 miljoen sterftes wêreldwyd in 2014. Dit is dus die leidende oorsaak van dood as gevolg van 'n enkele aansteeklike agens en moet gevolglik as 'n baie hoë-prioriteit gesondheidsprobleem beskou word. Tot op hede is tien gene geassosieer met MVMS, waarvan vyf ook betrokke is by TB-vatbaarheid. In soveel as die helfte van alle MVMS-pasiënte is daar egter geen mutasies in enige van hierdie gene gevind nie. Dit beklemtoon die belang daarvan om steeds onbekende gene, wat MVMS veroorsaak, te identifiseer.

Die doelwit van die huidige studie was om nuwe mutasies wat MVMS veroorsaak te identifiseer, deur gebruik te maak van volledige eksoomvolgordebepaling (VEV) met gevorderde tegnieke, en om vas te stel of die gene, wat hierdie variante bevat, geassosieer kan word met verhoogde vatbaarheid vir TB. Dit het ook 'n potensiële skakel tussen MVMS en tuberkulose meningitis (TBM) ondersoek, deur die teenwoordigheid van variante in die gene wat voorheen met MVMS geassosieer is in TBM pasiënte te ondersoek. Drie MVMS-pasiënte en vyf van hulle gesonde ouers, asook 10 TBM pasiënte en 10 kontroles, is gewerf. VEV is uitgevoer op die Illumina HiSeq en het ~60 000 - 75 000 variante per pasiënt opgelewer. Verskeie filter- en variantprioritiseringsmetodes is gebruik drie potensiële variante wat moontlik MVMS in hierdie pasiënte kan veroorsaak, te identifiseer gebaseer op *in silico* voorspellings van hul potensiaal om siekte te veroorsaak, asook die funksie van die gene waarin hulle voorkom. Dit sluit in twee variante in *TAP1* en een in *MAP3K14*. *In vitro* funksionele studies is uitgevoer om die betrokkenheid van hierdie twee gene in MVMS te bewys. Gevallekontrole assosiasiestudies is uitgevoer deur gebruik te maak van bekende enkelaminosuur-polimorfismes in die twee nuut-ontdekte MVMS gene in 'n studiegroep van TB-gevalle en kontroles. Vier variante teenwoordig in drie van die gene wat MVMS veroorsaak is ook in vier van die TBM pasiënte ontdek, en was afwesig in die kontrolegroep. Hulle is voorspel om siekte te veroorsaak.

Een variant is in elke MVMS pasiënt identifiseer. Die heterosigotiese I296M-variant in *TAPI* was afwesig in kontroles van dieselfde etniese groep en is nog nooit tevore beskryf volgens “ExAC Browser”, die “1000 Genomes Project” en “ESP6500” nie. Die variant is voorspel om nadelig te wees en mag die voorstelling van antigeen aan molekule van die Klas I Major Histoverenigbaarheidskompleks affekteer. Die homosigotiese V345M-variant in *MAP3K14* is afwesig van die voorgenoemde databasisse en word voorspel om nadelig te wees. Dit kan moontlik effektiewe seintransduksie deur kernfaktor-kappa B verhoed en dus lei tot afwykende limfoïede immuniteit. Die P67S-variant in *TAPI* is nog nooit tevore in homosigotiese vorm beskryf nie, alhoewel dit in heterosigotiese vorm in drie Asiate en een Europeër opgeteken is. Hierdie heterosigotiese variant word voorspel om ‘n benigne polimorfisme te wees. Homosigotiese variante kan egter meer ernstige uitwerkings op proteïenfunksie hê en kan verskillende etniese populasies op verskillende maniere beïnvloed.

Ter opsomming identifiseer hierdie resultate drie nuwe variante wat vermoedelik betrokke mag wees by MVMS en dus verhoogde vatbaarheid vir TB. Dit identifiseer ook ‘n potensiële skakel tussen MVMS en TBM, wat verder ondersoek moet word. Die identifisering van mutasies wat MVMS veroorsaak kan behandelingstrategieë beïnvloed deur by te dra tot die implementering van pasiënt-spesifieke inentingstrategieë en behandelingsregimes. Dit kan lei tot die identifisering van hoe-risiko familieleden, en ook nuwe kandidaatgene identifiseer wat evalueer kan word vir hul betrokkenheid in verhoogde TB-vatbaarheid onder die algemene bevolking.

## ACKNOWLEDGEMENTS

A PhD degree is not something any person can obtain as a single entity. This thesis would absolutely not have been written without the support of so many.

To my two supervisors Drs. Craig Kinnear and Marlo Möller: thank you for your unconditional support, guidance and patience. Since day 1 you have made me feel like part of the family. Thank you for always trusting and believing in me. You both contributed immensely to the researcher, and the person, I am today. I am extremely proud to be associated with you.

My sincerest gratitude to Professors Paul and Eileen van Helden for giving me the opportunity to be a part of such a world-class research unit. I am privileged to have studied under your leadership. Your love and enthusiasm for research and for the students and staff in your research groups are inspiring and absolutely contagious.

Annika Neethling, Brigitte Glanzmann and Caitlin Uren: thank you for making every day in the lab worthwhile and for all your knowledgeable input into my research project. I am privileged to have shared these past three years with you.

A special thanks to Drs. Andre Loxton and Andrea Gutschmidt for all their immunology and flow cytometry expertise, and to Haiko Schurz and Nicholas Bowker for making statistics a little more understandable.

Thank you Mom and Dad for always giving me the best in life and for supporting my every choice. I owe all I have, all I am and all I ever will be to the two of you. Arlen de Villiers: thank you for being my lifeline, for loving me when I was unlovable, for suffering silently through all my moods, and for your unconditional support. I am truly blessed to share my life with you.

This research project would not have been possible without financial support from the National Research Foundation, the Harry Crossley Foundation and the Stella and Paul Loewenstein Charitable and Educational Trust.

## TABLE OF CONTENTS

<b>INDEX</b>	<b>PAGE</b>
List of abbreviations	xi
List of figures	xxi
List of tables	xxiii
Outline of dissertation	xxv
Chapter 1: INTRODUCTION	1
1.1. INFECTIOUS DISEASES	2
1.1.1. Tuberculosis	3
1.2. <i>MYCOBACTERIUM TUBERCULOSIS</i>	4
1.2.1. Characteristic features	4
1.2.2. The emergence of a human pathogen	7
1.3. IMMUNE RESPONSE AGAINST <i>M.tb</i> INFECTION	8
1.3.1. Innate immune responses to <i>M.tb</i> infection	10
1.3.1.1. <i>Macrophages</i>	10
1.3.1.2. <i>Neutrophils</i>	12
1.3.1.3. <i>Dendritic cells</i>	13
1.3.1.4. <i>Natural killer cells</i>	14
1.3.2. Adaptive immune response to <i>M.tb</i> infection	16
1.3.3. The type I cytokine response	18
1.3.4. Disease outcome	19
1.4. SUSCEPTIBILITY TO <i>M.tb</i> INFECTION	21
1.5. PRIMARY IMMUNODEFICIENCY DISORDERS	24
1.5.1. Mendelian susceptibility to mycobacterial disease	27
1.6. PRESENT STUDY	31
Chapter 2: MATERIALS AND METHODS	32
2.1. STUDY PARTICIPANTS	34
2.1.1. MSMD patients	34
2.1.1.1. <i>Case descriptions</i>	35
2.1.2. TBM cohort	36

2.1.3.	PTB cases and controls	37
2.2.	WHOLE EXOME SEQUENCING	38
2.2.1.	Exome sequencing	39
2.2.2.	Data analysis and bioinformatics	39
2.2.3.	Polymerase chain reaction (PCR)	43
2.2.4.	Gel electrophoresis	44
2.2.5.	Gel clean-up	44
2.2.6.	Sanger sequencing	45
2.3.	VARIANT FREQUENCY DETECTION IN CONTROLS	45
2.3.1.	Allele-specific restriction enzyme digest	45
2.3.2.	Database searches	46
2.4.	<i>IN SILICO</i> VARIANT PREDICTIONS	47
2.4.1.	HOPE	47
2.4.2.	MutationTaster2	48
2.4.3.	PolyPhen-2	50
2.4.4.	SIFT	52
2.5.	FUNCTIONAL STUDIES	54
2.5.1.	Generation of mutant plasmid constructs	55
2.5.1.1.	<i>Bacterial plasmids</i>	55
2.5.1.2.	<i>Site-directed mutagenesis</i>	55
2.5.2.	Bacterial strain and cell line	56
2.5.2.1.	<i>Bacterial Strain</i>	56
2.5.2.2.	<i>Cell line</i>	56
2.5.2.3.	<i>Bacterial plasmid isolation</i>	57
2.5.3.	Generation of <i>E.coli</i> DH5 $\alpha$ competent cells	57
2.5.4.	Culturing of the HEK293T Cell Line from frozen stocks	57
2.5.4.1.	<i>Thawing the cells</i>	57
2.5.4.2.	<i>Removing DMSO from stocks and culturing cells</i>	58
2.5.4.3.	<i>Splitting of cell cultures</i>	58
2.5.5.	Bacterial plasmid transformation	58
2.5.6.	Plasmid transfection into human cells	59

2.5.7.	Bafilomycin treatment of cells	59
2.5.8.	Protein isolation via cell lysis	60
2.5.9.	Bradford protein concentration determination	60
2.5.10.	Western blot	61
2.5.10.1.	<i>Sodium dodecyl sulphate polyacrylamide gel electrophoresis</i>	61
2.5.10.2.	<i>Membrane blocking</i>	61
2.5.10.3.	<i>Addition of primary antibody</i>	61
2.5.10.4.	<i>Addition of secondary antibody</i>	61
2.5.10.5.	<i>Chemiluminescent visualization of membrane proteins</i>	62
2.6.	FLOW CYTOMETRY	62
2.6.1.	Peripheral blood mononuclear cell isolation and growth	62
2.6.2.	Absolute cell counts	64
2.7.	CASE-CONTROL ASSOCIATION STUDIES	65
2.7.1.	Genotyping	65
2.7.2.	Data analysis	65
Chapter 3:	RESULTS	67
3.1.	WHOLE EXOME SEQUENCING	68
3.1.1	Sanger sequencing	70
3.2.	VARIANT FREQUENCY DETECTION IN CONTROLS	70
3.2.1.	Allele-specific restriction enzyme digest	70
3.2.2	Database searches	72
3.3.	<i>IN SILICO</i> VARIANT PREDICTIONS	72
3.3.1.	<i>TAP1 I296M</i>	73
3.3.2.	<i>MAP3K14 V345M</i>	74
3.3.3.	<i>TAP1 P67S</i>	75
3.3.4.	<i>ISG15 L54P</i>	75
3.3.5.	<i>TYK2 C378F</i>	76
3.3.6.	<i>IL12RB1 T432M</i>	77
3.3.7.	<i>TYK2 A53T</i>	77
3.4.	FUNCTIONAL STUDIES	79
3.4.1.	Immunological workup of PID_040	79

3.4.2.	Investigation of the effect of <i>MAP3K14 V345M</i>	79
3.4.2.1.	<i>Site-directed mutagenesis</i>	79
3.4.2.2.	<i>Plasmid transfection into human cells</i>	79
3.4.2.3.	<i>Phosphorylation assay</i>	81
3.4.2.4.	<i>Downstream functional effects of MAP3K14 V345M</i>	82
3.4.2.5.	<i>Effect of MAP3K14 V345M on autophagy</i>	83
3.5.	FLOW CYTOMETRY	85
3.6.	CASE-CONTROL ASSOCIATION STUDIES	85
Chapter 4:	DISCUSSION	88
4.1.	OVERVIEW OF STUDY FINDINGS	90
4.2.	WHOLE EXOME SEQUENCING	91
4.2.1.	MSMD patients	91
4.2.1.1.	<i>MAP3K14: The NF-<math>\kappa</math>B signaling pathway</i>	92
4.2.1.2.	<i>TAP1: MHC class I activation</i>	93
4.2.2.	TBM cohort	94
4.2.2.1.	IL12RB1	95
4.2.2.2.	ISG15	96
4.2.2.3.	TYK2	97
4.3.	<i>IN SILICO</i> VARIANT PREDICTIONS	99
4.3.1.	MSMD patients	100
4.3.1.1.	TAP1 I296M	100
4.3.1.2.	MAP3K14 V345M	104
4.3.1.3.	TAP1 P67S	106
4.3.2.	TBM patients	108
4.3.2.1.	ISG15 L54P	108
4.3.2.2.	TYK2 C378F	109
4.3.2.3.	TYK2 A53T	112
4.3.2.4.	IL12RB1 T432M	113
4.4.	FUNCTIONAL STUDIES	115
4.4.1.	Effect of <i>MAP3K14 V345M</i> on the function of NIK	115
4.4.2.	Downstream effects of <i>MAP3K14 V345M</i>	117

4.4.3.	Effect of <i>MAP3K14 V345M</i> on autophagy	118
4.5.	FLOW CYTOMETRY	120
4.6.	CASE-CONTROL ASSOCIATION STUDIES	123
4.6.1.	Population genetics	125
4.6.2.	Complex genetic effects	125
4.6.3.	Categorizing cases and controls	126
4.6.4.	<i>M.tb</i> strain diversity	127
4.7.	STUDY LIMITATIONS	128
4.7.1.	Whole exome sequencing	128
4.7.2.	Patients	130
4.7.3.	Functional studies	131
4.7.4.	Case-control association studies	132
4.7.4.1.	<i>Number of markers tested</i>	132
4.7.4.2.	<i>Phenotypic resolution</i>	133
4.7.4.3.	<i>Population stratification</i>	133
4.7.4.4.	<i>Statistical power and sample size</i>	134
4.8.	FUTURE STUDIES	135
4.9.	CONCLUDING REMARKS	136
	Reference List	138
Appendix I:	PID warning signs	195
Appendix II:	FastQC results for exome sequencing data	196
Appendix III:	Supplementary Table 1	212
Appendix IV:	ANNOVAR explained	215
Appendix V:	Reagents	216
Appendix VI:	Cell counting with Hemocytometer	219
Appendix VII:	Supplementary Table 2	221
Appendix VIII:	List of suppliers	222
Appendix IX:	Supplementary Figure 1	224
Appendix X:	Supplementary Figure 2	227
Appendix XI:	Supplementary Figure 3	228
Appendix XII:	Case-control association study results	230

## LIST OF ABBREVIATIONS

1KGP	1000 Genomes Project
3D	Three dimensional
ABC	ATP-binding cassette
AD	Autosomal dominant
AFB	Acid-fast bacilli
AIDS	Acquired immune deficiency syndrome
Ala	Alanine
AMP	Antimicrobial peptides
AnR	Androgen receptor
APC	Antigen presenting cells
APCDR	African Partnership for Chronic Disease Research
AR	Autosomal recessive
Arg	Arginine
ASN	Alpha-synuclein
Asn	Aspartic acid
AVGP	African Genome Variation Project
ATP	Adenosine triphosphate
B-LCL	Epstein-Barr virus-immortalized lymphoblastoid cell lines
BAFF	Bafilomycin A1
BAFFR	B cell-activating factor belonging to the TNF family
BAM	Binary alignment/map
BC	Before Christ
BCG	Bacille Calmette Guerin
BDNF	Brain-derived neurotrophic factor
BLAST	Basic local alignment search tool
BR	Basic region
BSA	Bovine serum albumin
CAAPA	Consortium on Asthma among African-ancestry Populations in the Americas

CADASIL	Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy
cADPR	Cyclic-adenine dinucleotide phosphate ribose
CAF	Central Analytical Facility
cAMP	Cyclic adenosine monophosphate
CBD	Cytokine binding domains
CCDS	Consensus coding sequence
CD	Cluster of differentiation
CFU	Colony forming units
CH	Congenital hypo-thyroidism
CGD	Chronic granulomatous disease
CIITA	Class II trans-activator
CMT	Charcot-Marie-Tooth
CNS	Central nervous system
CO <sub>2</sub>	Carbon dioxide
COMT	Catechol-O-methyltransferase
CREB	Cyclic adenosine monophosphate response element binding
CTL	Cytolytic T lymphocytes
CYBB	Cytochrome b beta
Cys	Cysteine
DAS	Distributed annotated system
dbSNP	Single nucleotide polymorphism database
DC	Dendritic cells
DC-SIGN	Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DISC	Death-inducing signaling complexes
DMEM	Dulbecco's modified eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxy ribonucleotide triphosphate
DosR	Dormancy survival regulator

DSB	DNA double-strand break
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
EHR	Enduring hypoxic response
EM	Environmental mycobacteria
ESID	European Society for Immunodeficiencies
ESP6500	Exome Sequencing Project 6500
ETA	Ethionamide
ExAC	Exome Aggregation Consortium
FABP2	Fatty acid binding protein
FACS	Fluorescence-activated cell sorting
FasL	Fas ligand
FATHMM	Functional analysis through hidden Markov models
FBS	Fetal bovine serum
FERM	Band 4.1, ezrin, radixin, moesin-homology
FMF	Familial Mediterranean fever
Fn	Fibronectin
GATK	Genome analysis toolkit
GERP	Genomic evolutionary rate profiling
Gly	Glycine
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GoNL	Genome of the Netherlands
GWAS	Genome-wide association studies
H <sub>2</sub> O <sub>2</sub>	Hydrogen peroxide
H3Africa	Human Heredity and Health in Africa
HA	Hemagglutinin
HapMap	Haplotype Map
HEK293	Human embryonic kidney
HGMD	Human Gene Mutation Database
HI	Hearing impairment
HIES	Hyper-IgE syndrome

HIF-1 $\alpha$	Hypoxia-inducible factor 1 $\alpha$
HIR	Human insulin receptor
HIV	Human immunodeficiency virus
HMERF	Hereditary myopathy with early respiratory failure
HOPE	Have (y)Our Protein Explained
HRP	Horseradish peroxidase
HRR	Homologous recombination repair
HSP	Heat shock proteins
HTS	High throughput system
HWE	Hardy-Weinberg equilibrium
ICAM-1	Intercellular adhesion molecule 1
ICMB	Institute for Clinical Molecular Biology
IFN- $\gamma$	Interferon gamma
IFNGR	Interferon gamma receptor
Ig	Immunoglobulin
IGRA	Interferon gamma release assay
IKBKG	Inhibitor of nuclear factor kappa-B kinase subunit gamma
IL	Interleukin
IL12B	Interleukin 12 beta
IL12RB1	Interleukin 12 receptor beta 1
Ile	Isoleucine
IM	Integral membrane
INH	Isoniazid
IRF8	Interferon regulatory factor 8
ISG15	Interferon-stimulated genes 15
ITIM	Immunoreceptor tyrosine-based inhibitory motif
JAK3	Janus kinase 3
KF- $\kappa$ B	Kernfaktor-kappa B
KIR	Killer cell Ig-like receptors
KLD	Kinase, Ligase and DpnI
LAGID	Latin American Group for Primary Immunodeficiencies

LAM	Lipoarabinomannans
LB	Luria-Bertani broth
LCK	Lymphocyte-specific protein tyrosine kinase
LD	Linkage disequilibrium
LIF	Leukemia inhibitory factor
LIR	Leukocyte Ig-like receptors
LM	Lipomannans
LT $\beta$ R	Lymphotoxin beta receptors
LTA4H	Leukotriene A4 hydrolase
LTBI	Latent tuberculosis infection
Lys	Lysine
mAbs	Monoclonal antibodies
MAF	Minor allele frequency
ManLAM	Mannosylated Lipoarabinomannans
MAP3K14	Mitogen-activated protein kinase kinase kinase 14
MARCH1	Membrane associated ring-CH-type finger 1
Mb	Mega-bases
MDC	Macrophage-derived chemokine
MDR	Multi-drug resistance
Met	Methionine
MHC	Major Histocompatibility complex
MICAL1	Microtubule associated monooxygenase, calponin and LIM domain containing 1
MIP	Macrophage inflammatory protein
mL	Milliliters
MPs	Micro-particles
MR	Mannose receptor
mRNA	Messenger ribonucleic acid
MSMD	Mendelian susceptibility to mycobacterial disease
MT	Memory T
MT2	MutationTaster2

<i>M.tb</i>	<i>Mycobacterium tuberculosis</i>
MTBC	<i>Mycobacterium tuberculosis</i> complex
MVMS	Mendeliese vatbaarheid vir mikobakteriële siekte
MYD88	Myeloid differentiation primary response gene 88
NAADP	Nicotinic acid adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate
NAG	N-acetyl glucosamine
NaHep	Sodium heparin
NAM	N-acetylmuramic acid
NBD	Nucleotide binding domain
NCBI	National Center for Biotechnology Information
NCR	Non-catalytic region
NEB	New England BioLabs
NEMO	NF-kappa-B essential modulator
NF-κB	Nuclear factor-kappa B
ng	Nano-gram
NHLS	National Health Laboratory Service
NIH	National Institutes of Health
NIK	Nuclear factor-kappa B inducing kinase
NK	Natural killer
NMD	Nonsense-mediated mRNA decay
NOS	Nitric oxide synthase
NRD	Negative regulatory domain
NSCLC	Non-small-cell lung cancer
NTM	Nontuberculous mycobacteria
NQO1	NAD(P)H dehydrogenase, quinone 1
OMIM	Online Mendelian Inheritance of Man
OR	Odds ratio
PAMP	Pathogen-associated molecular patterns
PBMC	Peripheral blood mononuclear cells
PBP2	Penicillin-binding protein 2

PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PD	Parkinson's disease
PD-1	Programmed cell death-1
PDB	Protein Data Bank
PD-L1	Programmed death ligand 1
PHA	Phytohemagglutinin
Phe	Phenylalanine
PI3K	Phosphatidylinositol 3-kinases
PID	Primary immunodeficiency disorder
PIM	Phosphatidylinositol mannosides
PK	Pseudo-kinase
PKC $\gamma$	Protein kinase C gamma
PMP22	Peripheral myelin protein 22
PNPLA3	Patatin-like phospholipase-3
PP-2	PolyPhen-2
PPAR $\gamma$	Peroxisome proliferator activated receptor gamma
PPD	Purified protein derivative
RB	Retinoblastoma
PAGE	Population Architecture in Genomics and Epidemiology
Pro	Proline
PRR	Pattern-recognition receptor
P-RR	Proline-rich repeat
PSI-BLAST	Position-specific iterative basic local alignment search tool
PTB	Pulmonary tuberculosis
QTL	Quantitative trait loci
RBC	Red blood cells
RIF	Rifampicin
RNA	Ribonucleic acid
RNI	Reactive nitrogen intermediates
ROI	Reactive oxygen intermediates

ROS	Reactive oxygen species
Rpm	Revolutions per minute
RR	Reactive
SA	South Africa
SAC	South African Coloured
SAM	Sequence alignment/map
SCA14	Spinocerebellar ataxia type 14
SCID	Severe combined immunodeficiency disorders
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Ser	Serine
SIFT	Sorting intolerant from tolerant
SNAP	Single nucleotide polymorphism annotation and proxy search
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SOC	Super optimal broth
STAT	Signal transducer and activator of transcription
STB	Smooth tubercle bacilli
SUFCU	Stellenbosch University Flow Cytometry Unit
TAP1	Transporter associated with antigen processing 1
TAPER	Tool for automated selection and prioritization for efficient retrieval of sequence variants
TB	Tuberculosis
TbD1	Tuberculosis deletion event one
TBM	Tuberculosis meningitis
TBST	Tris-buffered saline, 0.1% Tween 20
TC	Computerized tomography
TCR	T cell receptor
TGIF1	Transforming growth $\beta$ -induced factor 1
Th	T helper cells
Thr	Threonine

TIRAP	Toll-interleukin receptor domain containing adaptor protein
TLR	Toll-like receptor
TMD	Transmembrane domain
TMHs	Transmembrane helices
TNF	Tumor necrosis factor
TNFR	Tumor necrosis factor receptors
tRNA	Transporter ribonucleic acid
TRAF3	Tumor necrosis factor receptor-associated factor 3
TSPAN33	Tetraspanin 33
TST	Tubercle skin test
TYK	Tyrosine kinase
UBL	Ubiquitin-like
UCT	University of Cape Town
UCSC	University of California, Santa Cruz
UN	United Nations
US	United States
USIDNet	United States immunodeficiency network
UTR	Untranslated region
UU	Unreactive
UV	Ultra violet
Val	Valine
VCF	Variant call format
VEV	Volledige eksoomvolgordebepaling
VRL	Vitreoretinal lymphoma
WES	Whole exome sequencing
WGS	Whole genome sequencing
w/v	Weight per volume
XDR	Extremely drug resistance
XRCC3	X-ray cross-complementing group 3
$\alpha$	Alpha
$\beta$	Beta

$\gamma$	Gamma
$\gamma_c$	Common gamma chain
$\kappa$	Kappa
$\mu\text{L}$	Micro-liters
$\mu\text{M}$	Micro-molar
$^{\circ}\text{C}$	Degrees Celsius

## LIST OF FIGURES

	<b>PAGE</b>
Figure 1.1: The cell envelope of mycobacteria	6
Figure 1.2: Evolutionary association between MTBC members and selected mycobacteria	9
Figure 1.3: Macrophage response to <i>M.tb</i> infection	11
Figure 1.4: Neutrophil response to <i>M.tb</i> infection	12
Figure 1.5: Dendritic cell response to <i>M.tb</i> infection	14
Figure 1.6: <i>M.tb</i> elimination by direct functioning of natural killer cells	15
Figure 1.7: Adaptive immunity to TB infection	17
Figure 1.8: Type-1 cytokine pathway	20
Figure 1.9: Primary features of TB: initial infection to activated host defense	22
Figure 2.1: Overview of bioinformatics pipeline used for processing of WES data	41
Figure 2.2: Overview of filtering process used to determine potential disease-causing variants	42
Figure 2.3: Simple map of pWZL-Neo-Myr-Flag-MAP3K14 plasmid	55
Figure 2.4: Simple map of pCR-Flag-IKK-alpha plasmid	55
Figure 2.5: Position of cell types and density media after centrifugation of whole blood	63
Figure 3.1: Sanger sequencing results for <i>JAK3 R918P</i> in PID_012	70
Figure 3.2: Enzyme digest results for <i>TSPAN33 A111T</i> identified in PID_012	71
Figure 3.3: Representation of enzyme digest results for the <i>TAP1 I296M</i> variant identified in PID_012	71
Figure 3.4: Mutation of Ile to Met observed at position 296 in TAP1 of PID_012	73
Figure 3.5: Mutated and wild-type amino acids at position 296 of TAP1	74
Figure 3.6: Mutation of Val to Met, as observed at position 345 in MAP3K14 of PID_040	74
Figure 3.7: Mutated and wild-type amino acids at position 345 of MAP3K14	74
Figure 3.8: Mutation of Pro to Ser, as observed at position 67 in TAP1 of PID_060	75
Figure 3.9: Mutation of Leu to Pro, as observed at position 54 in ISG15 of D5371 and D5378	75
Figure 3.10: Mutated and wild-type amino acids at position 54 of ISG15	76

Figure 3.11: Mutation of Cys into Phe, as observed at position 378 in TYK2 of D5371	76
Figure 3.12: Mutated and wild-type amino acids in TYK2 of D5371	76
Figure 3.13: Mutation of Thr into Met, as observed at position 432 in IL12RB1 of D5374	77
Figure 3.14: Mutation of Ala into Thr, as observed at position 53 in TYK2 of D5377	78
Figure 3.15: Mutated and wild-type amino acids in TYK2 of D5377	78
Figure 3.16: Sanger sequencing of PWZL-NEO-FLAG-MAP3K14 (top) and PWZL-NEO-FLAG-MAP3K14-V345M (bottom) plasmids	80
Figure 3.17: Investigation of successful transfection of the <i>MAP3K14</i> and <i>IKK<math>\alpha</math></i> plasmids into HEK293T cells	80
Figure 3.18: Investigation of kinase activity of mutant versus wild-type NIK	81
Figure 3.19: Investigation of p52 and p100 levels of mutant versus wild-type NIK	83
Figure 3.20: Measurement of LC3-II levels before and after BAFF treatment	84
Figure 3.21: Percentage of CD4 <sup>+</sup> and CD8 <sup>+</sup> T cells in PID_012 and a healthy control	86
Figure 3.22: Percentages of activated CD4 <sup>+</sup> and CD8 <sup>+</sup> T cells	87
Figure 4.1: The non-canonical NF- $\kappa$ B pathway	93
Figure 4.2: Function of TAP1 in antigen processing via MHC class I molecules	94
Figure 4.3: Conserved motifs in the ABC	101
Figure 4.4: Domains of the ABC	101
Figure 4.5: Schematic representation of the TAP transporter	102
Figure 4.6: Schematic representation of NIK protein domain structures	105
Figure 4.7: The domain structure of JAK-family kinases	110
Figure 4.8: Basal autophagy and the effect of BAFF treatment	119

## LIST OF TABLES

	<b>PAGE</b>
Table 1.1: Inherited PIDs associated with mycobacterial susceptibility to BCG, NTM or <i>M.tb</i>	26
Table 1.2: Mutations in genes associated with different types of MSMDs	29
Table 2.1: List of MSMD patients and controls sent for whole exome sequencing	34
Table 2.2: Demographic variable of the TBM cases and controls	37
Table 2.3: Demographic variable of the PTB cases and controls	38
Table 2.4: Forward and reverse primers used for each specific variant	44
Table 2.5: Allele-specific restriction enzyme analysis of variants identified in PID_012	46
Table 2.6: Expected outcome of restriction enzyme analysis of variants identified in PID_012	47
Table 2.7: Summary of MT2 output	48
Table 2.8: Summary of PP-2 output	51
Table 2.9: Summary of SIFT output	53
Table 2.10: Comparison between PP-2, SIFT and MT2	54
Table 2.11: Q5® Site-directed mutagenesis primers	56
Table 2.12: Primary and secondary antibodies and their optimized concentrations used in Western blot assays	62
Table 2.13: Amount of antibodies added to PBMCs of PID_012 and control for flow cytometric analysis	64
Table 3.1: Variants identified in the MSMD patients after WES	68
Table 3.2: Potentially disease-causing variants identified in functionally relevant genes in the MSMD patients	69
Table 3.3: Potentially MSMD-causing variants identified in the TBM patients	69
Table 3.4: Population frequencies of the <i>TAP1 P67S</i> variant caused by a C/T nucleotide exchange	72
Table 3.5: <i>In silico</i> predictive results of the variants identified in the MSMD patients	73
Table 3.6: <i>In silico</i> predictive results of the variants identified in the TBM patients	73
Table 3.7: Population frequencies of the <i>IL12RB1 T432M</i> variant	77
Table 3.8: Population frequencies of the <i>TYK2 A53T</i> variant	78

Table 3.9: Total cell counts from 2010 until 2014, in units of cells per microliter	79
Table 3.10: Percentages of activated CD4 <sup>+</sup> and CD8 <sup>+</sup> T cells	87

## OUTLINE OF DISSERTATION

This dissertation involves the whole exome sequencing of three South African patients diagnosed with Mendelian susceptibility to mycobacterial disease (MSMD) and five of their healthy parents, as well as 10 tuberculosis meningitis (TBM) patients and 10 ethnically matched controls. It identifies three putative MSMD-causing variants and makes predictions with regards to the cause of these variants. It also functionally characterizes one of them. Furthermore, four variants are identified in MSMD-associated genes of the TBM patients, potentially providing a link between MSMD and TBM.

This dissertation is divided into four chapters:

**Chapter One** gives a background and overview of tuberculosis (TB) with specific focus on the human host's immune response against *Mycobacterium tuberculosis* infection. It gives an overview of primary immunodeficiency disorders and describes MSMD in more detail. It introduces the present study by detailing its aims and objectives.

**Chapter Two** describes the methodological approaches used in this study. It gives information with regards to recruitment of patients, identification of potentially disease-causing variants using whole exome sequencing (WES) and bioinformatics techniques, verification and frequency investigation of all interesting variants, as well as functional characterization of one of the variants.

**Chapter Three** is a description of results obtained throughout the study, including results from WES of all study participants, Sanger sequencing and restriction enzyme digestion of the interesting variants identified in MSMD patients, and results from functional experiments.

**Chapter Four** discusses the important findings of this study, the possible relevance of the findings and the role they might play in better understanding MSMD and TB susceptibility in the South African population. It advises on possible future work that may aid in early diagnosis of MSMD patients and identification of children at risk of developing TB.

## CHAPTER 1: INTRODUCTION

INDEX	PAGE
1.1. INFECTIOUS DISEASES	2
1.1.1. Tuberculosis	3
1.2. <i>MYCOBACTERIUM TUBERCULOSIS</i>	4
1.2.1. Characteristic features	4
1.2.2. The emergence of a human pathogen	7
1.3. IMMUNE RESPONSE AGAINST <i>M.tb</i> INFECTION	8
1.3.1. Innate immune responses to <i>M.tb</i> infection	10
1.3.1.1. <i>Macrophages</i>	10
1.3.1.2. <i>Neutrophils</i>	12
1.3.1.3. <i>Dendritic cells</i>	13
1.3.1.4. <i>Natural killer cells</i>	14
1.3.2. Adaptive immune response to <i>M.tb</i> infection	16
1.3.3. The type I cytokine response	18
1.3.4. Disease outcome	19
1.4. SUSCEPTIBILITY TO <i>M.tb</i> INFECTION	21
1.5. PRIMARY IMMUNODEFICIENCY DISORDERS	24
1.5.1. Mendelian susceptibility to mycobacterial disease	27
1.6. PRESENT STUDY	31

## 1.1. INFECTIOUS DISEASES

In the early 1900's Charles Nicolle, a French bacteriologist, fundamentally changed the way researchers thought about infectious diseases by discovering asymptomatic infections – causing microbiologists worldwide to investigate the specific mechanisms leading to unique variability in disease-outcome among patients<sup>1</sup>. The theory that disease is significantly less common than asymptomatic infections for most microbes became increasingly popular, and led researchers to ask the question: on what other factors do the pathogenesis of infectious diseases depend? To date, it has been consistently proven that environmental factors (microbial, ecological, socio-economic) as well as host factors (both genetic and non-genetic) influence the complexity of each individual infection to varying degrees<sup>2</sup>. In the context of homogenous populations where most other factors play an equal role and where just a few infected individuals develop clinical disease, the genetic background of the host started enjoying more attention as one of the main determinants of disease outcome<sup>3,4</sup>.

Accumulating evidence from as early as the 1930's gave rise to the theory that susceptibility to infectious diseases could be inherited<sup>3</sup>. Biological children of patients suffering from infectious diseases are significantly more at risk of developing the same infections than adoptees living in the same homes, with an odds ratio of 5.81<sup>5</sup>. The concordance rate of infectious diseases in monozygotic twins are also significantly higher than in dizygotic twins<sup>6,7</sup>. Both ethnic and familial aggregation of common and rare infections, which in some cases include Mendelian inheritance patterns, have also been observed. These findings strongly suggest the existence of a genetic component to infectious phenotypes<sup>2,6,8,9</sup>.

Two main schools of thought emerged in the 1950's with regards to infectious diseases, the first consisting of Garrodian clinicians and scientists, and the second of Galtonian population geneticists. The former focussed mainly on monogenic “inborn errors” of immunity which they termed primary immunodeficiency disorders (PIDs). They have had great success over the years in identifying and describing over 200 of these “errors”<sup>8,10,11</sup>. Infectious diseases from a human genetics point of view have simultaneously been studied by the second group<sup>6</sup>. Both groups contributed immensely to the current knowledge that rare weakly-virulent microbes and/or

multiple recurring infections in single patients are consequences of monogenic PIDs<sup>12</sup>. Individuals with PIDs can either show rare infectious phenotypes associated with rare alleles, or multiple infectious diseases caused by one defective gene product, which supports the idea that rare alleles cause rare inherited disorders, while common diseases are associated with common variants<sup>13</sup>.

One of the main goals of human infectious disease research for the last decade has been to identify and investigate as yet undiscovered genes rendering individuals, families or entire populations more resistant or increasingly susceptible to different infectious agents<sup>14</sup>. Typically, common major alleles predispose entire populations to an infectious disease while rare, monogenic inborn errors of immunity render single individuals susceptible to specific infectious diseases<sup>15</sup>. However, as the number of studies conducted in this regard increased, so did the different theories of potential modes of inheritance, leaving the field of genetic predisposition to infectious diseases without a stable, predictive architecture. More research with regards to genes and alleles involved in increased susceptibility to specific infectious diseases as well as the modes of inheritance of these genes and alleles is crucial.

Tuberculosis (TB) is ranked alongside human immunodeficiency virus (HIV) as a leading cause of death due to an infectious disease, annually claiming the lives of millions of adults and children<sup>16</sup>. Interestingly, some PIDs, such as Mendelian susceptibility to mycobacterial disease (MSMD), increase sufferers' susceptibility to mycobacterial infections, even those considered normally non-pathogenic<sup>17</sup>. Understanding the genetic factors that play a role in PIDs predisposing individuals to TB may shed more light on the genetic factors and immunological pathways involved in TB susceptibility.

### **1.1.1. Tuberculosis**

Several mycobacterial strains can cause TB in humans, of which *Mycobacterium tuberculosis* (*M.tb*) is the main causative agent. Although it is estimated that only 5-10% of infected individuals develop active TB disease<sup>18</sup>, it remains one of the leading causes of death worldwide due to a single infectious agent<sup>16</sup>. For this reason, TB was declared a major global health problem more than a decade ago. Still, an estimated 9.6 million people get infected with *M.tb*

and approximately 1.5 million die due to the disease annually<sup>16</sup>. Although the numbers of infections and fatalities have decreased significantly over the last decade, and the Millennium Development Goal of decreasing the number of TB cases worldwide set by the United Nations (UN) in 2000 has been reached, most infections and fatalities could have been prevented with effective treatment regimens and correct treatment adherence<sup>16</sup>. South Africa (SA) is currently rated amongst the six highest-burdened countries in the world along with China, India, Indonesia, Nigeria and Pakistan, with more than 800 000 new infections each year<sup>19</sup>. Despite the increased amount of research conducted to identify TB cases and effective treatment strategies, and the increased funds spent by governments to control this epidemic, very little has actually improved<sup>16,20</sup>. With the emergence of multi- and extremely-drug resistant (MDR and XDR) mycobacterial strains, the lack of effective vaccines and new antibiotics, and the worldwide HIV pandemic, which increases an individual's risk of developing active TB<sup>20</sup>, the situation has worsened.

It is of great global health importance to investigate and understand why only a small subset of the population exposed to the bacteria will develop active disease<sup>21</sup>. It is also fascinating that in densely-populated, highly endemic countries where the majority of inhabitants are chronically exposed to a variety of extremely infectious *M.tb* strains, approximately 20% of the population repeatedly test negative for infection throughout their lives. This most probably indicates a natural inborn resistance against *M.tb* in this specific fraction of the population, highlighting the determinant role host genetics plays in disease outcome, and is worth investigating<sup>22</sup>. Individual disease development is influenced by a variety of inherent biological components, and genetic studies in human hosts have shown the involvement of several common variants that increase disease susceptibility as well as some rare variants causing severe forms of TB<sup>23</sup>. A defined cause is, however, absent in most infected individuals developing active disease<sup>24</sup>.

## **1.2. MYCOBACTERIUM TUBERCULOSIS**

### **1.2.1. Characteristic features**

*M.tb* belongs to the *Mycobacterium* genus, a gram-positive bacteria part of the *Actinomycetes* family<sup>25</sup>. The *Mycobacterium tuberculosis* complex (MTBC), capable of causing TB in humans, consists of *M.africanum*, *M.bovis*, *M.canetti*, *M.microti* and *M.tb*. Although the predecessor for

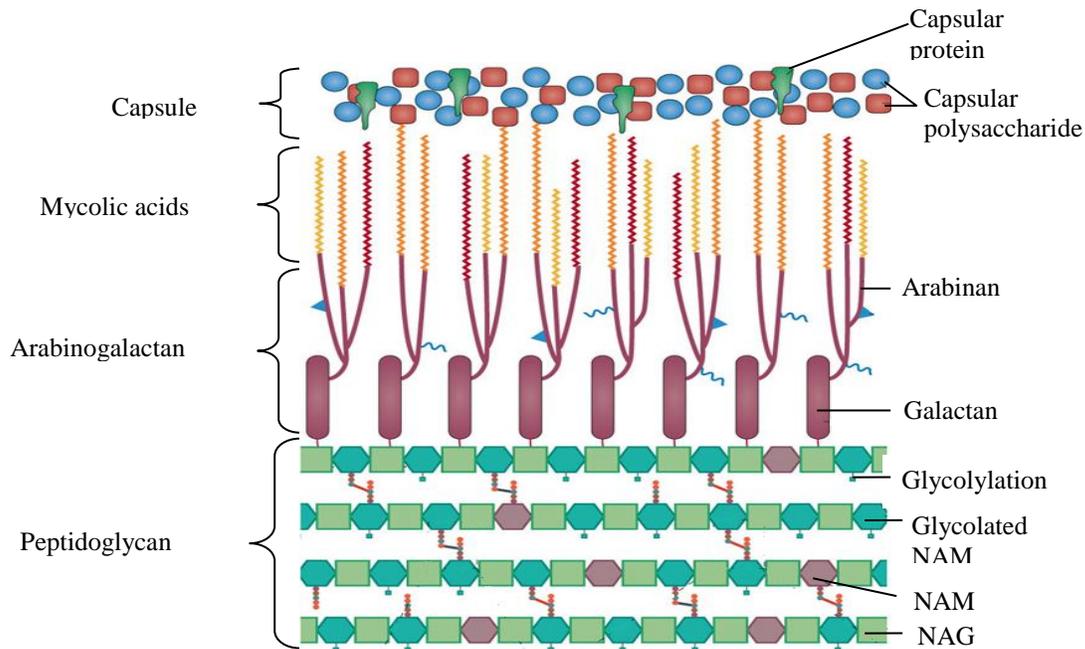
the MTBC as a human pathogen remains speculative, one theory is that these members originated from a progenitor related to *M.canetti*<sup>26</sup>. By estimations, the latest shared ancestor lived 5 000 - 40 000 years ago and most probably originated in Africa<sup>27,28</sup>. Very little genetic diversity exists between these different strains and most share identical proteins. The *M.tb* genome is thought to be either abnormally inert or evolutionary relatively young<sup>29</sup>. The tubercle bacillus has several characteristic features contributing to the uniqueness of TB, such as a complex cell envelope, intracellular pathogenesis, slow growth, dormancy and genetic homogeneity<sup>30</sup>.

*M.tb* has a tough cell wall that inhibits the movement of nutrients in and out of the cell, causing a typically slow growth rate. Its intricate cell envelope (Figure 1.1) is comprised of three macromolecules: peptidoglycan, arabinogalactan and mycolic acids. This is essential for antibiotic resistance, mycobacterial virulence and cell growth<sup>31,32</sup>. The peptidoglycan layer consists of long polymers of disaccharide N-acetyl glucosamine–N-acetylmuramic acid (NAG–NAM) heavily crosslinked by peptide bridges, and surrounds the plasma membrane<sup>33</sup>. It is surrounded by the arabinogalactan layer, which consists of arabinose and galactose monosaccharides. Polymers of arabinose, known as arabinan, are bound by long-carbon-chain mycolic acids to give mycobacteria its characteristic wax-like lipid coat and contributes towards virulence and cell wall impermeability<sup>34</sup>. Mycolic acids consist of shorter, saturated  $\alpha$ -branches bound to meromycolate branches.

A cascade of enzymes is involved in the processing and maturation of mycolic acids to form three variants of meromycolate, all of which are needed for complete virulence during infection<sup>35</sup>. Mycobacterial cells are capable of adapting to their surroundings in order to survive. The growth and division of bacterial cells lead to the formation of subsequent heterogeneous daughter cells<sup>36,37</sup>. During infection some of these cells can remodel their cell walls to enable survival<sup>38</sup>.

Cell proliferation begins with the elongation of the original cell followed by division into two subsequent daughter cells. These cells grow and divide asymmetrically to form unequally sized

daughter cells with different growth rates and cell wall compositions, which could be advantageous for survival in extreme host environments<sup>39</sup>.



**Figure 1.1: The cell envelope of mycobacteria.** The mycobacterial envelope consists of four layers made up of different substances<sup>33</sup>. NAM = N-acetylmuramic acid; NAG =N-acetyl glucosamine.

In the presence of unfavorable stimuli such as respiratory poisons, oxidative stress and hypoxia, *M.tb* goes into a quiescent physiological state known as latency or dormancy. This is controlled by the transcription factor dormancy survival regulator (DosR)<sup>40</sup>. During latency, the mycobacteria remain viable for long periods of time without undergoing any morphological differentiations or growth, and are phenotypically resistant to drugs<sup>41</sup>. During hypoxia, DosR levels increase up to five-fold compared to aerobic conditions<sup>42</sup>. DosR has to be activated before complete hypoxia if the mycobacteria is able to survive, to allow the transcription and translation of survival and adaptation mechanisms before energy depletion<sup>43</sup>. A group of genes known as enduring hypoxic response (EHR) maintains the mycobacteria during latency, until conditions become aerobic and favorable<sup>44</sup>. Initial alterations in gene expression during the chronic phase can reduce the amount of protective CD4<sup>+</sup>IFN- $\gamma$  T cells in the lungs. Down-regulation of genes expressed during growth followed by the up-regulation of dormancy genes lead to improper, inadequate stimulation of CD4<sup>+</sup> T cells and insufficient IFN- $\gamma$  production<sup>45</sup>. DosR activation also inhibits the establishment of a sufficient Th1 cell response against the mycobacteria<sup>46</sup>. However,

dormant mycobacteria are incapable of blocking maturation of phagosomes and can thus not disrupt host immunity<sup>47</sup>.

### 1.2.2. The emergence of a human pathogen

*Homo sapiens* and *M.tb* have been interacting with each other for thousands of years. This is supported by morphological evidence in human fossils, the presence of mycobacterial deoxyribonucleic acid (DNA) in human remains and literary descriptions of the disease recorded throughout human history<sup>48</sup>. The characteristic skeletal changes associated with TB (e.g. collapsed vertebrae, osteomyelitis, periosteal lesions) have been reported in the earliest cattle breeders of Egypt 3500 - 2650 BC and Sweden 3200 - 2300 BC respectively<sup>49-51</sup>. Although reports of older cases based exclusively on skeletal changes exist<sup>52,53</sup>, these are the first confirmed using ancient DNA. Polymerase chain reaction (PCR) was used to amplify the insertion sequence IS6110 – a segment consisting of 123 base pairs thought to be specific to the MTBC<sup>5,54</sup>. *M.tb* was detected in four of 11 specimens studied, which included bones between 300 and 1400 years old originating from Europe, Turkey and Borneo. Sequence data has subsequently been used to confirm these results<sup>55</sup>. Identical methods were used in 1994 to detect *M.tb* in the lung tissue of a 1000-year-old Peruvian mummy<sup>56</sup>.

Although sequences similar to IS6110 have subsequently been identified in other mycobacteria, the use of this insertion sequence for MTBC ancient DNA detection is still one of the most popular methods<sup>57</sup>. IS6110 is highly conserved and *M.tb* usually contain multiple, easily detectable copies. Spacer oligonucleotide typing (spoligotyping) has also been used in several studies to genotype MTBC strains<sup>58-63</sup>. It is, however, difficult to distinguish between the different strains of the MTBC using spoligotyping and IS6110 insertion elements. With the advances in sequencing technologies leading to new genomic methods such as whole genome sequencing (WGS) in combination with bioinformatics, the accumulation of several deletions over time have been observed in the MTBC, providing methods to distinguish between the individual species and the different lineages of this complex<sup>64</sup>. Researchers subsequently had to change their initial hypotheses regarding the origin of human pathogens, namely that TB was transferred from animals to human hosts during domestication, to the now more commonly

accepted theory that the two strains capable of infecting humans and animals emerged around the same time<sup>65,66</sup>.

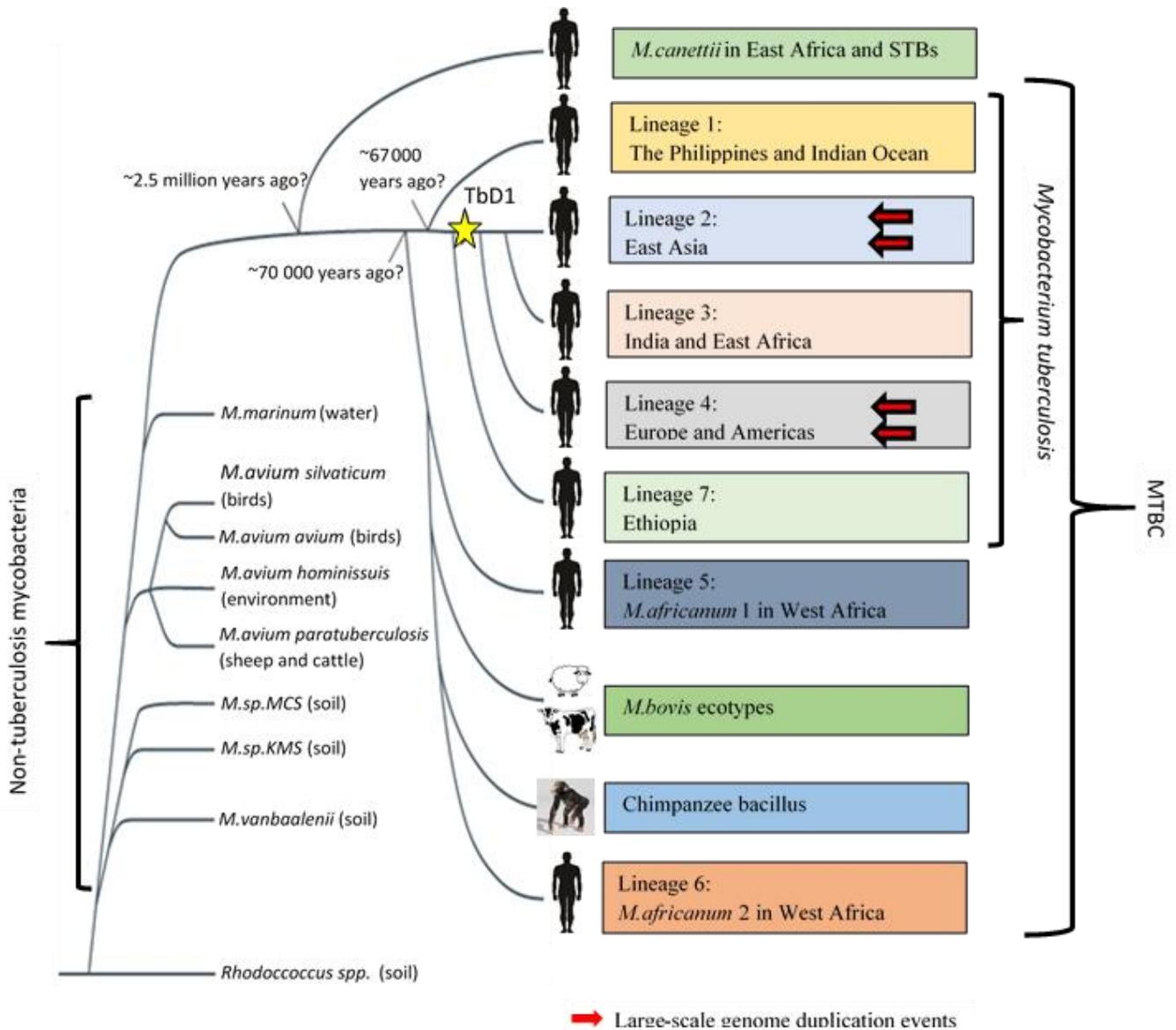
In 1905 Robert Koch isolated H37Rv, which today is a commonly-used laboratory *M.tb* strain. The genome of H37Rv is 4.4 Mb large, rich in GC-sequences and consists of approximately 4000 genes<sup>67</sup>. It contains large amounts of repetitive DNA regions, in particular insertion sequences, duplicated housekeeping genes and several multi-gene families<sup>67</sup>. Sequencing of the *M.tb* genome led to the development of a DNA microarray that enabled the investigation of the genomic composition of closely related mycobacteria via comparative genome hybridizations<sup>68</sup>. In combination with pulse-field gel electrophoresis approaches, these studies led to the identification of regions of variable genomic composition present in some of the MTBC strains<sup>69,70</sup> and a phylogenetic tree of the MTBC (Figure 1.2) was constructed based on the distribution of these regions. The phylogeny supports the new hypotheses that animal-infecting mycobacterial species diverged from human-adapted TB strains, since *M.bovis* actually represent species derived from and related to *M.tb*. This suggests that human *M.tb* strains did not arise ~10 000 years ago with animal domestication, but rather much earlier<sup>28,71,72</sup>.

Since *M.tb* and smooth tubercle bacilli (STB) strains are both associated with human TB, it could be argued that their last common ancestor potentially infected humans. With the use of a molecular clock analysis it was suggested that a common ancestor for both strains existed as early as 2.8 million years ago, implying that mycobacterial disease in hominids predates *Homo sapiens* as far back as *Homo habilis*<sup>73</sup>. The MTBC has low genetic diversity and shows evidence of single clonal expansion from the progenitor population<sup>74</sup>, potentially indicating the emergence of these pathogens from a genetic bottleneck when the MTBC split from its progenitor population<sup>29</sup>.

### **1.3. IMMUNE RESPONSE AGAINST *M.tb* INFECTION**

The role of the innate immune response and its mediators, natural killer (NK) cells, in TB immunity is not completely understood despite compelling evidence of its involvement. *M.tb* is most commonly transmitted through the respiratory system via aerosol deposition of the bacilli in lung alveoli<sup>75</sup>. Alveolar macrophages phagocytose the invading pathogen. Upon stimulation of

toll-like receptors (TLRs), the innate immune response is activated and its inflammatory mediators released to activate NK cells, neutrophils and later, T cells to provide early defense<sup>76</sup>.



**Figure 1.2: Evolutionary association between MTBC members and selected mycobacteria.** The specific deletion event for lineages 2, 3 and 4 of *M.tb* are indicated by TbD1. All indicated species form part of the *Mycobacterium* genus<sup>77</sup>. MTBC = *Mycobacterium tuberculosis* complex; Spp. = Species pluralis; *M* = *Mycobacterium*; M.sp.KMS = *Mycobacterium* species KMS; M.sp.MCS = *Mycobacterium* species MCS; TbD1 = *Mycobacterium tuberculosis* specific deletion; STBs = Smooth tubercle bacilli.

However, these responses will frequently not be able to effectively eradicate the infections, leading to spread of the infection to the draining lymph nodes via dendritic cells (DCs) and

macrophages and to other organs through the bloodstream. In immune-competent individuals, cluster of differentiation (CD) 4<sup>+</sup> and CD8<sup>+</sup> T cells will drive the formation of granulomas to contain the infection<sup>76</sup>. Although *M.tb* bacilli are contained in granulomas, they are not destroyed which leads to chronic, sub-clinical infection known as latent TB infection (LTBI). It is clear that both the innate and adaptive immune pathways are critical in fighting *M.tb* infections<sup>78-80</sup>.

### **1.3.1. Innate immune responses to *M.tb* infection**

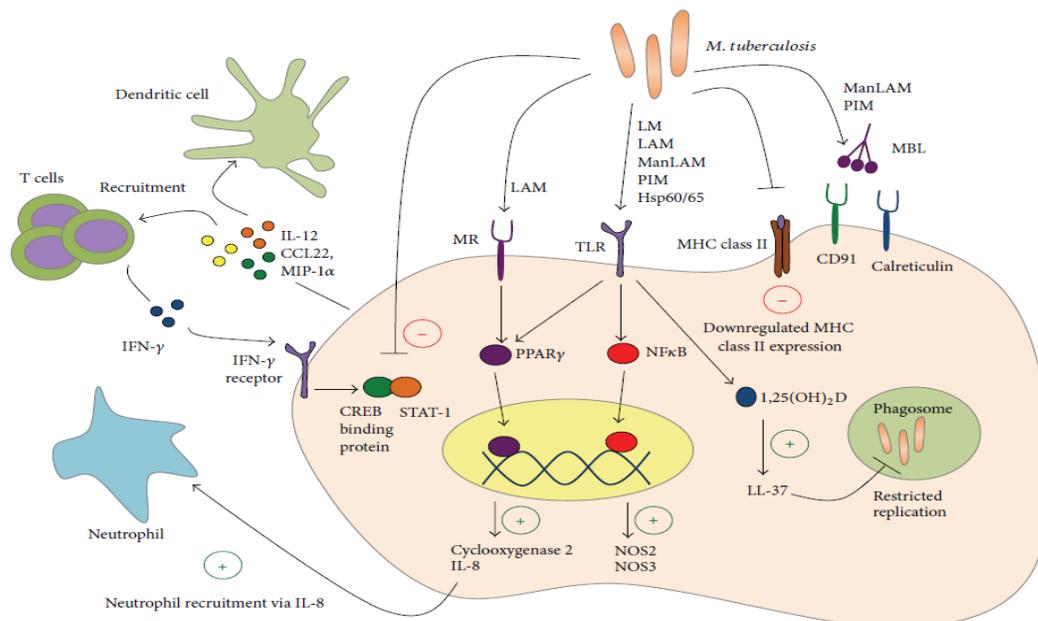
Macrophages, neutrophils, DCs and NK cells are the four major components of innate immunity in humans<sup>81</sup>.

#### ***1.3.1.1. Macrophages***

Following aerosol infection of the lungs, *M.tb* is phagocytosed by alveolar macrophages (Figure 1.3). Macrophages serve as *M.tb* replication sites and reservoirs in alveolar granulomas during early interactions. Several cell surface receptors recognize mycobacteria and are implicated in phagocytosis of *M.tb*, including TLRs (-2,-4 and -9), collectins (mannose-binding lectin and surfactant proteins A and D) and C-type lectins (Dectin-1, mannose receptor and Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin [DC-SIGN])<sup>82</sup>. The C-type lectin Mannose Receptor (MR) activates the expression of macrophage peroxisome proliferator activated receptor gamma (PPAR $\gamma$ ) upon binding to *M.tb*<sup>83,84</sup>. This interaction, along with the accumulation of neutrophils, regulates inflammatory responses during infection<sup>83</sup>. Initial receptor-mediated interaction between macrophages and *M.tb* can thus manipulate inflammatory responses, and the bacteria could potentially influence these responses to favor its dissemination and survival.

Collectins recognize and remove apoptotic cells<sup>85</sup>. *M.tb* accumulates in non-acidified sections of functioning macrophages, suggesting that the mycobacteria manipulates phagosomes in macrophages to provide preferential niches for itself<sup>86,87</sup>. The recognition of the pathogen associated molecular patterns (PAMPs) of *M.tb* by TLR-2 activates anti-microbial functions such as nuclear factor kappa-B (NF- $\kappa$ B) signaling in macrophages<sup>88</sup>. A critical and well-established component of human immunity against *M.tb* infection is the IFN- $\gamma$ /IL12 axis<sup>85-88</sup>. Gene profiling studies provided evidence of the crucial role of interferon gamma (IFN- $\gamma$ ) during infection with

*M.tb*. It should however be noted that IFN- $\gamma$  alone cannot inhibit replication of *M.tb* in macrophages. Additional immunological factors are required for optimal anti-mycobacterial immunity, because *M.tb* decreases responsiveness to IFN- $\gamma$  stimulation by inhibition of the crucial STAT-1-Cyclic adenosine monophosphate (cAMP) response element binding (CREB) protein-protein interactions<sup>93-95</sup>.



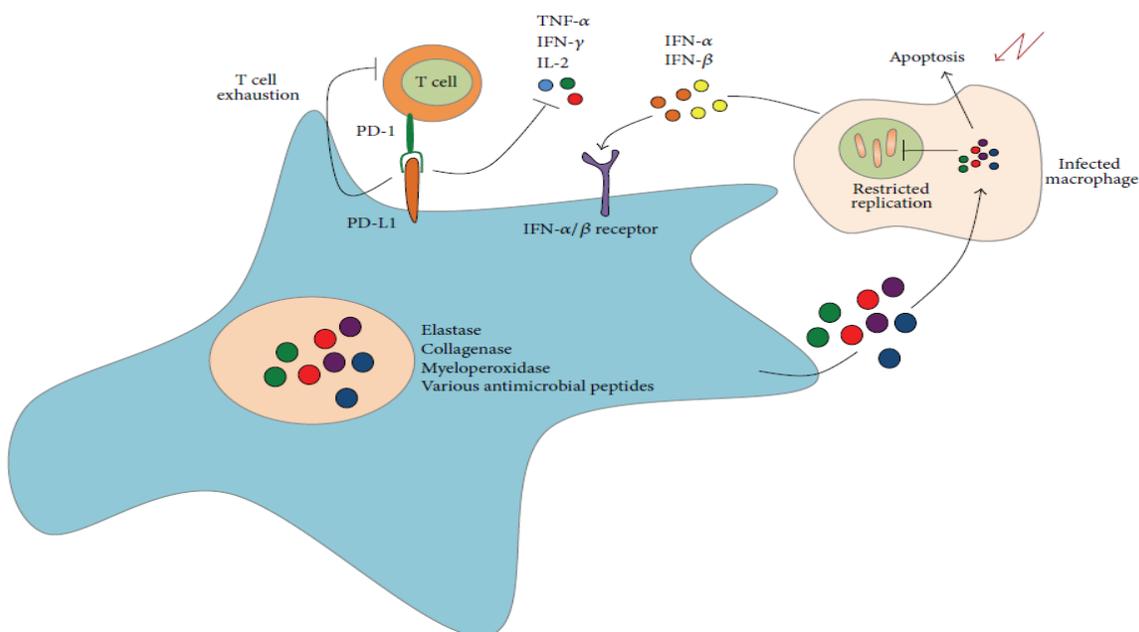
**Figure 1.3: Macrophage response to *M.tb* infection.** Various receptors on alveolar macrophages recognize *M.tb* to recruit other immune cells to the infection site and inhibit bacterial replication<sup>81</sup>. IFN- $\gamma$  = interferon gamma; IL = interleukin; CCL = Chemokine ligand; MIP-1 $\alpha$  = Macrophage inflammatory protein -1 alpha; MR = Mannose receptor; LAM = Lipoarabinomannans; TLR = Toll-like receptor; MHC = Major histocompatibility complex; Hsp = Heat shock protein; PIM = phosphatidylinositol mannosides; ManLAM = Mannosylated lipoarabinomannans; LAM = Lipoarabinomannans; LM = Lipomannans; *M.tuberculosis* = *Mycobacterium tuberculosis*; MBL = Mannose-binding lectin; CD = Cluster of differentiation; NF $\kappa$ B = Nuclear factor kappa B; NOS = Nitric oxide synthase; CREB = Cyclic adenosine monophosphate response element binding; STAT = Signal transducer and activator of transcription; PPAR $\gamma$  = Peroxisome proliferator activated receptor gamma.

Other immunological factors important for fighting *M.tb* infections include IL1 $\beta$  and similar pro-inflammatory cytokines at different stages of infection<sup>96-101</sup>, IL12, macrophage inflammatory protein-1 $\alpha$  (MIP-1 $\alpha$ /CCL3) and macrophage-derived chemokine CCL22 (MDC)<sup>96</sup>. Furthermore, *M.tb* infection downregulates the expression of major histocompatibility (MHC) class II molecules on macrophages through lower class II trans-activator (CIITA) expression, which may lead to the decreased adaptive immune response caused by weakened recognition of

macrophages by T cells<sup>102</sup>. In the presence of suitable activating stimuli from T cells, it is clear that macrophages are sufficient for *M.tb* restriction, although basic questions regarding infected macrophages in human TB remain unanswered.

### 1.3.1.2. Neutrophils

Following infection, mycobacteria are also phagocytosed by neutrophils which causes the release of anti-microbial effectors from their granules<sup>81</sup>. This very potent population of effector cells is able to mediate the immune system's anti-mycobacterial activity, as well as TB immunopathology (Figure 1.4). Neutrophils undergo respiratory burst, releasing factors such as collagenase, myeloperoxidase and elastase which damage both host and bacterial cells<sup>81</sup>.



**Figure 1.4: Neutrophil response to *M.tb* infection.** A selection of anti-microbial enzymes are produced by neutrophils upon *M.tb* infection to limit growth of the bacteria in macrophages through apoptosis of infected cells<sup>81</sup>. PD-1 = Programmed cell death; PD-L = Programmed death ligand; TNF-α = Tumor necrosis factor alpha; IFN-γ = Interferon gamma; IL = Interleukin; IFN-β = Interferon beta.

In addition to their role in release of antimicrobial peptides (AMPs), neutrophils also play immune-regulatory roles during TB infection. During chronic infection, the lymphocyte response becomes exhausted or dysfunctional upon binding of programmed death ligand 1 (PD-L1), present on myeloid cells, to their receptors on lymphocytes<sup>103–105</sup>. The elevated expression of

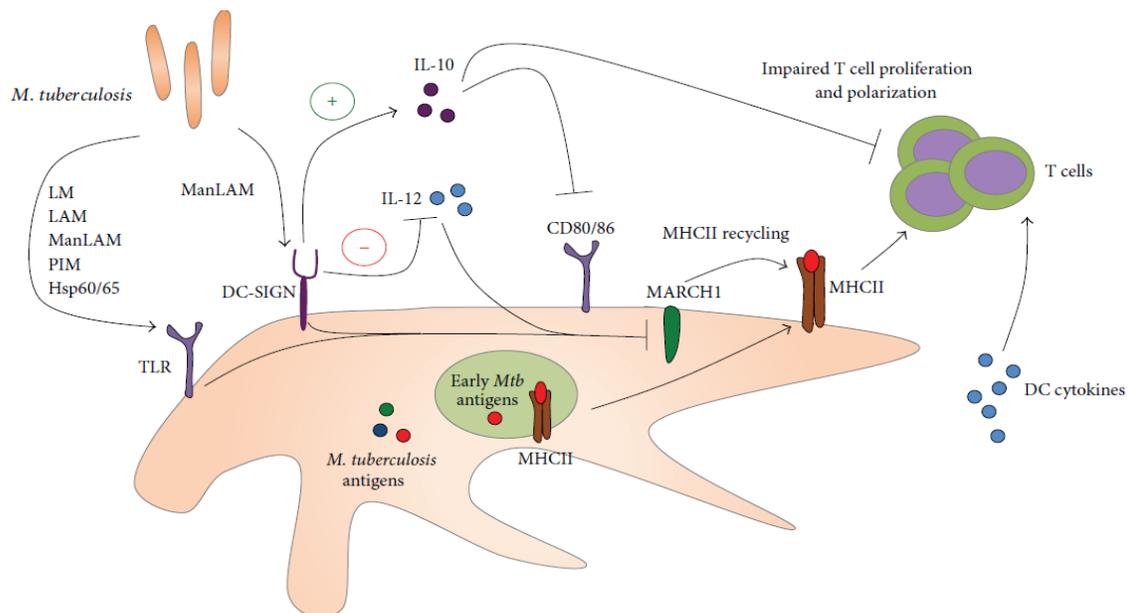
PD-L1 in whole blood of patients with active TB is caused by the surface expression of these ligands on neutrophils. Interestingly, a recent transcriptional profiling study found a 393 blood-based transcript profile that was able to differentiate between active and latent TB. The authors subsequently derived an 86-gene transcriptional profile corresponding to neutrophil expression of type I and II IFN-inducible genes<sup>106</sup>. Different populations of T cells have been found in different tissue sections of *M.tb*-infected humans, leading researchers to hypothesize that different populations of innate immune cells exist in the different tissues at various time-points during infection. A large proportion of the cells infiltrating the lungs during TB infection consists of neutrophils, highlighting the importance of determining the roles of these cells in lung tissue, immunopathology and uncontrolled inflammation<sup>81</sup>.

### ***1.3.1.3.Dendritic cells***

Dendritic cells link innate and adaptive immunity. These primary antigen-presenting cells (APCs) activate the adaptive immune response by presenting antigens to effector cells, their costimulatory capacity and by producing cytokines. DCs derived from human monocytes recognize *M.tb* ligands through expression of receptors such as DC-SIGN, mannose receptors, CD11b and CD11c. Following infection, *M.tb* binds to DC-SIGN through Mannosylated Lipoarabinomannans (ManLAM) which results in the production of the pro-inflammatory cytokine IL10, and the impairment of DC maturation (Figure 1.5). It is thought that these unfavorable, immunosuppressive conditions may contribute to *M.tb* survival<sup>107</sup>. Furthermore, ManLAM also prevents the production of IL12 via DC-SIGN and mannose receptor binding<sup>108</sup>. This suggests that *M.tb* is capable of inhibiting activation of the host's adaptive immune response to infection through subversion of DC function. IL12 orchestrates the innate and adaptive immune responses, and is crucial for the eradication of intracellular *M.tb* due to its role in the activation of Th1 cells<sup>109</sup>. Mutations in genes affecting IL12 or its receptor are associated with susceptibility to mycobacterial infections in humans<sup>110–113</sup>, while mice deficient in IL12 are susceptible to *M.bovis* and BCG infections<sup>114</sup>.

The extent, nature and kinetics of T cell responses are influenced by modulating antigen presentation, DC maturation and production of cytokines, giving *M.tb* the necessary time to establish itself in the lungs. Studies have shown that the ability of *M.tb*-infected DCs to stimulate

lymphoproliferation of memory and naïve CD4 and CD8 cells<sup>115</sup> is impaired, which suggests the infection of DCs by *M.tb* weakens T-cell responses. Additionally, since *M.tb* is slow-growing, the availability of antigens during early infection may also contribute to inadequate responses of T cells. It is speculated that excess T cells specific to irrelevant antigens at the different stages of infection might be present due to inappropriate presentation of *M.tb* antigens during activation of adaptive immunity<sup>116–118</sup>. Taken together, this suggests that DCs are crucial in defense against *M.tb* infection, and future studies should focus on these cells to elucidate why specific antigens are absent or under-represented at T cell levels.



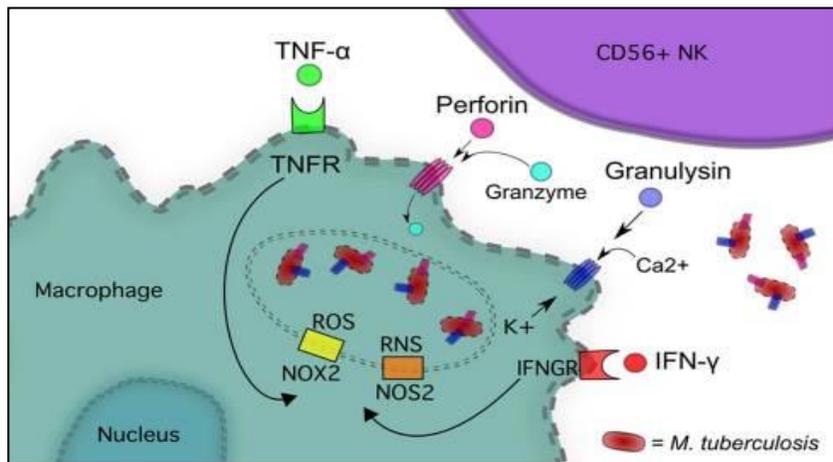
**Figure 1.5: Dendritic cell response to *M.tb* infection.** DCs are very important for the activation and differentiation of T cells through the presentation of antigens<sup>81</sup>. MHCII = Major histocompatibility complex class II; DC = Dendritic cell; MARCH1 = Membrane associated ring-CH-type finger 1; CD = Cluster of differentiation; IL = Interleukin; ManLAM = Mannosylated lipoarabinomannans; LAM = Lipoarabinomannans; LM = Lipomannans; DC-SIGN = Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin; TLR = Toll-like receptor; Hsp = Heat shock protein; PIM = Phosphatidylinositol mannosides.

#### 1.3.1.4. Natural killer cells

Natural killer cells are potent cytotoxic granular lymphocytes responsible for non-MHC-restricted cytotoxicity of pathogens<sup>66,67</sup>. Nucleated cells express MHC class I molecules which bind to inhibitory receptors such as leukocyte Ig-like receptors (LIRs), killer cell Ig-like receptors (KIRs) and CD94/NKG2 receptors on NK cells to inhibit cytotoxicity. Several activating

receptors (NKp30, -44, and -46) are also present on NK cells and their expression is increased upon lysis of target cells<sup>121–123</sup> (Figure 1.6). NK cells are activated by IL2 and IL12<sup>78</sup>.

The release of IL2 from neutrophils, DCs and macrophages following *M.tb* infection stimulates NK cells to produce, amongst others, IFN- $\gamma$  and IL22 which are able to restrict the growth of *M.tb* through enhancement of phagosomal fusion<sup>124</sup>. NK cells produce granules in the cytoplasm that contain granzymes, perforin and granulysin. Perforin enters the membranes of target cells and creates pores through which granzymes and granulysins are transported<sup>125</sup>. Granzymes induce target cell apoptosis either by directly activating Caspases 3 and 7 or via Bid proteolysis. Bid is a pro-apoptotic protein part of the Bcl-2 family and contains the BH3 domain, also known as the death domain<sup>126</sup>. Cytosolic Bid is cleaved by Caspase 8. The C-terminal end moves to the mitochondria where it is integrated into its outer membrane together with Bak and Bax<sup>127</sup>, to allow cytochrome c to exit into the cytoplasm and form the caspase reactivating complex<sup>128</sup>.



**Figure 1.6: *M.tb* elimination by direct functioning of natural killer cells.** NK cells inhibit growth of *M.tb* through TNF- $\alpha$ , IFN- $\gamma$ , perforin, granzyme and granulysin<sup>141</sup>. TNF- $\alpha$  = Tumor necrosis factor alpha; TNFR = Tumor necrosis factor receptor; ROS = Reactive oxygen species; NOX = NADPH oxidase; NOS = Nitric oxide synthase; RNS = Reactive nitrogen species; IFNGR = Interferon gamma receptor; IFN- $\gamma$  = Interferon gamma; CD = Cluster of differentiation; *M.tuberculosis* = *Mycobacterium tuberculosis*.

Granulysin interacts with lipids and, although it can effectively destroy extracellular *M.tb* on its own, it is not sufficient for intracellular *M.tb* killing without perforin – which enables it to disrupt oxidative phosphorylation and lipid metabolism<sup>129,130</sup>. Discharge of cytochrome c upon damage of lipid membranes initiates reactive oxygen species (ROS) production and apoptosis<sup>131</sup>. A variety of death receptor ligands are present on the surfaces of NK cells which, when bound to their receptors, activate apoptosis. Fas ligand (FasL) is one of these ligands. Its receptor (Fas) causes cell lysis when activated and is present on most cell types in the human body including

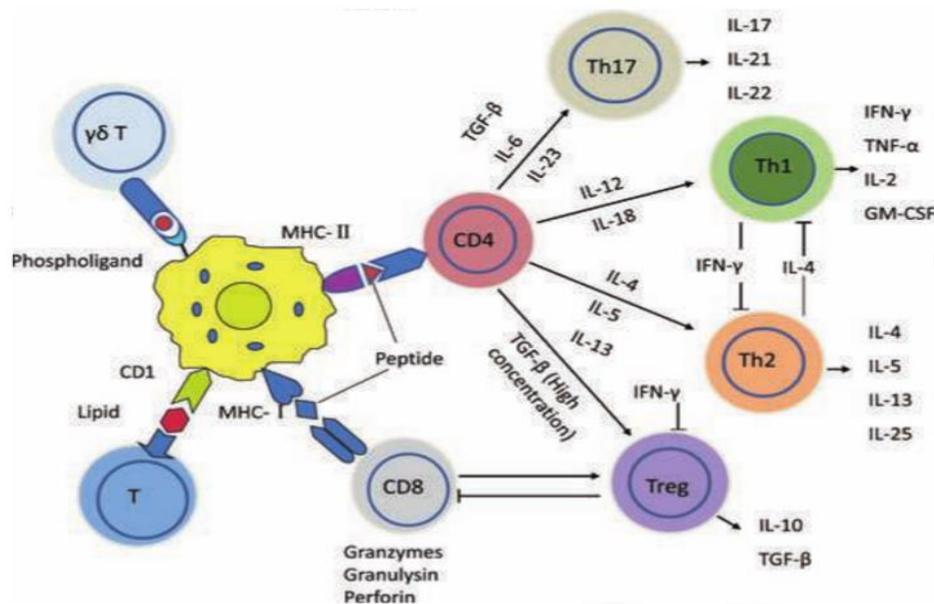
macrophages. When Fas and FasL bind, death-inducing signaling complexes (DISCs) are formed and can activate apoptotic pathways outside the mitochondria<sup>132</sup>. The viability of *M.tb* in infected macrophages is limited through this apoptotic Fas pathway<sup>133</sup>. CD40 ligand (CD40L) is also present on NK cells and, when bound to CD40 on macrophages and other antigen presenting cells, causes increased expression of CD80 and CD86 on macrophage surfaces<sup>134,135</sup>. NK cells also produce TNF- $\alpha$  and IFN- $\gamma$ , which activate the NF- $\kappa$ B pathway to up-regulate intercellular adhesion molecule 1 (ICAM-1) in the targeted cells, which in turn up-regulates cytotoxicity by increasing adhesion of NK cells to target cells<sup>136</sup>.

NK cells are responsible for the lysis of *M.tb*-infected monocytes and the reduced growth rate of mycobacteria<sup>124,137</sup>, and are present in advanced granulomatous lesions in the lungs of TB patients<sup>138</sup>. NKp44 expressed on the surfaces of these cells in the presence of IL2 and IL12 increases the activity of infected macrophages and leads to IFN- $\gamma$  production<sup>138,139</sup>. Expression of the natural cytotoxicity receptors NKG2D, NKp46 and NKp30 on NK cell surfaces also increases upon *M.tb* infection and, along with NKp44, up-regulates perforin and granulysin expression via the MAP kinase (MAPK) signaling pathway to increase *M.tb* degradation<sup>140</sup>.

### **1.3.2. Adaptive immune response to *M.tb* infection**

Macrophages and DCs infected with *M.tb* activate the adaptive immune system by presenting antigens to B and T cells (Figure 1.7). Activated neutrophils, DCs and macrophages produce IL-12p40, which in turn activates pulmonary DCs<sup>141</sup>. Apoptosis by macrophages leads to the release of more apoptotic vesicles capable of carrying *M.tb* antigens to healthy DCs to thus increase the frequency of antigen presentation<sup>142</sup>. The protective immunity consists of CD4 and CD8 T cells,  $\gamma$ - $\delta$  T lymphocytes and CD1-restricted specific T lymphocytes that produce IFN- $\gamma$ . CD4<sup>+</sup> T helper 1 (Th1) cells are crucial for protection against *M.tb* infection and elicit a much more potent IFN- $\gamma$  response than CD8<sup>+</sup> T cells<sup>143,144</sup>. Immune protection can be delayed in some organs due to limited amounts of CD4 T cells that result in decreased dissemination of active CD8 T cells from lymph nodes<sup>145</sup>. Due to the secretion of perforin, granzymes and granulysin by CD8 cytolytic T lymphocytes (CTLs), immune protection is provided against secondary challenges without the presence of CD4<sup>+</sup> T cells<sup>143</sup>.

Four types of CD4 Th cells exist, namely Th1, Th2, Th17 and Treg cells. Th1 cells cause granulocyte and macrophage activation and maturation, as well as CTL and Th1 cell stimulation by producing granulocyte-macrophage colony-stimulating factor (GM-CSF), lymphotoxin, TNF- $\alpha$ , IFN- $\gamma$  and IL2. Although Th2 cells stimulate antibody production by B cells by producing IL4, IL5, IL10 and IL13, they also inhibit Th1 immune responses. Th17 cells are involved in the early stages of host immunity and influence the production of defensin and the transferal of monocytes and neutrophils to inflammation sites by producing IL17, IL17F, IL21 and IL22<sup>146-148</sup>. The three subsets of Th cells are regulated by a large variety of Treg cells that also produce IFN- $\gamma$ , while CD8<sup>+</sup> Treg cells are involved in the inhibition of cytokine production and T cell proliferation<sup>149</sup>. Sets of T cells can be influenced in different ways by a variety of cytokines, as evidenced by Th2 cell development induction caused by IL4, IL5 and IL13 compared to the promotion of Th1 cell development mediated by IL12, IL18 and IFN- $\gamma$ .



**Figure 1.7: Adaptive immunity to TB infection.** A network of cells are involved in adaptive immunity, with the main function of antigen presentation to T cells to cause cytokine production by these cells as well as their targets<sup>150</sup>.  $\gamma\delta$  T = gamma delta T cells; CD = Cluster of differentiation; MHC = Major histocompatibility; TGF- $\beta$  = Tumor growth factor beta; Th = T-helper; IFN- $\gamma$  = Interferon gamma; TNF- $\alpha$  = Tumor necrosis factor alpha; GM-CSF = Granulocyte macrophage colony-stimulating factor; Treg = T regulatory.

Although not traditionally acknowledged, B cells also play a significant role in effective protection against *M.tb* infection via complement activation and cellular immune response interactions<sup>151,152</sup>. Lectin as well as the classical and alternative pathways are activated directly

by *M.bovis* BCG to cause C3b to bind to the macromolecules on the surfaces of mycobacteria and thus lead to killing of the mycobacteria<sup>153</sup>. Upon *M.tb* infection, memory T (MT) cells form, which develop into either central (CCR7<sup>hi</sup>, CD62L<sup>hi</sup>, CD69<sup>lo</sup>) or effector (CCR7<sup>lo</sup>, CD62L<sup>lo</sup>, CD69<sup>hi</sup>) MT cells<sup>144</sup>. Directly after interaction with antigens, MT cells start proliferating and production of TNF- $\alpha$ , IL2, GM-CSF, lymphotoxin and IFN- $\gamma$  is initiated<sup>146</sup>. However, a precise biomarker signature to indicate protective immunity is still not defined.

### 1.3.3. The type I cytokine response

The manner in which the human host responds to mycobacterial infection is largely dependent on the integrity of the host's type I cytokine response<sup>154</sup>. Mycobacterial- or salmonella-associated molecular patterns activate phagocyte-expressed pattern-recognition receptors (PRRs), leading to the production of cytokines like interleukins. IL23, -18 and -1 $\beta$  bind to their receptors situated on NK or T cells to cause IFN- $\gamma$  production<sup>155</sup>. IFN- $\gamma$  activates the production of IL12 and enhances IL18 production by binding to its receptors on monocytes and macrophages. Together with TNF- $\alpha$ , it also stimulates infected macrophages to initiate several microbicidal mechanisms to control the bacterial infection.

IL12 and -18 also stimulates the Th1 pathway to cause activation and differentiation of naïve T cells<sup>156</sup>. During the subsequent adaptive immune response IFN- $\gamma$ , responsible for control of the chronic stage of mycobacterial infection, is mainly produced by Th1 cells. This means that the type-1 cytokine pathway, with IL23, -12 and IFN- $\gamma$  as the main cytokines, creates an innate immune signaling loop linked to the adaptive immune response via Th1 activity<sup>156</sup>. IL12 and -23, as well as their receptors, are very closely related: IL12 phosphorylates STAT4, while IL23 causes phosphorylation of STAT1, STAT3, STAT4 and STAT5 to a lesser extent, all of which finally results in the production of IFN- $\gamma$  and IL10<sup>157,158</sup>. IL23 is responsible for the proliferation of memory T cells, proliferation and survival of CD4 cells (Th17 cells), which produce IL17, and maintaining this IL17 production via Th17 cells<sup>159,160</sup>.

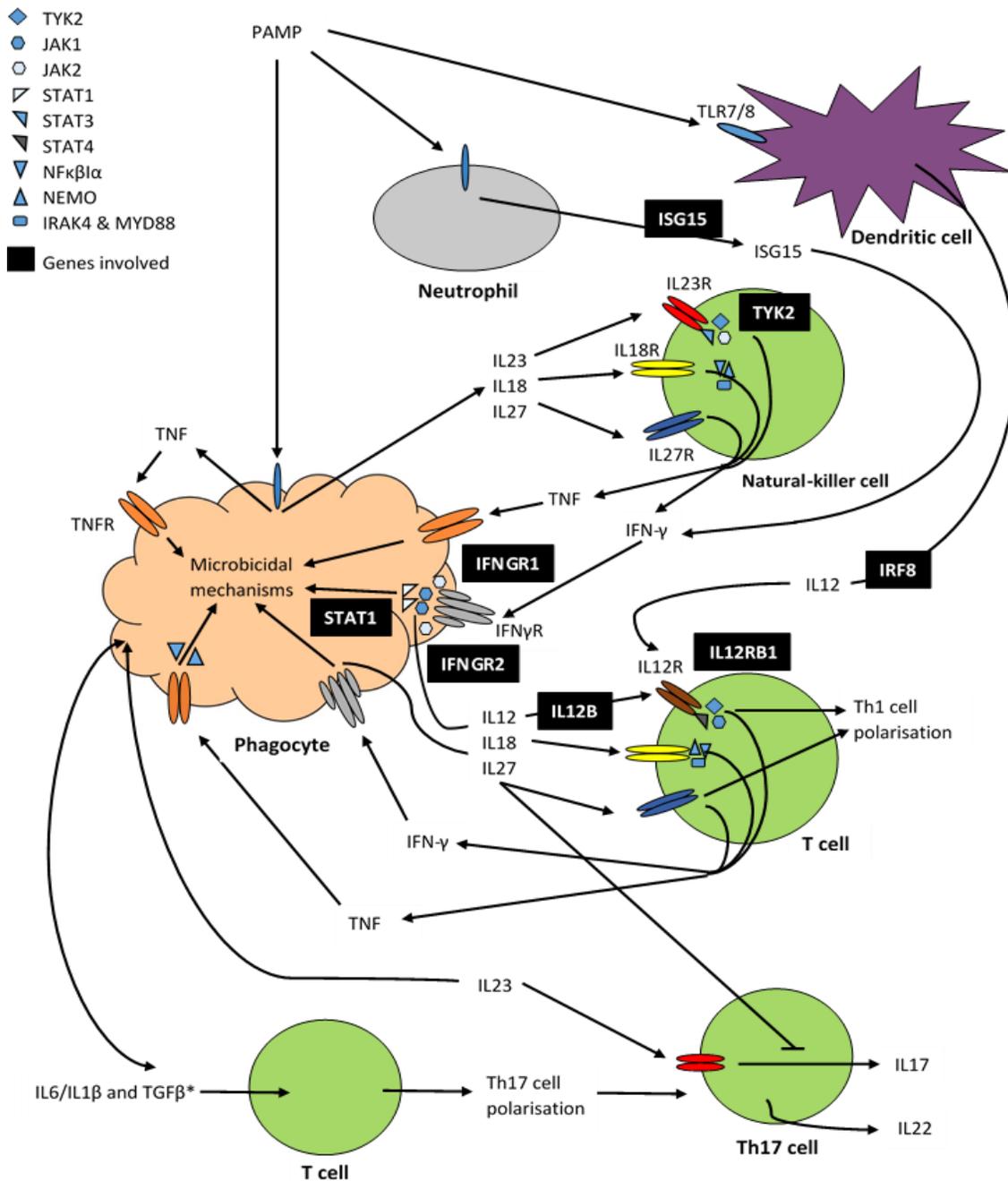
The IL23/IL17 pathway is involved in several inflammatory and autoimmune disorders. Th17 cells produce IL17 and -22, both pro-inflammatory cytokines regulated by IL23<sup>161</sup>. These cytokines play a critical role in mucosal host defense<sup>162</sup>, and are important factors in the host's

adaptive immune response. IL27 binds to its receptors situated on, amongst others, B and T lymphocytes, mast cells, monocytes, NK cells, endothelial cells and DCs<sup>163</sup>.

IFN- $\gamma$  alone can stimulate IL27 production by DCs and macrophages, although it is usually produced in response to certain TLR ligands without the need for IFN- $\gamma$ <sup>164</sup>. IL27 combined with IL18 and IL12 have the function of activating naïve T cells to proliferate, causing differentiation of Th1 cells and leading to IFN- $\gamma$  production. IL27 can also antagonize IL23 by suppressing the production of pro-inflammatory cytokines via activated CD4 cells to inhibit IL17 production<sup>165</sup>, thus also inhibiting excessive inflammation. The phosphorylation of different STAT pathways is the main cause of these opposing roles of IL23<sup>166</sup>. The type I cytokine response to mycobacterial infection is described in detail in Figure 1.8.

#### **1.3.4. Disease outcome**

Through thousands of years of co-evolution between human hosts and *M.tb*, the mycobacterium developed extensive survival mechanisms allowing it to remain latent inside the host for decades without causing active disease, but remaining in constant interaction with the host's immune system<sup>167</sup>. If the immune system is then impaired or compromised at any time due to HIV infection, diabetes, aging, smoking or silicosis for example, the *M.tb* is reactivated and causes active disease<sup>168</sup>. Only approximately 5% of individuals exposed to *M.tb* develop active disease within 24 months of infection, either without or after a very short latency stage. This is known as primary TB<sup>169</sup> and is associated with severe progression and dissemination to extra-pulmonary organs through the bloodstream. This is known as extra-pulmonary TB, the most life-threatening of which manifests either as miliary TB or infection of the central nervous system (CNS). The largest risk factor is age, with children below the age of one year at biggest risk<sup>170</sup>. BCG vaccines currently used in some countries to vaccinate infants against TB provide limited protection to children against severe primary TB, while antibiotics provide treatment for already infected children<sup>171</sup>. Unfortunately, an estimated 140 000 children still die annually due to this disease<sup>16</sup>.



**Figure 1.8: Type-1 cytokine pathway.** Several cytokines are produced upon recognition of PAMPs by PRRs to initiate the type-1 cytokine pathway. Most of the genes associated with MSMD is present in this pathway, and indicated in black<sup>156</sup>. TYK = Tyrosine kinase; JAK = Janus kinase; STAT = Signal transducer and activator of transcription; NF $\kappa$ B $\alpha$  = Nuclear factor kappa B inhibitor alpha; NEMO = NF-kappa-B essential modulator; IRAK = Interleukin 1 receptor associated kinase; MYD = Myeloid differentiation primary response gene; PAMP = Pathogen-associated molecular patterns; TLR = Toll-like receptor; ISG = Interferon-stimulated genes; IL = Interleukin; ILR = Interleukin receptor; TNF = Tumor necrosis factor; TNFR = Tumor necrosis factor receptor; IFN- $\gamma$  = Interferon gamma; IFN $\gamma$ R = Interferon gamma receptor; IFNGR = Interferon gamma receptor; IRF = Interferon regulatory factor; TGF $\beta$  = Tumor growth factor beta; Th = T helper.

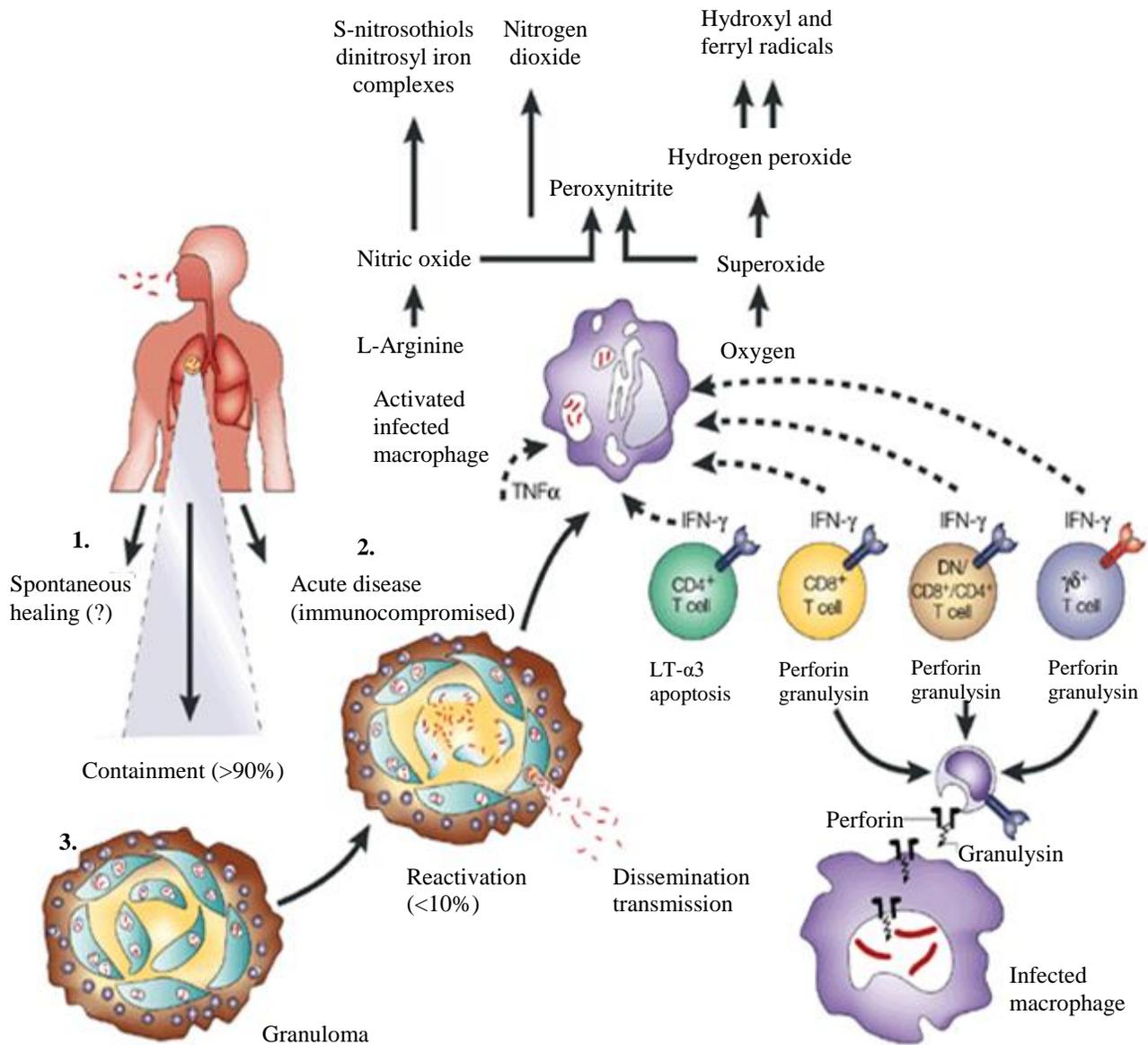
The other main form of clinical TB is pulmonary TB (PTB) in adults, usually occurring due to reinfection or reactivation of a latent infection years after initial exposure<sup>172</sup>. Most individuals exposed to *M.tb* develop LTBI, testing positive for the tuberculin skin test (TST) and/or interferon gamma release assay (IGRA) used as TB diagnostic tools, while presenting with no symptoms<sup>75</sup>. The majority of LTBI will never progress to active disease, with only about 5-10% of cases resulting in chronic PTB in adults<sup>169</sup>. The development of active TB in adults is clearly influenced by several complex genetic mechanisms thought to differ significantly from those involved in childhood primary TB. Understanding the process on molecular level can be immensely beneficial in disease treatment and prevention<sup>23</sup>.

Figure 1.9 illustrates the three different potential outcomes for a human host once infected with *M.tb*: 1) unsuccessful infection leading to spontaneous healing, although the frequency of this is miniscule; 2) disease development directly after infection in immune-compromised hosts; 3) initial containment of mycobacteria with disease developing later in life as a result of reactivation.

Granulomas act as the sites of infection, pathology, persistence as well as protection, while effector T cells and macrophages aid in controlling TB. T cells produce IFN- $\gamma$  and TNF- $\alpha$ , which are important in activating macrophages. This activation stimulates phagosome maturation and lead to the production of antimicrobial molecules like RNI and ROI<sup>173</sup>. PTB, involving only the lungs, is the most common type of TB and is responsible for ~90% of all cases<sup>174</sup>. It typically causes chest pain and extensive coughing that produces sputum and in some cases blood. Chronic or sustained TB may result in scarring of the upper lung lobes.

#### **1.4. SUSCEPTIBILITY TO *M.tb* INFECTION**

Extensive research with regards to TB infection and the development of active disease has been conducted since the 1910's, the majority of which epidemiologically and clinically proved the involvement of host genetic factors<sup>175-177</sup>. The existence of 5-10% of individuals who develop clinical disease in the absence of risk factors such as drug and alcohol abuse, old age, the use of immune suppressors, diabetes mellitus and HIV infection led to the investigation and identification of genetic factors that increase susceptibility to TB<sup>178</sup>.



**Figure 1.9: Primary features of TB: initial infection to activated host defense**<sup>179</sup>. TNF- $\alpha$  = Tumor necrosis factor alpha; IFN- $\gamma$  = Interferon gamma; CD = Cluster of differentiation; DN = Double-negative;  $\gamma\delta$  = gamma delta; LT = Lymphotoxin.

The variation in clinical responses to infection is dependent on the immune responses which, to a very large extent, is governed by the genetics of the host<sup>180–182</sup>. The involvement of host genetic factors has been shown in family aggregation and twin studies, with the latter showing a 2.5-fold higher concordance for clinical TB in monozygotic twin pairs compared to dizygotic twins<sup>183,184</sup>. Moreover, family studies have shown that in married couples, where one is sputum-positive for

*M.tb* infection, that the spouses with a family history of TB developed the disease more frequently than those without<sup>177</sup>.

The role of genetic factors in TB susceptibility was inadvertently proven in the 1928 Lübeck disaster when BCG contaminated with virulent *M.tb* was mistakenly administered to newborns as a TB vaccine. Of the 251 neonates vaccinated, 173 developed clinical and/or radiological signs of TB but survived, while 72 died from TB<sup>24,172</sup>. Although tragic, this accident proved the remarkable resilience of humans to *M.tb* infection. Symptoms spontaneously resolved in 68% of all infants who developed clinical disease. Different degrees of vaccine contamination was deduced, which resulted in a variety of symptoms in the neonates. This also highlighted the importance of dose in disease outcome. Overall, low contamination levels were associated with proper innate resistance to TB, while higher doses resulted in severe disease susceptibility. However, two neonates who received the lowest dose of *M.tb* died of the disease shortly after vaccination, probably indicating increased disease susceptibility<sup>24</sup>. Taken together, these findings point to variability of individual innate immunity, and potentially genetically-controlled resistance or susceptibility to TB<sup>172</sup>.

It was proven as early as the 1950's that, phenotypically, TB infection has an estimated heritability of more than 50%<sup>23</sup>. Anti-mycobacterial immunity has been reported to be very high in TB endemic areas, estimated at 71% for the TST, 68% for TNF- $\alpha$  production and 20-40% for IFN- $\gamma$  production<sup>185,186</sup>. It is thus evident that high heritability plays a crucial role in anti-mycobacterial immunity within the human host. Geographical origin also correlates directly with an individual's degree of resistance to *M.tb* infection, rendering those from ancestral origin free of TB increasingly vulnerable<sup>187</sup>. In cases where sudden outbreaks in populations with no or very little ancestral contact with *M.tb* took place, such as the case with native Americans from the Qu'Appelle Indian Reservation, the TB incidence was initially very high<sup>188</sup>. Individuals with European ancestry are also less likely to develop active TB than individuals with African ancestry<sup>189</sup>.

Several previous linkage, candidate gene and genome-wide association studies (GWAS) have successfully identified a number of TB susceptibility genes in different populations. However, it

is evident that many more unidentified susceptibility genes exist. The identification of these genes remains a daunting task given our incomplete understanding of host factors involved in protection against and susceptibility to TB. Disease development depends on interactions between several host, pathogen, risk and environmental factors<sup>190</sup>. Much research have been conducted to identify the specific combinations of pathogen and host<sup>191–199</sup> genotypes associated with TB susceptibility. Studies showing increased TB susceptibility in patients with HIV/AIDS<sup>200,201</sup>, patients on anti-TNF treatment<sup>202</sup> and in individuals with genetically impaired IFN- $\gamma$  immunity<sup>196</sup> proved the importance of the IL12/IFN- $\gamma$  immune response in controlling *M.tb* infection in humans. It also highlighted the significant role of CD4<sup>+</sup> T cells, TNF, IFN- $\gamma$  and IL12<sup>196,197,200,202–205</sup>.

Several environmental and other host risk factors to TB also exist. A multivariate analysis in the Gambian population identified family TB history as the highest risk factor, followed by the co-habitation of more than 10 adults, HIV infection, single marital status, smoking, male sex, house ownership and asthma<sup>206</sup>. However, several patients develop TB in the absence of any risk factors, implying the involvement of host genetic variation in disease susceptibility.

The identification of primary and acquired immunodeficiency disorders provided the basis for two studies showing that the cause of TB, especially in children, may be due to inborn errors of immunity<sup>196,207</sup>. These deficiencies are typically associated with increased susceptibility to infections, which includes mycobacterial infections. The most conclusive subsequent progress came from studies on patients diagnosed with MSMD, a PID characterized by infections with weakly virulent mycobacteria such as environmental mycobacteria (EM) and BCG vaccines. The main determinant of the human host's response to a mycobacterial infection is the type I cytokine pathway and its integrity<sup>172</sup>.

## **1.5. PRIMARY IMMUNODEFICIENCY DISORDERS**

Since the first description of PIDs in the early 1950's, more than 260 clinical entities have been characterized, most of which have well-defined molecular genetic causes<sup>208</sup>. PIDs are a diverse group of inherited disorders that affect the adaptive and innate immune systems to cause increased susceptibility to a broad range of different pathogens<sup>209</sup>. Mutations in about 130

different genes involved in impaired immune responses have been associated with more than 260 different Mendelian conditions<sup>210,211</sup>, with all the different aspects of immunity involved. The investigation of PIDs led to the identification of new genes and their products and pathways involved, as well as the clarification of the functions of previously discovered genes. The study of naturally-occurring mutants have effectively bettered our current understanding of the immune system as a whole<sup>10</sup>.

Because PIDs render patients vulnerable to infection with a wide range of microorganisms, they provide good candidates for *in vivo* assessment of the functions of effectors of immune responses during infectious challenges. The degree of predisposition, the range of infecting microorganisms and the fact that susceptibility to infectious diseases varies over time, are very important in detecting protective effectors in healthy individuals<sup>175</sup>. Several PID patients are extremely vulnerable to infection with *M.tb* and other related species<sup>212</sup>, indicating the involvement of essential components in immunity to mycobacterial infection such as phagocytic nicotinamide adenine dinucleotide phosphate (NADPH) oxidase, T cell-associated CD154, T cells and NEMO-dependent NF- $\kappa$ B. It is currently not possible to determine exactly which PIDs increase patients' susceptibility to *M.tb* infection<sup>212</sup>. The first evidence of inherited faults of the immune system causing increased susceptibility to TB was found when a large number of young children with rare PIDs all developed severe TB<sup>212</sup>. Table 1.1 provides a summary of the inherited disorders predisposing patients to mycobacterial infections.

Although the prevalence of PIDs is currently still speculative, Bousfiha and co-workers through the use of registry data for entire continents<sup>213-215</sup> and data from two large United States (US) surveys<sup>216,217</sup> projected PID prevalence to populations not yet investigated<sup>218</sup>. The prevalence data from Australian/New Zealand registries (5.6 per 100 000) was used to estimate the existence of at least 390 546 PID patients world-wide, with Africa contributing 58 572. Using the US surveys data, however, they calculated that over 6 million people worldwide currently live with a PID, with 902 631 individuals in Africa alone. In 2012, the available registries (ESID<sup>213</sup>, LAGID<sup>214</sup>, USIDNet<sup>219</sup>, North Africa<sup>220</sup>, Japan<sup>221</sup>, Iran<sup>222</sup> and Australia/New Zealand<sup>223</sup>) listed a total of 27 243 PID cases worldwide, only a small percentage of the expected PID cases<sup>218</sup>.

**Table 1.1: Inherited PIDs associated with mycobacterial susceptibility to BCG, NTM or *M.tb*.**

	NTM	BCG	TB	TB only*	Other infections†	Physiopathology
SCID	-	+	+	No	Yes	T cell defect
AD GATA2 deficiency	+	-	+	No	Yes	Quantitative defect of monocytes
CGD	+/-	+	+	No	Yes	Respiratory burst defect in all phagocytic cells
EDA-ID	+	+	+	No	Yes	Impaired CD40-dependent IL12 production
XR CD40L deficiency	+	+	+	No	Yes	Impaired CD40-dependent IL12 production
AR STAT1 deficiency	+	+	-	No	Yes	Impaired IFN- $\gamma$ response
AR IRF8 deficiency	-	+	-	No	Yes	Absence of monocytes and DC
AR TYK2 deficiency	-	+	+	Yes	Yes	Impaired IFN-c production
MSMD:						
- IFN- $\gamma$ R deficiencies	+	+	+	Yes	No	Impaired IFN-c response
- AD STAT1 deficiency	+	+	+	Yes	No	Impaired IFN-c response
- XR gp91phox deficiency	-	+	+	Yes	No	Respiratory burst defect in macrophages
- AD IRF8 deficiency	-	+	-	No	No	Absence of CD11c+ SC1c+ DC
- XR NEMO deficiency	+	+	+	No	No	Impaired CD40-dependent IL12 production
- IL12 and IL12R deficiencies	+	+	+	Yes	Yes	Impaired IFN-c production
- AR ISG15 deficiency	-	+	-	No	No	Impaired IFN-c production

AD = autosomal dominant inheritance; AR = autosomal recessive inheritance; BCG = Bacille Calmette Guerin vaccinated; CGD = chronic granulomatous disease; CD = Cluster of differentiation; EDA-ID = anhidrotic ectodermal dysplasia with immunodeficiency; ID = immunodeficiency; IFN- $\gamma$ R = interferon gamma receptor; IL = interleukin; IL12R = interleukin 12 receptor; MSMD = Mendelian susceptibility to mycobacterial disease; NTM = non-tuberculous mycobacteria; PID = primary immunodeficiency disorder; SCID = severe combined immunodeficiency disorder; TB = tuberculosis; XR = X-linked inheritance; GATA = globin transcription factor binding protein; STAT = signal transducer and activator of transcription; IRF = interferon regulatory factor; TYK = Tyrosine kinase; NEMO = nuclear factor-kappa-B essential modulator; ISG = Interferon-stimulated gene. \* = patients with only tuberculosis; † = The clinical phenotype of the patients may (yes) or may not or only rarely (no) include other infectious diseases<sup>172</sup>.

Many plausible reasons exist for this discrepancy, such as the under-reporting in developing countries due to limited resources available for accurate PID diagnosis; that the acquisition of registry data is based on voluntary reporting of PID patients by treating physicians dependent on the willingness of the specialist to participate in the process; and that many PIDs may not yet have been discovered<sup>218</sup>. In SA, the prevalence of PIDs is still unknown. However, based on a mid-2013 population of 52.9 million and on the assumption that the prevalence of PIDs in SA is

similar to that of well-resourced settings, it is estimated that the total of PID patients ranges between 2 850 and 45 723<sup>224</sup>. To date, fewer than 500 PID cases have been reported in SA with only about 300 listed on the SA National PID registry. This is partly due to the limitations of registry data mentioned, but also because PIDs are being masked by the endemic TB and AIDS pandemics.

In a preliminary investigation, 28 TB episodes were identified in 15 PID patients on the SA PID registry, translating to 5% of PID patients at risk of developing TB in SA. If the prevalence of PID in South Africa is indeed similar to what is observed in Australian/New Zealand and US registries, thus ranging between 2 850 and 45 723 PID patients, this means that a minimum of 142 to 2 286 PID patients will contract TB in their lifetime. There is thus a crucial need for increased PID awareness in SA and for patients with known or novel PID-causing mutations to be correctly identified – who would otherwise be labelled as non-compliant with TB treatment or as repeatedly exposed in their environment<sup>225</sup>.

### **1.5.1. Mendelian susceptibility to mycobacterial disease**

MSMD is a rare primary immunodeficiency affecting < 1 in 1 000 000 individuals worldwide, causing increased susceptibility to weakly virulent mycobacterial infections such as *M.bovis* BCG, systemic non-typhoidal salmonellosis and nontuberculous mycobacteria (NTM)<sup>175,226,227</sup>. Some individuals are also susceptible to more virulent infecting agents such as *M.tb*<sup>196,228,229</sup>, possibly implying a monogenic predisposition to TB infection<sup>230–232</sup>. MSMD patients are not prone to other severe infections and are usually otherwise healthy<sup>233,234</sup>.

Since 1996 seven autosomal genes and two X-linked genes carrying MSMD-causing mutations have been identified. The autosomal genes include interferon-gamma receptor 1 (*IFNGR1*)<sup>235–238</sup> and interferon-gamma receptor 2 (*IFNGR2*)<sup>239</sup>, which encode the two chains of the IFN- $\gamma$  receptor known as IFN $\gamma$ R1 and IFN $\gamma$ R2 respectively; interleukin 12B (*IL12B*)<sup>240</sup>, encoding the p40 subunits of IL12 and -23; interleukin 12 receptor beta-1 (*IL12RB1*)<sup>229,241–245</sup> which encodes the  $\beta$ 1 chain shared by the receptors for IL12 and IL23; signal transducer and activator of transcription 1 (*STAT1*)<sup>246,247</sup>; interferon regulatory factor 8 (*IRF8*)<sup>248</sup>; and interferon-stimulated gene 15 (*ISG15*)<sup>249,250</sup>. The two X-linked genes are nuclear factor-kappa-B essential modulator/

inhibitor of nuclear factor kappa-B kinase subunit gamma (*NEMO/IKBKG*)<sup>251,252</sup> and cytochrome b beta (*CYBB*)<sup>253</sup>. Defects in these genes are associated with impaired IFN- $\gamma$ -mediated immunity, indicating its importance in host immunity against mycobacterial infections. Based on either dominant or recessive forms of inheritance, partial or complete deficiencies and the expression of non-functional molecules or the complete loss of expression<sup>226,251,254</sup>, 18 different genetic aetiologies have been characterized<sup>17,209</sup>. Although not unanimously agreed upon, mutations in tyrosine kinase 2 (*TYK2*) have also been associated with MSMD<sup>255–257</sup>. The seven autosomal genes can be characterized into two main groups according to their different effects in the host, with group one (*IFNGR1*, *IFNGR2*, *STAT1*<sup>258</sup>) altering cellular responses to IFN- $\gamma$  produced by T lymphocytes and NK cells, and group two (*IL12B*, *IL12RB1*, *ISG15*, *IRF8*<sup>209,243</sup>) impairing this IFN- $\gamma$  production. Diseases associated with mutations in these genes are shown in Table 1.2.

Interestingly, several MSMD-causing genes have been associated with TB susceptibility. IL12R $\beta$ 1 deficiency was established in several children with TB, usually where a family history of infections with *Salmonella* species or weakly virulent mycobacteria was involved. Two girls who suffered from abdominal and disseminated TB respectively both had siblings with histories of non-typhoid salmonellosis<sup>259,260</sup>, while another suffered from disseminated TB with no applicable family or personal history<sup>261</sup>. *M.tb* infection was observed in a child with partial autosomal recessive (AR) IFNGR1 deficiency<sup>262</sup>, and informally reported in patients with IL12B deficiency<sup>263</sup>. A heterozygous *STAT1* mutation was identified in a TB patient with two deceased relatives reported as having suffered from TB<sup>264</sup>. *M.tb* infection has also been documented in a patient with X-linked recessive MSMD<sup>265</sup>.

Another study investigated the genetic causes and clinical features of MSMD patients from Japan, and described six MSMD patients with mutations in *IFNGR1* and one with a *NEMO* mutation<sup>266</sup>. These studies provide significant evidence of Mendelian predisposition to TB.

**Table 1.2: Mutations in genes associated with different types of MSMDs.**

Disease	Genetic defect/ presumed pathogenesis	Inheritance	Affected cells	Affected function	Associated features
IL12p40 deficiency	Mutation in <i>IL12B</i> : subunit p40 of IL12/IL23	AR	M	IFN- $\gamma$ secretion	Susceptibility to <i>Mycobacteria</i> and <i>Salmonella</i>
IL12 and IL23 receptor b1 chain deficiency	Mutation in <i>IL12RB1</i> : IL12 and IL23 receptor b1 chain	AR	L + NK	IFN- $\gamma$ secretion	Susceptibility to <i>Mycobacteria</i> and <i>Salmonella</i>
IFN- $\gamma$ receptor 1 deficiency	Mutation in <i>IFNGR1</i> : IFN- $\gamma$ R ligand binding chain	AR, AD	M + L	IFN- $\gamma$ binding and signaling	Susceptibility to <i>Mycobacteria</i> and <i>Salmonella</i>
IFN- $\gamma$ receptor 2 deficiency	Mutation in <i>IFNGR2</i> : IFN- $\gamma$ R accessory chain	AR	M + L	IFN- $\gamma$ signaling	Susceptibility to <i>Mycobacteria</i> and <i>Salmonella</i>
STAT1 deficiency (AD form)	Mutation in <i>STAT1</i> (loss of function)	AD	M + L	IFN- $\gamma$ signaling	Susceptibility to <i>Mycobacteria</i>
ISG15	Mutation in <i>ISG15</i> ; an IFN a/b- inducible, ubiquitin-like intracellular protein	AR	M + N + L	IFN- $\gamma$ secretion	Susceptibility to <i>Mycobacteria</i>
IRF8- deficiency (AD form)	Mutation in <i>IRF8</i> : IL12 production by CD1c + MDC	AD	CD1c + MDC	Differentiation of CD1c + MDC subgroup	Susceptibility to <i>Mycobacteria</i>
EDA-ID, X- linked (NEMO deficiency)	Mutations of <i>NEMO/IKBKG</i> , a modulator of NF- $\kappa$ B activation	XL	M+L	NF- $\kappa$ B signaling pathway	Various infections (bacteria, <i>Mycobacteria</i> , viruses, and fungi) Colitis EDA (not in all patients) Hypogammaglobine mia to specific antibody polysaccharides deficiency
Macrophage gp91 phox deficiency	Mutation in <i>CYBB</i> : electron transport protein (gp91phox)	XL	Mf only	Killing (faulty O <sub>2</sub> production)	Isolated susceptibility to <i>Mycobacteria</i>

AD = autosomal dominant inheritance; AR = autosomal recessive inheritance; EDA-ID = anhidrotic ectodermal dysplasia with immunodeficiency; IFN- $\gamma$  = interferon gamma; IFN- $\gamma$ R = interferon gamma receptor; IL =

interleukin; IL12B = interleukin 12 subunit beta; IL12p40 = interleukin 12 subunit p40; IL12RB1 = interleukin 12 receptor beta 1; IRF8 = interferon regulatory factor 8; ISG15 = interferon-stimulated gene 15; L = lymphocytes; M = monocytes-macrophages; MDC = myeloid dendritic cells; Mf = macrophages; N = neutrophils; NEMO = nuclear factor-kappa B essential modulator; CD = Cluster of differentiation; NK = natural killer cells; STAT1 = signal transducer and activator of transcription 1; XL = X-linked inheritance<sup>209</sup>.

The existence of an extensive group of children suffering from TB, with Mendelian predisposition to this disease as the underlying cause, is therefore probable<sup>267</sup>. In almost half of all MSMD patients, however, no genetic etiology has been described to date, thus excluding the involvement of the mentioned genes in disease outcome<sup>266</sup>. This highlights the crucial need to identify yet undiscovered novel candidate genes involved in MSMD and thus TB susceptibility. Siblings of another MSMD patient also presented with the same variant as the patient, but had severe TB as the only phenotype. It could thus be hypothesized that a monogenic predisposition to the disease may be present as the underlying cause in a large group of children with severe forms of TB<sup>236</sup>.

The most severe form of extra-pulmonary TB is tuberculous meningitis (TBM), with almost 80% of the affected population consisting of children less than five years old<sup>268</sup>. Approximately 2-5% of all active *M.tb* cases are due to TBM<sup>269</sup>. Tubercles form in the subpial (innermost layer of the meninges) or subependymal (layer below the ependyma in the lateral brain ventricles) areas due to TB dissemination. TBM is a result of these tubercles bursting into the subarachnoid space<sup>270</sup>. TBM is associated with very high mortality rates, rapid progression and usually permanent neurological damage<sup>271</sup>. This rare form of TB is more than three times more prevalent in children under the age of six years compared to adults<sup>272</sup>, similar to findings in MSMD. Although polymorphisms in *TLR2*<sup>273</sup>, toll-interleukin receptor domain containing adaptor protein (*TIRAP*)<sup>274</sup> and leukotriene A4 hydrolase (*LTA4H*)<sup>275</sup> have previously been associated with TBM susceptibility, relatively little is known about the host genetic factors associated with this disease. It is thus worth investigating TBM infected children for variants present in known MSMD genes.

## 1.6. PRESENT STUDY

One of the primary features of MSMD is increased susceptibility to TB. MSMD thus offers an opportunity to investigate TB susceptibility genes in the context of a Mendelian disorder. To date, mutations in 10 genes have been associated with phenotypically different forms of MSMD. Identifying MSMD-causing mutations enables exact diagnosis of these MSMDs and provides guidelines to successfully treat it. It also emphasizes the main molecular immune mechanisms involved in a host's immune reaction to TB<sup>168</sup>, since several of these MSMD-causing genes have also been associated with increased TB susceptibility (*IL12RB1*<sup>259–261</sup>, *IFNGR1*<sup>262</sup>, *IL12B*<sup>263</sup>, *STAT1*<sup>264</sup> and *NEMO*<sup>265</sup>). The existence of an extensive group of children suffering from TB, with Mendelian predisposition to this disease as the underlying cause, is thus very probable<sup>267</sup>. Since as much as 2286 South African PID patients are expected to contract TB in their lifetime, it is crucial that there is increased PID awareness in SA and that patients with known or novel PID-causing mutations are correctly identified, who would otherwise be labelled as non-compliant with TB treatment or as repeatedly exposed in their environment.

The present study hypothesizes that by identifying novel MSMD-causing mutations, one would effectively also be identifying novel candidate TB susceptibility genes that can be evaluated in TB cases and controls. As part of a larger study, which aims to identify disease-causing mutations in PIDs that present with increased susceptibility to TB, the present study aims to identify MSMD-causing mutations using whole exome sequencing (WES) and to determine whether the genes in which these mutations are found are associated with increased susceptibility to PTB. The objectives of this study are (1) to use WES to identify disease-causing mutations in MSMD patients; (2) to perform functional studies of the mutations identified to verify their causality in MSMD; (3) because it has been speculated that a significant number of children presenting with severe TB may have a monogenic predisposition to the disease, to screen the 10 MSMD genes (*IFGR1*, *IFGR2*, *IL12B*, *IL12RB1*, *STAT1*, *ISG15*, *IRF8*, *NEMO/IKBKG*, *CYBB* and *TYK2*) as well as any genes identified in aims 1 and 2, of 10 TBM patients for any variations using WES; and (4) to perform case-control association studies using variants in the gene/genes identified in our MSMD patients in a cohort of TB cases and control individuals.

## CHAPTER 2: MATERIALS AND METHODS

INDEX	PAGE
2.1. STUDY PARTICIPANTS	34
2.1.1. MSMD patients	34
2.1.1.1. <i>Case descriptions</i>	35
2.1.2. TBM cohort	36
2.1.3. PTB cases and controls	37
2.2. WHOLE EXOME SEQUENCING	38
2.2.1. Exome sequencing	39
2.2.2. Data analysis and bioinformatics	39
2.2.3. Polymerase chain reaction (PCR)	43
2.2.4. Gel electrophoresis	44
2.2.5. Gel clean-up	44
2.2.6. Sanger sequencing	45
2.3. VARIANT FREQUENCY DETECTION IN CONTROLS	45
2.3.1. Allele-specific restriction enzyme digest	45
2.3.2. Database searches	46
2.4. <i>IN SILICO</i> VARIANT PREDICTIONS	47
2.4.1. HOPE	47
2.4.2. MutationTaster2	48
2.4.3. PolyPhen-2	50
2.4.4. SIFT	52
2.5. FUNCTIONAL STUDIES	54
2.5.1. Generation of mutant plasmid constructs	55
2.5.1.1. <i>Bacterial plasmids</i>	55
2.5.1.2. <i>Site-directed mutagenesis</i>	55
2.5.2. Bacterial strain and cell line	56
2.5.2.1. <i>Bacterial Strain</i>	56
2.5.2.2. <i>Cell line</i>	56
2.5.2.3. <i>Bacterial plasmid isolation</i>	57

2.5.3.	Generation of <i>E.coli</i> DH5 $\alpha$ competent cells	57
2.5.4.	Culturing of the HEK293T Cell Line from frozen stocks	57
2.5.4.1.	<i>Thawing the cells</i>	57
2.5.4.2.	<i>Removing DMSO from stocks and culturing cells</i>	58
2.5.4.3.	<i>Splitting of cell cultures</i>	58
2.5.5.	Bacterial plasmid transformation	58
2.5.6.	Plasmid transfection into human cells	59
2.5.7.	Bafilomycin treatment of cells	59
2.5.8.	Protein isolation via cell lysis	60
2.5.9.	Bradford protein concentration determination	60
2.5.10.	Western blot	61
2.5.10.1.	<i>Sodium dodecyl sulphate polyacrylamide gel electrophoresis</i>	61
2.5.10.2.	<i>Membrane blocking</i>	61
2.5.10.3.	<i>Addition of primary antibody</i>	61
2.5.10.4.	<i>Addition of secondary antibody</i>	61
2.5.10.5.	<i>Chemiluminescent visualization of membrane proteins</i>	62
2.6.	FLOW CYTOMETRY	62
2.6.1.	Peripheral blood mononuclear cell isolation and growth	62
2.6.2.	Absolute cell counts	64
2.7.	CASE-CONTROL ASSOCIATION STUDIES	65
2.7.1.	Genotyping	65
2.7.2.	Data analysis	65

## 2.1. STUDY PARTICIPANTS

### 2.1.1. MSMD patients

Individuals suffering from MSMD were recruited via the Immunology Unit of the National Health Laboratory Service (NHLS) at Tygerberg hospital. Patients were clinically assessed by an immunologist and a diagnosis was made based on internationally recognized PID warning signs (Appendix I). Patients were examined for any signs of dysmorphology, and a PID category diagnosis was made according to the IUIS criteria of 2011<sup>276</sup>. MSMD symptoms include atypical recurrence of mycobacterial infection, disseminated BCG infection and systemic non-typhoidal salmonellosis. All patients were treated according to their specific symptoms.

Each participant gave written informed consent. Children under the age of 18 years gave informed assent, and informed written consent was provided by their parents/legal guardians. WES provided large amounts of information which, in some cases, included information about other disease states patients were not aware of. This was explained to participants prior to the study by a clinical geneticist, and each received counselling according to their unique circumstances from a genetic counsellor. Table 2.1 provides information of the three MSMD cases recruited for this study as well as their healthy parents. This study was approved by the Health Research Ethics Committee of Stellenbosch University (N13/05/075).

**Table 2.1: List of MSMD patients and controls sent for whole exome sequencing.**

Patient ID	Sex	Date of birth	Clinical diagnosis	Ethnicity
PID_012	F	19/02/1994	MSMD	SAC
PID_043	F	Unknown	Healthy mother of PID_012	SAC
PID_040	F	21/01/2008	MSMD	White
PID_041	M	Unknown	Healthy father of PID_040	White
PID_042	F	Unknown	Healthy mother of PID_040	White
PID_060	M	10/12/2005	MSMD	Xhosa
PID_061	F	05/01/1983	Healthy mother of PID_060	Xhosa
PID_062	M	24/07/1968	Healthy father of PID_060	Xhosa

ID = Identity; F = Female; M = Male; MSMD = Mendelian susceptibility to mycobacterial disease; SAC = South African Colored.

### **2.1.1.1. Case descriptions**

Patient PID\_012 is a 22 year old female from the South African Colored (SAC) population. She is one of four children of which two have a history of TB treatment. There have been multiple deaths in the family due to TB, all apparently from a single episode in late adulthood not related to recurrences. PID\_012 was first diagnosed with TB at the age of three years and suffered five subsequent episodes, including one *M.avium* infection. All but one of these episodes were caused by drug-sensitive strains, and the patient was placed on MDR TB treatment for the drug-resistant episode. The Laboratory of Human Genetics of Infectious Diseases at the Necker Medical School in Paris assisted with immunological workup. They reported normal full blood count, IgE levels, complement pathway activation and neutrophil burst responses, with elevated IgG levels. Lymphocyte subsets persistently showed significant reduction of CD3, CD4, CD8, CD19 and NK cells. Lymphocyte stimulations were consistently low for Phytohemagglutinin (PHA), Protein A and *Candida*. Although malfunctioning of the IL12/23-IFN- $\gamma$  pathway was expected given the patient's symptoms combined with good social circumstances and treatment adherence, BCG stimulation showed normal IL12p40 and IFN- $\gamma$  production. The patient did not present with any other infections.

Patient PID\_040 is an 8 year old white South African female initially diagnosed with humoral immunodeficiency after presenting with a BCG abscess on the upper leg at the age of two years. She tested negative for the Herpes simplex virus. She presented with severe hypogammaglobulinaemia with decreased Ig levels (IgA<0.06 g/L, IgM = 0.15 g/L, IgG = 0.53 g/L), increased lymphocyte subsets and normal lymphocyte proliferation to mitogens and recall antigens. She received intravenous Ig (IVIg) replacement therapy. T cell proliferation to mitogens and recall antigens as well as IFN- $\gamma$  production by T cells were normal. Up-regulation of CD69 on NK cells after IL2 stimulation was slightly decreased. CD40L was detected as present. CD27 expression appeared decreased as gated on CD19<sup>+</sup> B cells. The CD27<sup>+</sup>IgD<sup>+</sup> cell population was decreased at 2.96%. T cell receptor excision circles (TRECs) and kappa-deleting recombination excision circles (KRECs) were clearly visible. Class-switched memory B-cells (CD27<sup>+</sup>IgD<sup>-</sup>) remained persistently low at 0.12%. Memory T cells were low in relation to naïve T cells. Reduced levels of  $\gamma/\delta$  T cells were observed – cells involved in the innate immune reaction against mycobacteria. The patient developed BCG meningitis at the age of three and

was treated in the acute phase with Isoniazid (INH), Rifampicin (RIF), Ethionamide (ETA) and Decadron, and thereafter with only INH and RIF. At the age of four, diffuse granulomas were identified in her brain. INH, RIF, ETA and Levofloxacin was prescribed for a year, after which the Levofloxacin was discontinued. In 2014 she presented with acute loss of consciousness and raised intracranial pressure. *M.bovis* BCG genotypically sensitive to INH and RIF was subsequently cultured from the patient. She was started on a very aggressive 18 month treatment regimen of Levofloxacin, Terizidone, Amikacin IV, Linezolid, RIF, INH, Para-aminosalicylic acid (PAS) and IVIG replacement therapy. Genetic testing identified no mutations in *STAT1*, *LRBA*, *IL12RB1*, *IL12B* or *IFNGR1*.

Patient PID\_060 is an 11 year old Xhosa boy first diagnosed with PTB at the age of two. He suffered two subsequent episodes of confirmed pulmonary *M.tb* infection. No default from treatment was documented. Upon admission, abnormally high levels of  $\gamma/\delta$  T cells and decreased T cell proliferation against anti-CD3, which normalized upon IL2 addition, was detected. After TB treatment the patient consistently tested negative for TB (purified protein derivative [PPD] as well as TB T-spot test) i.e. achieved successful cure. Although this still needs to be confirmed on a subsequent sample, it implies decreased IL2 production by T cells and the treating physician postulated that the patient has decreased lymphocyte IFN- $\gamma$  (with or without IL2) production.

### **2.1.2. TBM cohort**

Samples from 10 TBM patients and 10 ethnically matched controls were previously obtained from the pediatric neurology ward at Tygerberg Children's hospital (Stellenbosch University) as part of a separate study (Ethics approval number N12/07/041). All TBM patients and controls self-identified as part of the South African Colored (SAC) population and were unrelated. Controls were over the age of 18 years, with an average age of about 50 years. Control samples were selected for older age as the risk of TBM development decreases as age increases. The TBM patients were much younger than controls (average age of five years) due to the age of onset of the disease. The age, sex and ancestry proportions of the 10 cases and 10 controls are given in Table 2.2. Patients were classified as "proven TBM cases" when *M.tb* was isolated from cerebrospinal fluid (CSF), or if post-mortem Ziehl-Neelsen staining was positive in brain tissue, with a positive test result on granuloma formation. Clinical signs of meningitis in the presence of

characteristic CSF findings, additional clinical findings or imaging criteria, with exclusion of the most likely alternative diagnoses, indicated “probable TBM”<sup>277</sup>. Patients were HIV negative and did not present with any other acquired cause of immune-deficiency. Upon admission, all patients were submitted to Computerized Tomography (CT) imaging. The size and location of all infarcts were noted, and an Evans ratio<sup>278</sup> of >0.3 was used to indicate hydrocephalus. Short-term, intense TB treatment was given to all patients.

**Table 2.2: Demographic variable of the TBM cases and controls.**

	TBM Cases	Controls
Number of participants	10	10
Number of males	5 (0.50)	5 (0.50)
Age (mean ± SD)	5.1 ± 3.93	49.65 ± 12.79
African San [IQR]	0.27 [0.24-0.35]	0.31 [0.27-0.45]
African non-San [IQR]	0.42 [0.31-0.74]	0.20 [0.15-0.24]
European [IQR]	0.09 [1x10 <sup>-5</sup> -0.14]	0.24 [0.22-0.29]
South Asian [IQR]	0.08 [1x10 <sup>-5</sup> -0.14]	0.11 [0.08-0.15]

IQR = Interquartile range; SD = Standard deviation; TBM = Tuberculosis meningitis.

### 2.1.3. PTB cases and controls

Three hundred and eighty-two (382) smear and/or culture positive PTB patients and 389 healthy controls were recruited from two metropolitan areas of Cape Town selected due to the high TB incidence and low HIV prevalence at the time of sampling<sup>279</sup>. PTB cases and controls were over the age of 18 years. Controls had never had a case of active TB. However, this does not exclude the possibility of latent infection, since 80% of children older than 15 years living in this region presented with positive TSTs<sup>280</sup>, indicating significant exposure of control individuals to *M.tb*. All cases and controls self-identified as part of the SAC population, were HIV negative and unrelated. The age, sex and ancestry proportions of the cases and controls are given in Table 2.3. One-hundred of these controls were used to investigate the prevalence of any potentially disease-causing variants identified in the MSMD patients. Another healthy individual from the SAC population, to ethnically match PID\_012, was included for flow cytometric comparisons.

**Table 2.3: Demographic variable of the PTB cases and controls.**

	PTB Cases	Controls
Number of participants	382	389
Number of males	212 (0.55)	118 (0.30)
Age (mean $\pm$ SD)	36.32 $\pm$ 11.04	30.88 $\pm$ 13.10
African San [IQR]	0.31 [0.20-0.40]	0.25 [0.19-0.34]
African non-San [IQR]	0.24 [0.15-0.36]	0.27 [0.19-0.38]
European [IQR]	0.17 [0.12-0.22]	0.12 [0.05-0.20]
South Asian [IQR]	0.15 [0.10-0.21]	0.25 [0.18-0.31]

IQR = Interquartile range; SD = Standard deviation; PTB = Pulmonary tuberculosis.

The present study aimed to identify novel TB susceptibility genes by finding MSMD-causing mutations through whole exome sequencing. We hypothesized that, since one of the primary features of MSMD is an increased susceptibility to mycobacterial disease, MSMD-causative genes are good candidates to investigate for increased TB susceptibility in the general population. It is also speculated that a number of children presenting with severe TB, such as TBM, may have a monogenic predisposition to this disease<sup>281</sup>. For this reason, 10 genes previously associated with MSMD, as well as the genes in which the putative MSMD-causing variants were identified in the present study, were screened for potential TBM-causing mutations in 10 TBM patients.

## 2.2. WHOLE EXOME SEQUENCING

DNA was extracted and purified from the MSMD patients and their parents as well as the TBM patient and control cohorts using the Nucleon BACC3 kit (GE Healthcare, UK) according to the manufacturer's instructions. In total, 28 DNA samples were whole exome sequenced at the Institute for Clinical Molecular Biology (ICMB) in Kiel, Germany. The TBM cases and controls were sequenced as part of another ongoing study (project numbers 95/072 and N09/07/185), and only variants situated in known MSMD-causing genes were investigated for potentially disease-causing variants in these cases.

### 2.2.1. Exome sequencing

Targeted enrichment and sequencing of the exomes, the first step in WES, was conducted with Illumina's TruSeq Exome Enrichment Kit with the unique feature of using biotinylated probes, streptavidin beads and magnets to capture and extract only the exonic DNA regions from genomic DNA samples. This kit targets >20 000 genes with >200 000 exons based on the gene annotations from CCDS (<https://www.ncbi.nlm.nih.gov/CCDS/CcidsBrowse.cgi>), RefSeq (<http://www.ncbi.nlm.nih.gov/refseq/>) and Encode/Gencode (<https://genome.ucsc.edu/ENCODE;> <http://www.encodegenes.org/>) as well as 9 Mb of predicted microRNA targets, and results in a total target size of 62 Mb. Preference was given to this kit as it includes untranslated regions (UTRs) frequently implicated in TB susceptibility. The enrichment was carried out in batches of four samples which were sequenced together on one lane of the Illumina HiSeq 2000 sequencing platform. Paired-end sequencing was used, yielding an average of 120 million reads per sample, with an average coverage of the target regions of 60-80× after duplicate removal. Patient and control samples were loaded onto flow cells at random as to compensate for any dissimilarities between different runs.

### 2.2.2. Data analysis and bioinformatics

Exome sequencing was followed by bioinformatics analysis of the sequencing data obtained from the MSMD patients and the TBM cases. The pipeline used for both of these groups is shown in Figure 2.1. A summary of the quality control results, as well as the important graphs and plots generated by FastQC for the sequencing data of each patient, is shown in APPENDIX II. The software packages used in this section, together with the versions and important parameters used, is indicated in Appendix III.

The sequencing output in FASTQ format underwent quality control in FastQC<sup>282</sup>. FastQC is a simple but effective tool that provides a quick impression of the quality of raw sequence data produced by high throughput sequencing pipelines, such as the Illumina HiSeq. The main functions of FastQC are: (1) to import data from BAM, SAM or FastQ files; (2) to provide a quick overview of the quality of the sequence data to determine whether the sequencing of any regions were problematic; (3) to summarize data in easy assessable tables and graphs; (4) to

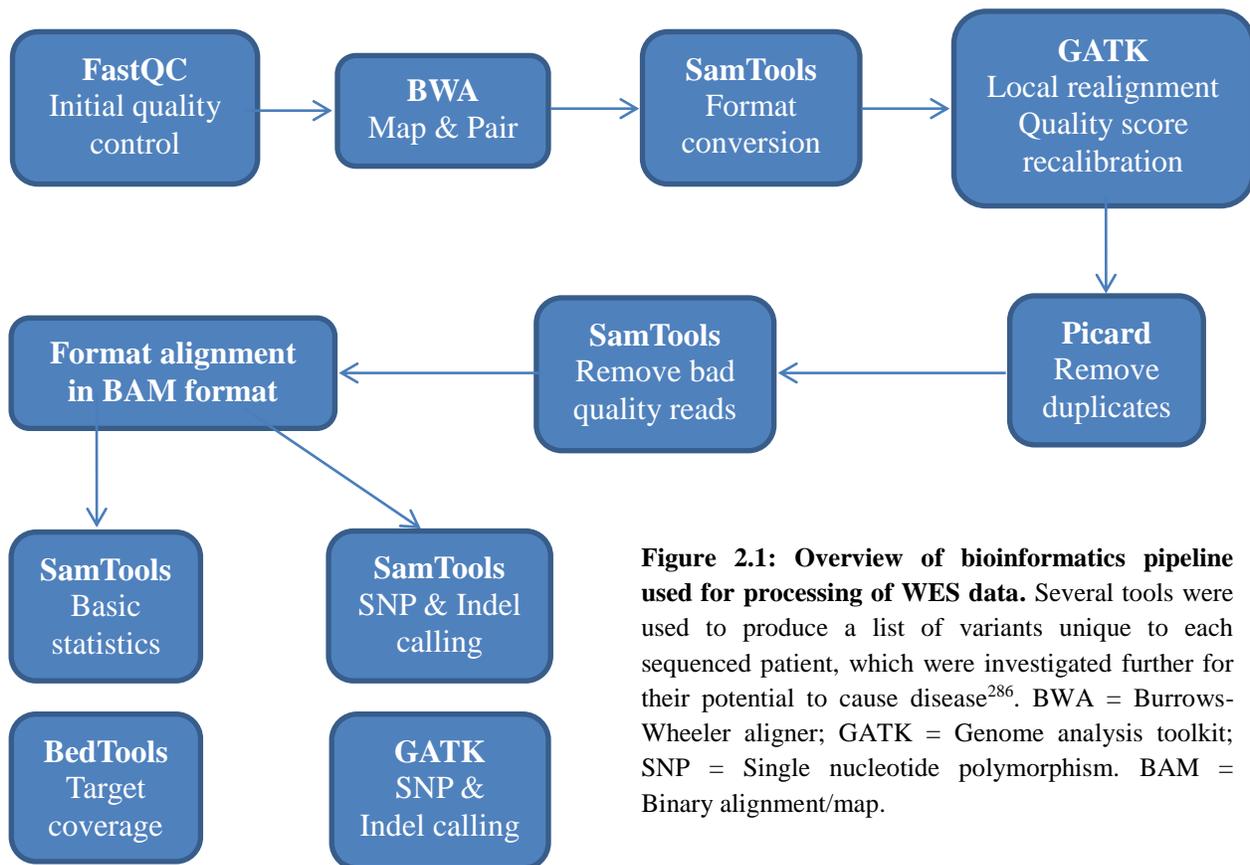
export results to an HTML-based permanent report; and (5) offline operation to allow automated generation of reports without running the interactive application.

The data that passed quality control was then mapped against the National Center for Biotechnology Information (NCBI) Human Reference Genome hg19, the most up-to-date version of the human genome at the time, using Burrows-Wheeler Aligner (BWA). BWA is a read alignment package based on backward search with Burrows-Wheeler Transform (BWT) that efficiently align short sequencing reads against a large reference sequence<sup>283</sup>. It allows gaps and mismatches, and supports base space reads, such as what is observed in sequence data from Illumina sequencing machines. BWA outputs alignment in Sequence Alignment/Map (SAM) format, and the SamTools software package was then used for format conversion. PCR duplicates, which can lead to false positive single nucleotide variant (SNV) calls through the overestimation of PCR errors, were removed using Picard<sup>284</sup>. Picard is a publically available set of tools for processing and analyzing next generation sequencing data.

Each step in the bioinformatics analysis generated statistics which were summarized in an overview for each sample showing the amount and quality of the raw sequencing data, percentage of uniquely mapped reads and duplicates, and calculation of the average coverage, as well as the percentage of the target covered  $\geq 1\times$ ,  $8\times$ ,  $10\times$  and  $20\times$ . SamTools<sup>285</sup> and GATK<sup>286,287</sup> were used in parallel to detect SNVs in order to reduce the false negative calls while achieving high specificity. The outcome of the alignment file was in VCF format, which was generated using the Mpileup function in SamTools. SamTools imports SAM formats, does sorting, merging and indexing, and can swiftly retrieve reads in any regions. The IndelRealigner function in GATK was used to perform local realignment of reads around indels, after which BaseRecalibrator was used to detect any errors in base quality scores. UnifiedGenotyper was used for variant calling, with a minimum phred-scaled confidence threshold of 10 at which variants were emitted and a target coverage threshold of 1000 for down-sampling to coverage.

The variant calls from both callers, SamTools and GATK, were pooled into a single set per sample. The exhaustive SNP/SNV annotation tool ANNOVAR was then used. The online version thereof (wANNOVAR) was accessed via a platform-independent web-interface

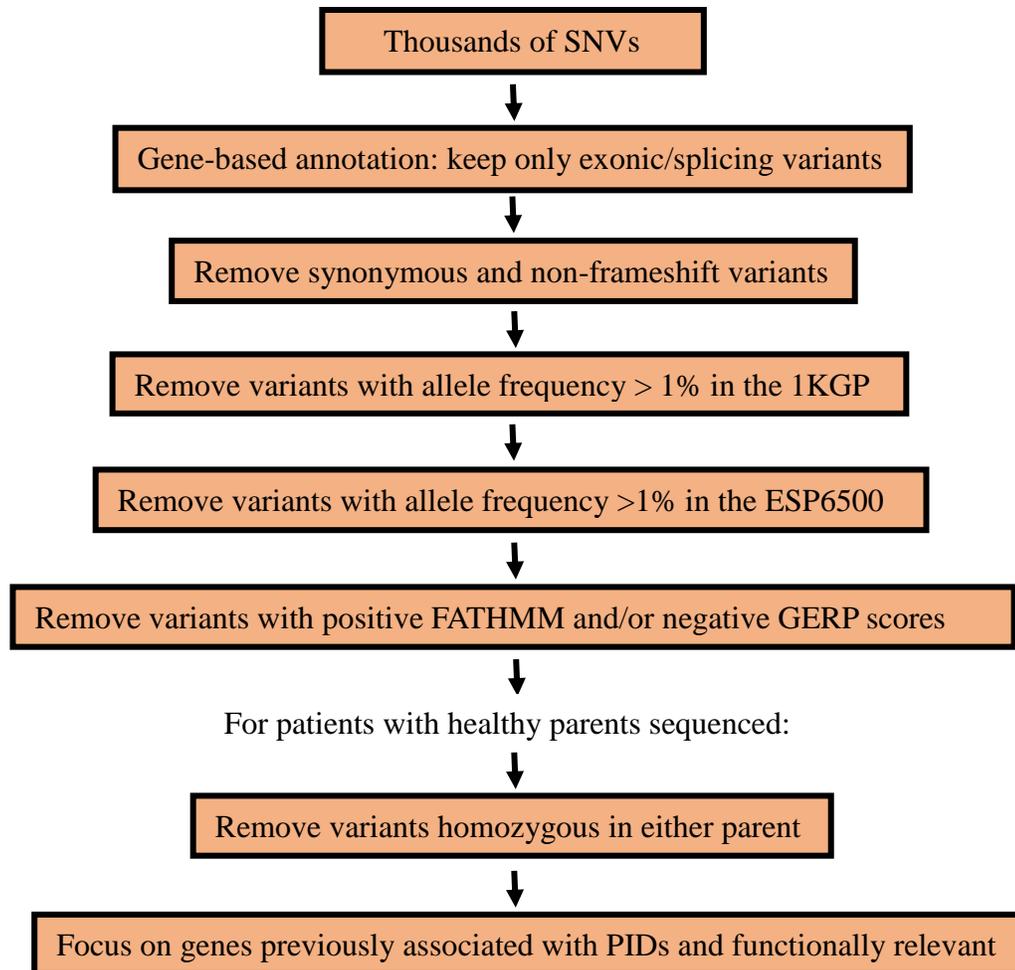
(<http://wannovar.usc.edu/index.php>). After importing the SNV data for each sample into ANNOVAR, additional information was summarized for each variant. Apart from annotation and classification of the SNVs (e.g. intronic/exonic missense/splice-site), a number of programs and databases were interrogated for each called position to determine whether a particular SNV was deleterious. An explanation of the predictive scores and values generated for the variants by ANNOVAR is given in Appendix IV. The combination of these different sources allowed for thorough assessment of the SNV data.



**Figure 2.1: Overview of bioinformatics pipeline used for processing of WES data.** Several tools were used to produce a list of variants unique to each sequenced patient, which were investigated further for their potential to cause disease<sup>286</sup>. BWA = Burrows-Wheeler aligner; GATK = Genome analysis toolkit; SNP = Single nucleotide polymorphism. BAM = Binary alignment/map.

The basic annotation options in ANNOVAR used in this project are (1) identifying and removing common SNVs unlikely to be disease-causing based on a minor allele frequency (MAF) threshold of >1% for the 1000 Genomes Project (1KGP) data, the Exome Sequencing Project 6500 (ESP6500) data, as well as our own healthy control data set; (2) restricting the SNVs to those causing amino acid changes in the protein: missense, nonsense, stopcodon-gain, stopcodon-loss and splice-site SNVs as well as frameshift and splice-site indels; (3) assessing the

impact on protein structure through the combined results of the prediction tools; and (4) the presence of a gene or SNV in Online Mendelian Inheritance in Man (OMIM) or Human Gene Mutation Database (HGMD), which shows known disease associations. This combination of annotation steps for the identification of potential novel candidate genes involved in rare Mendelian disorders are shown in Figure 2.2.



**Figure 2.2: Overview of filtering process used to determine potential disease-causing variants.** These filtering steps were used to predict which of the identified variants might be deleterious and disease-causing. SNVs = Single nucleotide variants; 1KGP = 1000 Genomes Project; ESP = Exome Sequencing Project; FATHMM = Functional analysis through hidden Markov models; GERP = Genome evolutionary rate profile; PIDs = Primary immunodeficiency disorders.

Synonymous and non-frameshift variants do not alter the amino acids they are situated in, and are thus highly unlikely to have any harmful effects on the proteins containing them. Variants

present in healthy controls, such as those present on the 1KGP and ESP6500 databases, at a MAF of larger than 1% are also unlikely to be rare, disease-causing mutations. Functional analysis through hidden Markov Models (FATHMM) determines species-specific weightings for predictions of functional effects of protein missense variants. It has been shown to outperform conventional prediction methods, such as SIFT, Polyphen and MutationTaster, and is currently the most trusted prediction tool. Positive FATHMM scores indicate tolerated variants unlikely to be damaging to the protein and thus cause disease, while negative scores indicate pathogenicity. Genomic Evolutionary Rate Profiling (GERP) scores are conservation scores obtained from the database for nonsynonymous SNPs functional predictions (dbNSFP). Scores of larger than 0 indicate conserved areas more likely to contain deleterious variants.

Further experimental work, such as variant verification using Sanger sequencing, frequency detection in the general population and functional investigation of the involvement of variants in disease causality, was only carried out using the variants identified in some of the MSMD patients. No TBM patients were investigated further.

### **2.2.3. Polymerase chain reaction (PCR)**

Oligonucleotide primers were designed to amplify only 100 – 600 nucleotides surrounding each identified variant. For each variant, a forward and reverse primer was designed using the NCBI GenBank database (<http://www.ncbi.nlm.nih.gov/>) and Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>). All primers (Table 2.4) were synthesized at the University of Cape Town (UCT).

In a final 50µL reaction, 20ng of genomic DNA was used along with 400nM of each of the forward and reverse primers. Five µL of 10× NH<sub>4</sub> Reaction Buffer, 2mM MgCl<sub>2</sub> solution, 3mM dNTP mix and 0.5 Units of BIOTAQ DNA polymerase were added. For the *TSPAN33 A111T* variant, 1.3µL dimethyl sulfoxide (DMSO) was also added. The volumes were filled with deionised water. The amplification reactions were performed in a Mastercycler® ep PCR system (Eppendorf, GER) at temperatures optimal for each fragment. The cycle parameters consisted of a denaturing step of three minutes at 95°C, followed by 35 cycles consisting of 30 second

intervals of 95°C, optimal annealing temperatures for 30 seconds (Table 2.4), and 72°C for 30 seconds, after which an extension step followed for one minute at 72°C.

**Table 2.4: Forward and reverse primers used for each specific variant.**

Patient ID	Gene	Variant	Forward Primer	Reverse Primer	Amplicon	Annealing Temp
PID_012	<i>TAP1</i>	p.I296M	cgcctcactgactggattct	aacacctctccctgcaagtg	275 bp's	63°C
	<i>JAK3</i>	p.R918P	ctgggcaagggcaacttt	tgagtgctttgaggatctg	117 bp's	60°C
	<i>TSPAN33</i>	p.A111T	cgtggtgggtgtcctcat	tgatgtttctcaggggaagg	544 bp's	63°C
PID_040	<i>MAP3K14</i>	p.V345M	agccctggaacctcacc	tgagattggcggataagaga	455 bp's	65°C
PID_060	<i>TAP1</i>	p.P67S	Cgccagtagggaggact	agtcggcgagaagtagcagt	159 bp's	62°C

ID = Identity; *TAP1* = Transporter associated with antigen processing 1; *JAK3* = Janus activating kinase 3; *TSPAN33* = Tetraspanin 33; *MAP3K14* = Mitogen activated protein kinase kinase kinase 14; bp's = base pairs; °C = Degrees Celsius.

#### 2.2.4. Gel electrophoresis

Gel electrophoresis was performed for each PCR product. One percent agarose gels (Appendix V) were prepared with SeaKem® LE agarose and SB buffer (Appendix V). The 20× SB buffer stock was dissolved 20× with deionized water before adding 1% agarose.

Two µL ethidium bromide (Appendix V) was added to every 50mL buffer/agarose mixture to enable visualization of the DNA fragments under ultra violet (UV) light. After the gels set, 10µL of each PCR product was added to individual wells. Ten µL of 100ng/µL KAPA Universal DNA Ladder (Kapa Biosystems (Pty) Ltd., RSA) was added to the first well of each gel. Gel-electrophoresis was conducted at 180V to separate the different sizes of amplified DNA fragments containing the variants of interest. The bands that formed on the gels were visualized under UV light. The Ladder served as a reference to determine the sizes of each PCR product, which were compare to the amplicon sizes in Table 2.4.

#### 2.2.5. Gel clean-up

The amplified DNA fragments needed to be purified from the agarose gels in order to be Sanger sequenced efficiently. DNA bands of the correct sizes were cut out of the gels under UV light using sterile blades. The Wizard® SV Gel and PCR clean-up System (Promega Corp., USA) was used to purify the DNA from these gel slices, following the manufacturers' instructions.

### **2.2.6. Sanger Sequencing**

Sanger sequencing was used to validate the WES results and verify whether the potentially disease-causing variants identified were true variants or sequencing artefacts. PCR fragments containing each of the variants identified through WES were bi-directionally sequenced using the BigDye® Terminator v3.1 Cycle Sequencing kit (Perkin-Elmer, Applied Biosystems Inc., Foster City, California, USA), followed by electrophoresis on an ABI 3130XL Genetic Analyzer (Perkin-Elmer, Applied Biosystems Inc., Foster City, California, USA). All automated DNA sequencing reactions were performed at the Core Sequencing Facility of the Department of Genetics at Stellenbosch University, Stellenbosch.

## **2.3. VARIANT FREQUENCY DETECTION IN CONTROLS**

The variants identified in PID\_012 was genotyped in SAC control individuals to determine the frequency thereof in the general population. Genotyping was used since no database containing the genome or exome sequencing data of healthy individuals from this population existed at the time this study was conducted. The exome data of Bantu-speaking as well as European and American controls were investigated for variants identified in PID\_060 and PID\_040, respectively. Variants were characterized as common polymorphism present in the general population if they presented with a minor allele frequency (MAF) of >1% in the controls. Only rare mutations with MAFs of <1% were investigated further. Polymorphisms are broadly defined as numerous forms of on single gene present either in an individual or among entire groups. Polymorphisms are common in a population group, and two or more equally acceptable alternatives of the DNA sequence exist<sup>288</sup>. Mutations occur when DNA sequences are irreversibly modified to vary from what is observed in other, healthy individuals<sup>289</sup>.

### **2.3.1. Allele-specific restriction enzyme digest**

Allele-specific restriction enzyme analysis (ASREA) was used to genotype 100 SAC control individuals for variants identified in patient PID\_012. Generally, 5µl of a PCR-amplified product was added to a reaction mix comprising 1-3 Units of the relevant enzyme, 1µl of the relevant restriction enzyme buffer and deionized water to a final volume of 10µl (Table 2.5). The mixture was subsequently incubated for 2-16 hours at optimal temperatures. Following digestion, 8µl of the digested sample was mixed with 1µl bromophenol blue loading dye (Appendix V) and loaded

onto 2% agarose gels and visualized on a long wave 3UV transilluminator (UVP, Inc. Upland, CA, U.S.A). Restriction enzymes that were used to genotype each variant were identified using NEBcutter v2.0 (<http://nc2.neb.com/NEBcutter2/>). The restriction enzyme cut sites as well as the expected band sizes are shown in Table 2.6.

**Table 2.5: Allele-specific restriction enzyme analysis of variants identified in PID\_012.**

Variant	Enzyme	Buffer	Optimal temperature
<i>TAP1 I296M</i>	<i>MscI</i>	Buff. 4	63°C
<i>TSPAN33 A111T</i>	<i>MscI</i>	Buff. 4	63°C

TAP1 = Transporter associated with antigen processing 1; TSPAN33 = Tetraspanin 33; Buff. = Buffer.

↑ indicates the cut site; °C = Degrees Celsius.

**Table 2.6: Expected outcome of restriction enzyme analysis of variants identified in PID\_012.**

Variant	Enzyme	Recognition site	Homozygous wild-type	Heterozygous	Homozygous mutant
<i>TAP1 I296M</i>	<i>MscI</i>	5'-TGG↓CCA-3' 3'-ACC↑GGT-5'	315 bp's	315 bp's 233 bp's 82 bp's	233 bp's 82 bp's
<i>TSPAN33 A111T</i>	<i>MscI</i>	5'-TGG↓CCA-3' 3'-ACC↑GGT-5'	343 bp's 239 bp's	343 bp's 239 bp's 192 bp's 151 bp's	239 bp's 192 bp's 151 bp's

TAP1 = Transporter associated with antigen processing 1; TSPAN33 = Tetraspanin 33; Bp's = base pairs.

### 2.3.2. Database searches

Databases containing the whole exome or whole genome sequence data of ethnically matched controls were interrogated for the variants identified in PID\_040 and PID\_060. The African Partnership for Chronic Disease Research (APCDR) granted us access to exome data of 100 healthy Bantu-speaking controls, which were investigated for the mutations found in PID\_060. Data from the 1KGP and the ESP6500 databases were investigated for variants identified in PID\_040. The Exome Aggregation Consortium (ExAC Browser Beta; <http://exac.broadinstitute.org/>), consisting of exome sequencing data from 60 706 unrelated individuals from different sequencing projects and ethnic backgrounds, was also interrogated to determine whether the variants have previously been identified in any population<sup>290</sup>. No representative panel of publically available healthy control exomes exists for the SAC population.

## 2.4. *IN SILICO* VARIANT PREDICTIONS

In the current study, Have (y)Our Protein Explained (HOPE; <http://www.cmbi.ru.nl/hope/>), MutationTaster2 (MT2; <http://www.mutationtaster.org/>), PolyPhen-2 (PP-2; <http://genetics.bwh.harvard.edu/pph2/>) and Sorting Tolerant From Intolerant (SIFT; <http://sift.jcvi.org/>) were used to predict the functional and structural consequences of each of the variants of interest. These specific tools were chosen since they are commonly used, and their efficacy has been proven<sup>291–294</sup>.

Several studies have used these four tools to predict the effects of amino acid exchanges on protein structure and function. Most have reported a very useful improvement in their understanding of biological pathways in which these proteins are situated, and thus a more effective design of downstream laboratory-based experiments, based on the results of these tools<sup>295–302</sup>.

### 2.4.1. HOPE

Have (Y)Our Protein Sequenced (HOPE) is a completely automatic program used for analyzing functional and structural effects of point mutations. It interrogates a large variety of sources to obtain information, such as the 3D coordinates of the protein, and sequence predictions or annotations. All this data is stored in a database and used to predict the mutation's effects on the function and 3D structure of a protein. Results are improved with regards to detail and reliability due to the use of 3D structures. A report containing text, figures and animations, which is easy to understand, is then generated<sup>303</sup>.

Six major aspects are explained in the report: contacts, structural domain, modifications, variants, conservation, and amino acid properties. “Contacts” explains interactions between the mutated residue and other atoms or molecules, and investigates interaction types (like disulfide bridges, hydrogen bonds and ionic interactions). “Structural domain” refers to parts of the protein with defined names, like regions, motifs, domains, etc. Structures that may influence post-translational processes without directly affecting protein structure are described by “modifications”, while all known splice variants, mutagenesis sites and polymorphisms are listed in the “variants” section. The frequency at which all other amino acids can occur at the mutated

residue relative to the amino acid most often observed, based on multiple sequence alignments, are indicated by “conservation”. All known properties of the wild-type and mutant amino acids are lastly listed<sup>304</sup>.

#### 2.4.2. MutationTaster2

MT2 rapidly and effectively evaluates the potential of DNA variations to cause disease by using reputable analysis tools and combining information from several biomedical databases. It takes into account splice-site changes, alterations that can influence mRNA expression, evolutionary conservation and loss of protein features. A Bayes classifier is used to predict the disease-causing potential of each variant<sup>305</sup> by calculating the probability of variants to be rare disease-causing mutations or innocuous polymorphisms based on characteristics of each variant and the results of all applied tests.

In order to make these predictions, the occurrence of all individual features for known disease-causing mutations or common polymorphisms are studied in a large set of >100 000 known disease mutations from HGMD Professional and >6 million harmless SNPs and Indel polymorphisms from the 1KGP<sup>306</sup>. Table 2.7 describes each prediction made by MT2.

**Table 2.7: Summary of MT2 output.**

Output	Description
Prediction	Disease causing: probably deleterious Disease causing automatic: known to be deleterious Polymorphism: probably harmless Polymorphism automatic: known to be harmless
Summary	List of most prominent features of analyzed alteration
Name of alteration	User-specified name
Alteration (phys. location)	Alteration on "physical" i.e. chromosomal level
HGNC symbol	The official HGNC symbol
Ensembl transcript ID	Ensembl transcript ID, starting with ENST.
UniProt peptide (SwissProt ID)	UniProt KB / SwissProt accession ID
Alteration type	Either base exchange, combination of insertion and deletion, insertion or deletion
Alteration region	Either 5'UTR, CDS, 3'UTR or intron
DNA changes	Alteration on nucleotide level. gDNA level (g.) is displayed always, cDNA level (cDNA.) for alterations located in exons, CDS level (c.) only for alterations residing in exon in coding sequence

AA changes	Any amino acid changes, displaying original versus new amino acid as well as position of substitution and a score for it.
Position(s) of altered AA	Lists positions of altered AA
Frameshift	Either yes or no
dbSNP/1KGP/ClinVar/HGMD	Any known polymorphism(s)/ disease variant found at position in question
Regulatory features	Regulatory features from Ensembl (histone modification sites, open chromatin or transcription factor binding sites)
phyloP/phastCons	Methods to determine the grade of conservation of nucleotide. <b>phastCons</b> : reflect probability that each nucleotide belongs to conserved element. 0 to 1, with 1 more probably conserved <b>phyloP</b> = -14 to +6. Conserved = positive scores; fast-evolving = negative scores.
Splice sites	Determines position of any splice site change relative to intron/exon borders: if loss/decrease of splice site occurs at intron/exon or exon/intron border: "real" splice site change
Kozak consensus sequence altered	Kozak consensus sequence (gccRccAUGG) upstream of start codon plays major role in translation initiation. R at -3 and G at +4 are highly conserved. Program checks whether previously strong consensus sequence has been weakened.
Conservation on AA level	Amino acid / nucleotide sequence homologues of 10 other species are aligned with corresponding human sequence. <b>All identical</b> : same amino acid in human and homologue amino acid sequence. <b>(Partly) conserved</b> : similar amino acids in human and homologous amino acid sequence. <b>Not conserved</b> : different amino acids in human and homologue amino acid sequence
Protein features	Protein features that are directly / indirectly affected by alteration. <b>Lost</b> : AA exchange is located in protein feature, which might get lost if whole exon is skipped due to splice site changes, or if protein is shortened because of premature termination codon
Length of protein	Checks if resulting protein will be elongated, truncated, or if NMD is likely to occur. <b>Prolonged</b> : original termination codon destroyed; translation stops later than normal <b>Slightly truncated</b> : <10% of wt protein length missing <b>Strongly truncated</b> : >10% of original protein length missing
AA sequence altered	Yes or no
Position(s) of altered AA	Shows position on amino acid level if alteration is in CDS
Position of stopcodon in wt/mu CDS	Position of last base of stop codon (TGA/TAA/TAG). Position 1 = A in start ATG codon
Position (AA) of stopcodon in wt/mu AA sequence	Position of stop asterisk (*) in amino acid sequence. Position 1 = first amino acid of protein
poly(A) signal	Analysis of polyadenylation signals
Conservation on nucleotide level	DNA sequences of different species are compared to human DNA sequence. <b>All identical</b> = same base(s) in human and species sequence <b>Not conserved</b> = different base(s) in human and species sequence <b>No alignment</b> = no local alignment around indicated position(s)
Position of start ATG in wt / mu cDNA	Position of A in start ATG, position 1 = first base of cDNA

Position of termination codon in wt/mu cDNA	Position of last bp of termination codon (TGA/TAA/TAG), position 1 = first bp of cDNA
Chromosome	Chromosome alteration is located on
Strand	1 = forward strand; -1 = reverse strand
Last intron/exon border	Last base of exon before last exon
Theoretical NMD border in CDS	Determines NMD border as last intron/exon junction minus 50bp and analyses if given premature termination codon occurs 5' to this border, thus eventually leading to NMD
Length of CDS	Length of coding sequence from A of initiation codon to last base of termination codon
cDNA position	Gives last wt base before alteration and first wt base after alteration in cDNA sequence context (relative to start of transcribed coding DNA reference sequence)
gDNA position	Gives last wt bp before alteration and first wt bp after alteration in gDNA sequence context (positions relative to start of gDNA reference sequence)
Chromosomal position	Gives last wt base before alteration and first wt base after alteration in chromosomal sequence context (position relative to start of chromosomal reference sequence)
gDNA and cDNA sequence snippet	Sequence surrounding alteration (20bp up- and downstream). Altered bases highlighted
wt and mu AA sequence	Complete AA sequences. * indicates STOP
Speed	Time MutationTaster needed for analysis & prediction

HGNC = HUGO Gene Nomenclature Committee; ID = Identity; UTR = Untranslated region; CDS = Coding sequence; gDNA = Genomic BDA; cDNA = Coding DNA; AA = Amino acid; dbSNP = Single nucleotide polymorphism database; 1KGP = 1000 Genomes Project; HGMD = Human Gene Mutation Database; R = Purine; NMD = Nonsense-mediated decay; wt = Wild-type; mu = Mutant; bp = base pairs<sup>306</sup>.

### 2.4.3. PolyPhen-2

PP-2 uses a collection of structure- and sequence-based automatically-selected predictive features to compare properties of a mutant allele to that of a wild-type allele. It (1) determines how probable the two alleles are to inhabit the specific site of interest; (2) how different the mutated protein is from the wild-type protein; and (3) if the altered allele originated at a hypermutable site<sup>307</sup>. These features are used to determine the functional importance of the alteration using a Bayes classifier<sup>291</sup>. PP-2 has been trained and tested using two pairs of datasets: HumDiv and HumVar. HumDiv consists of 6 321 differences between human proteins and closely related mammalian homologs assumed to be non-damaging, and 3 155 alleles in UniProt classified as influencing protein function and stability and causing Mendelian diseases. HumVar contains all mutations from UniProt classified as disease-causing and ~9 000 non-synonymous non-disease-causing SNPs. Table 2.8 describes the output generated by PP-2.

**Table 2.8: Summary of PP-2 output.**

Column No.	Column Name	Description
Original query (as copied from user input):		
1	o_acc	original protein identifier
2	o_pos	original substitution position in the protein sequence
3	o_aa1	original wild type (reference) amino acid
4	o_aa2	original mutant (substitution) amino acid
Annotated query:		
5	rsid	dbSNP reference SNP identifier if available
6	acc	UniProtKB accession if known protein, otherwise same as o_acc
7	pos	substitution position in UniProtKB protein sequence, otherwise same as o_pos
8	aa1	wild type amino acid in relation to UniProtKB sequence
9	aa2	mutant amino acid in relation to UniProtKB sequence
10	nt1	wild type (reference) allele nucleotide
11	nt2	mutant allele nucleotide
PolyPhen-2 prediction outcome:		
12	prediction	qualitative ternary classification appraised at 5%/10% (HumDiv) or 10%/20% (HumVar) FPR thresholds (“benign”, “possibly damaging”, “probably damaging”)
PolyPhen-1 prediction description (obsolete, please ignore):		
13	based_on	prediction basis
14	effect	predicted substitution effect on the protein structure or function
PolyPhen-2 classifier outcome and scores:		
15	pph2_class	probabilistic binary classifier outcome (“damaging” or “neutral”)
16	pph2_prob	classifier probability of the variation being damaging
17	pph2_FPR	classifier model False Positive Rate (1 - specificity) at the above probability
18	pph2_TPR	classifier model True Positive Rate (sensitivity) at the above probability
19	pph2_FDR	classifier model False Discovery Rate at the above probability
UniProtKB/Swiss-Prot derived protein sequence annotations:		
20	site	substitution SITE annotation
21	region	substitution REGION annotation
22	PHAT	PHAT matrix element for substitutions in the TRANSMEM region
Multiple sequence alignment scores:		
23	dScore	difference of PSIC scores for two amino acid variants (Score1-Score2)
24	Score1	PSIC score for wild type amino acid (aa1)
25	Score2	PSIC score for mutant amino acid (aa2)
26	MSAv	version of the multiple sequence alignment used in conservation scores calculations: 1 - pairwise BLAST HSP (obsolete), 2 - MAFFT-Leon-Cluspack (default), 3 - MultiZ CDS
27	Nobs	number of residues observed at substitution position in multiple alignment
Protein 3D structure features:		
28	Nstruct	initial number of BLAST hits to similar proteins with 3D structures in PDB
29	Nfilt	number of 3D BLAST hits after identity threshold filtering
30	PDB_id	PDB protein structure identifier
31	PDB_pos	position of substitution in PDB protein sequence
32	PDB_ch	PDB polypeptide chain identifier
33	ident	sequence identity between query sequence and aligned PDB sequence
34	length	PDB sequence alignment length

35	NormASA	normalized accessible surface area
36	SecStr	DSSP secondary structure assignment
37	MapReg	region of the phi-psi map (Ramachandran map) derived from the residue dihedral angles
38	dVol	change in residue side chain volume
39	dProp	change in solvent accessible surface propensity resulting from the substitution
40	B-fact	normalized B-factor (temperature factor) for the residue
41	H-bonds	number of hydrogen sidechain-sidechain and sidechain-mainchain bonds formed
42	AveNHet	number of residue contacts with heteroatoms, average per homologous PDB chain
43	MinDHet	closest residue contact with a heteroatom, Å
44	AveNInt	number of residue contacts with other chains, average per homologous PDB chain
45	MinDInt	closest residue contact with other chain, Å
46	AveNSit	number of residue contacts with critical sites, average per homologous PDB chain
47	MinDSit	closest residue contact with a critical site, Å
Nucleotide sequence context features:		
48	Transv	whether substitution is a transversion
49	CodPos	position of the substitution within a codon
50	CpG	whether substitution changes CpG context: 0 - non-CpG context retained, 1 - removes CpG site, 2 - creates new CpG site, 3 - CpG context retained
51	MinDJnc	substitution distance from closest exon / intron junction
Pfam protein family:		
52	PfamHit	Pfam identifier of the query protein
Substitution scores:		
53	IdPmax	maximum congruency of mutant amino acid to all sequences in multiple alignment
54	IdPSNP	maximum congruency of mutant amino acid to sequences in multiple alignment
55	IdQmin	query sequence identity with closest homologue deviating from wild type amino acid

No = Number; dbSNP = Single nucleotide polymorphism database; SNP = Single nucleotide polymorphism; PHAT = Predicted hydrophobic and transmembrane; PSIC = Position-specific independent counts; BLAST = Basic local alignment search tool; 3D = Three dimensional; PDB = Protein data bank; Pfam = Database of protein families<sup>291</sup>.

A comparative study found PP-2 to be superior to three other predictive tools<sup>293,308,309</sup>. HumVar is usually less accurate in its predictions due to the inclusion of mildly deleterious alleles in its dataset. However, HumVar is more appropriate for diagnosing Mendelian diseases, while HumDiv is used for dense mapping of regions identified via GWAS, to evaluate rare alleles and to analyze natural selection<sup>291</sup>. PP-2 is the most prominent prediction tool currently available that uses sequence as well as structural protein information to make predictions<sup>310</sup>.

#### 2.4.4. SIFT

SIFT predicts whether protein function is affected by amino acid variations based on the extent of amino acid conservation in closely related protein families<sup>311</sup>. It interrogates protein databases to create a dataset of protein sequences functionally related to a protein of interest. An alignment

built from these homologous sequences is used to determine the likelihood for each of the 20 amino acids to exist at a position of interest. The probability of the most common amino acid is used to normalize these probabilities, which is referred to as the SIFT scores<sup>312</sup>. Positions associated with low levels of conservation tolerate most alterations, while conserved positions do not. The confidence with which these predictions are made is indicated by conservation values of 0 (all amino acids can fill position) to  $\log_2 20$  (only 1 amino acid is observed)<sup>311</sup>. SIFT does not use protein structure to evaluate the consequence of an amino acid alteration. Table 2.9 describes the output generated by SIFT.

**Table 2.9: Summary of SIFT output.**

Column names	Description	Example
Coordinates	Chromosomal position of variant submitted	1,100624830,1,T/A
Codons	Codon that has been changed; + mRNA orientation	ATA-tTA
Transcript ID	According to Ensembl	ENST00000342895
Protein ID	According to Ensembl	ENSP00000344470
Substitution	Resulting AA substitution	I121L
Region	Region in which substitution is situated	EXON, CDS
dbSNP ID	If dbSNP has a variant overlapping at same position, rsID is displayed. However, alleles may differ	rs34920283:A
SNP Type	Type of SNP observed	Nonsynonymous
Prediction	Effect of the variant on protein is predicted to either be <b>tolerated</b> , <b>damaging</b> or <b>damaging with low confidence</b>	TOLERATED
Score	Ranges from 0 to 1. <b>Damaging</b> $\leq 0.05$ ; <b>tolerated</b> $> 0.05$	0.59
Median Info	<b>0 to 4.32</b> : ideally between 2.75 and 3.5. Measure diversity of sequences used for prediction. A warning will occur if $>3.25$ (prediction was based on closely related sequences)	3.06
# Seqs at position	Number of sequences with an AA at position of prediction. SIFT automatically chooses the sequence, but if substitution is at beginning/end of protein, there may be only a few sequences represented at that position.	28

ID = Identity; dbSNP = Single nucleotide polymorphism database; SNP = Single nucleotide polymorphism; Seqs = Sequences; AA = Amino acid; rs ID = Reference single nucleotide polymorphism cluster identifies; SIFT = Sorting intolerant from tolerant.

A comparative study between PP-2, MT2 and SIFT found that MT2 is the most accurate and sensitive of these tools, while HumVar of PP-2 is the most specific (Table 2.10). These four tools

are each clearly quite effective, as evident from Table 2.10. However, individually, any of them can lead to mislabeled variants and thus result in either false negative or false positive identification of potentially disease-causing variants. By using a combination of all four tools, we aimed to mitigate this.

**Table 2.10: Comparison between PP-2, SIFT and MT2.**

Tool	n	Sensitivity	Specificity	Accuracy
PP-2 HumVar	2200	0.789	0.887	0.838
PP-2 HumDiv	2200	0.858	0.821	0.840
SIFT	2200	0.827	0.858	0.843
MT2	2200	0.887	0.874	0.880

The sensitivity, specificity and accuracy of three of the prediction tools used in this study are compared. N = number of cases investigated in the comparison study by Schwartz and colleagues<sup>306</sup>. PP-2 = PolyPhen-2; SIFT = Sorting intolerant from tolerant; MT2 = MutationTaster2.

## 2.5. FUNCTIONAL STUDIES

In order to determine the functional involvement of the MAP3K14 V345M variant identified in PID\_040 in disease causality, several functional studies were performed. Ideally, one would do these functional analyses using peripheral blood mononuclear cells (PBMCs) isolated from the patient. However, for logistical reasons beyond our control, we were unable to obtain a follow-up blood sample from this patient for PBMC isolations. The effects of each mutation identified in PID\_040 were thus investigated by transfecting a mutant plasmid construct into an immortalized cell line. Transfection is a powerful tool for studying gene expression, and a very popular alternative in the absence of patient-derived cells. Cloned genes are transfected into cells and typically used for biochemical characterization, mutational analyses, investigation of the effects of gene expression on cell growth, investigation of gene regulatory elements, and to produce a specific protein<sup>313</sup>.

In the present study, plasmids containing either wild-type or mutant MAP3K14 were generated and transfected into HEK293T cells, along with plasmids containing wild-type IKK $\alpha$ , after which Western Blotting was used to determine if any differences existed with regards to the expression of several interesting proteins between the wild-type and mutant groups.

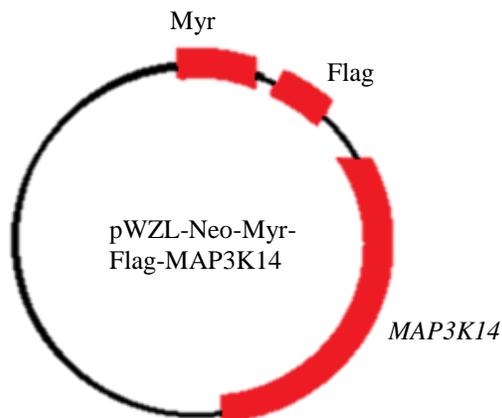
## 2.5.1. Generation of mutant plasmid constructs

### 2.5.1.1. Bacterial plasmids

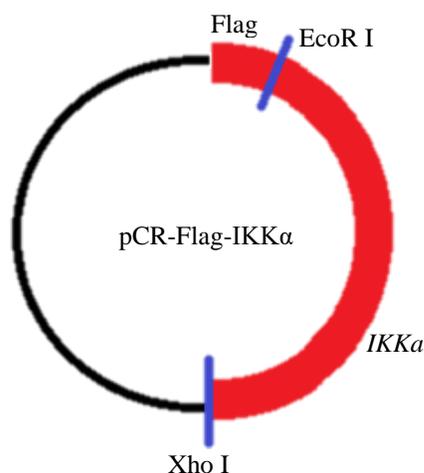
Bacterial plasmid constructs containing inserts encoding MAP3K14 (pWZL-Neo-Myr-Flag-MAP3K14) (Figure 2.3) and IKK $\alpha$  (pCR-Flag-IKK-alpha) (Figure 2.4) were purchased from the non-profit plasmid repository, Addgene (<https://www.addgene.org>).

### 2.5.1.2. Site-directed mutagenesis

The Q5<sup>®</sup> Site-Directed Mutagenesis Kit (New England BioLabs Inc., UK.) was used according to the manufacturer's instructions to create a pWZL-Neo-Myr-Flag-MAP3K14 mutant construct. The mutant construct, pWZL-Neo-Myr-Flag-MAP3K14-V345M, was created to mimic the mutation identified in patient PID\_040. This mutagenesis kit enables rapid, site-specific mutagenesis of double-stranded plasmid DNA. It uses the robust Q5 Hot Start High-Fidelity DNA Polymerase together with custom mutagenic primers to create insertions, deletions and substitutions in a wide variety of plasmids.



**Figure 2.3: Simple map of pWZL-Neo-Myr-Flag-MAP3K14 plasmid.** The MAP3K14 gene is introduced into a bacterial plasmid, preceded by the Tag/Fusion proteins Flag and Myr.



**Figure 2.4: Simple map of pCR-Flag-IKK-alpha plasmid.** The IKK $\alpha$  gene is introduced into a bacterial plasmid, preceded by the Tag/Fusion protein Flag and enclosed by two enzyme restriction sites EcoR1 and Xho1.

Briefly, a 25µL reaction was prepared per sample consisting of 12.5µL Q5 HotStart High-Fidelity 2× Master Mix, 200mM of both the forward and reverse primers (Table 2.11), and 20ng of template DNA. The volume was made up with nuclease-free water. The amplification reactions were performed in a Mastercycler® ep PCR system (Eppendorf, GER). The cycle parameters consisted of a denaturing step of 30 seconds at 98°C, followed by 25 cycles consisting of 10 second intervals of 98°C, 60°C for 30 seconds, and 72°C for six minutes. An extension step followed for two minutes at 72°C and the samples were kept at 4°C until further use. Following amplification, 1µL of the subsequent PCR product was added to 5µL 2× KLD Reaction Buffer, 1µL 10× KLD Enzyme Mix and 3µL nuclease-free water before being incubated for five minutes at room temperature. These plasmids were then Sanger sequenced to determine whether the mutagenesis worked, before being transformed into *E.coli* (section 2.5.5) and transfected into human cells (section 2.5.6).

**Table 2.11. Q5® Site-directed mutagenesis primers.**

5'-3' Sequence	Length	Start	Stop	Tm	GC%
F: AAGGCAGC <b><i><u>G</u></i></b> TGAGCTC	16	1150	1165	56.67	62.50
R: CAGAGCATGCACTAGGTAT	20	1149	1130	57.46	50.00

The mutated site is indicated in bold, italics and is underlined. F = Forward primer; R = Reverse primer; Tm = Melting temperature.

## 2.5.2. Bacterial strain and cell line

### 2.5.2.1. Bacterial Strain

To facilitate site-directed mutagenesis, the appropriate bacterial plasmids (section 2.5.1) were transformed into *E.coli* DH5α strain (Appendix V).

### 2.5.2.2. Cell line

The pWZL-Neo-Myr-Flag-MAP3K14, pWZL-Neo-Myr-Flag-MAP3K14-V345M and pCR-Flag-IKK-alpha plasmid constructs were transfected into human embryonic kidney 293 (HEK293T) cell lines. It would have been ideal to perform the functional work on patient-derived immune cells, but unfortunately this was not available. HEK293T cells were chosen due to their easily transfective nature and availability.

### **2.5.2.3. Bacterial plasmid isolation**

In order to isolate the plasmids of interest, one colony of *Escherichia coli* (*E.coli*) containing each of the plasmids of interest was picked from an appropriate agar slant and inoculated into 10mL LB media containing 25mg/mL Ampicillin. The inoculums were incubated overnight at 37°C in a YIH DER model LM-530 incubator (Scilab Instrument Co., Ltd., TW). This was followed by centrifugation for 10 minutes at 2400rpm in a Beckman model TJ-6 centrifuge (Beckman Coulter, Scotland, UK) and discarding of the supernatant. Pellets were re-suspended in residual supernatant, after which the PureYield™ Plasmid Miniprep System (Promega Corp., USA) was used according to the manufacturer's guidelines to purify the plasmid DNA.

### **2.5.3. Generation of *E.coli* DH5 $\alpha$ competent cells**

A scrape of an *E.coli* DH5 $\alpha$  frozen (-70°C) glycerol stock was inoculated into 10mL LB-media. The culture was incubated overnight at 37°C in a YIH DER model LM-530 shaking incubator (SCILAB Instrument CO. Ltd, Taipei, Taiwan) at 200rpm. Following incubation, a 1mL aliquot of this culture was inoculated into a 2L Erlenmeyer flask containing 200mL LB media (Appendix V). This culture was incubated at room temperature for 24 hours, while shaking at 200rpm, to mid-log phase (OD<sub>600nm</sub>=0.6) on a Labcon orbital shaker (Labcon Pty, Ltd, Maraisburg, RSA). At this point the culture was decanted into 4× 50mL polypropylene tubes, which were centrifuged at 3000rpm for 15 minutes at 4°C in a Multitex centrifuge (MSE instruments, England). The supernatant was removed and 8mL ice-cold CAP buffer (Appendix V) was used to re-suspend the pellets. The cells were re-pelleted by centrifugation at 3000rpm for 15 minutes at 4°C in a Multitex centrifuge. The supernatant was discarded and the pellets re-suspended in 4mL ice-cold CAP buffer. The suspended cells were subsequently transferred into 1.5mL microfuge tubes in 500 $\mu$ L aliquots and snap frozen by immersion in liquid nitrogen. The cells were stored at -70°C until they were needed.

### **2.5.4. Culturing of the HEK293T Cell Line from frozen stocks**

#### **2.5.4.1. Thawing the cells**

Frozen HEK293T cells were thawed rapidly by immersing the vial containing the frozen stock in a 37°C water bath (Mettmert®, Schwabach, Germany) for 10 minutes. Once the cells were thawed, the outside of the vial was sterilized with 70% ethanol.

#### ***2.5.4.2. Removing DMSO from stocks and culturing cells***

Frozen stocks contained DMSO, which had to be removed as soon as possible for maximum viability of cells upon plating. For this, 1mL growth media (Appendix V), pre-warmed to 37°C, was added to the thawed stock and mixed by gentle pipetting. The mixture was transferred to 15mL Greiner tubes (Greiner Bio-one, Frickenhausen, Germany) and another 5mL growth media was added. The cells were pelleted by centrifugation at 10000rpm for one minute using a Sorval® GLC-4 General Laboratory centrifuge (Separations Scientific, Johannesburg, South Africa), followed by removal of the supernatant. The pellet was re-suspended in 5mL growth media and the cells were again centrifuged at 10000rpm for one minute. The subsequent pellet was re-suspended in 10mL growth media and transferred to a T25 culture flask. The flask was gently swirled to distribute the cells evenly over the growth surface before being incubated at 37°C in a Farma-thermosteri-cycle 5% CO<sub>2</sub> humidified incubator (Farma International, Miami, Florida, U.S.A.).

#### ***2.5.4.3. Splitting of cell cultures***

Cell cultures were split every 2-4 days when they reached approximately 80%-90% confluency. Briefly, the growth media was removed from the flask and the cells were washed with sterile phosphate buffered saline (PBS) (Appendix V) containing no calcium or magnesium. To this, 2mL of 200mg/L trypsin/EDTA (Highveld Biological, Lyndhurst, South Africa) was added to facilitate the detachment of the cells from the growth surface of the flask. After three minutes, 5mL growth media was added and the cells were gently re-suspended. Cells from a single flask were transferred to four flasks each containing 10mL growth media.

#### **2.5.5. Bacterial plasmid transformation**

Prior to transformation, an aliquot of competent *E.coli* DH5 $\alpha$  (section 2.5.3) was thawed on ice for 20-30 minutes. Once thawed, 5 $\mu$ L plasmid preparation (section 2.5.2.3) was added. The mixture was incubated on ice for 20-30 minutes after which it was placed in a Lasec 102 circulating water bath (Lasec Laboratory and Scientific Company Pty Ltd, Cape Town, R.S.A) at 42°C for 45 seconds. The sample was removed from the water bath and left at room temperature for two minutes. One mL of LB media (Appendix V) was added to the mixture and the sample was incubated for one hour at 37°C, while shaking at 200rpm in a YIH DER model LM-530

shaking incubator. Next, 200µl of the sample was plated onto LB agar plates containing ampicillin (Appendix V). The remaining transformation reaction mixture was centrifuged at 13000rpm for two minutes in a Beckman Microfuge Lite, the supernatant discarded and the pellet re-suspended in 200µl LB media. This was also plated onto ampicillin LB agar plates. All plates were incubated, inverted, for 16 hours at 37°C in a model 329 stationary CO<sub>2</sub> incubator (Former Scientific, Marieta, Ohio, U.S.A.). All transformations were performed in triplicate.

#### **2.5.6. Plasmid transfection into human cells**

Forty-eight hours before transfecting the cells, approximately  $1-3 \times 10^4$  cells per well were plated in growth media in a six-well tissue culture plate (Appendix VI) and incubated at 37°C in a 5% Farma-thermosteri-cycle CO<sub>2</sub> humidified incubator (Farma, international, Miami, Florida, U.S.A.). Two days later, the cells were visualized under a Nikon TMS light microscope (Nikon, Tokyo, Japan) to determine the level of confluence. Cells were only transfected once they reached approximately 80% confluence. For each transfection reaction, a total of 2µg DNA (1µg purified pCR-Flag-IKK-alpha plasmid + 1µg purified pWZL-Neo-Myr-Flag-MAP3K14 plasmid / purified pWZL-Neo-Myr-Flag-MAP3K14-V345M plasmid) was diluted in 400µL Dulbecco's Modified Eagle's Medium (DMEM) in a 1.5mL Eppendorf tube. After mixing gently, 2µL PLUS™ Reagent (Invitrogen, USA) was added to each tube and incubated for 15 minutes at room temperature. Five µL Lipofectamine™ LTX Reagent (Invitrogen, USA) was added to each mixture and incubated at room temperature for a further 30 minutes. The DNA-Lipofectamine complexes were subsequently added to the ~80% confluent wells and mixed gently before being incubated for 16 hours at 37°C in a model 329 stationary CO<sub>2</sub> incubator. All transfections were performed in triplicate.

#### **2.5.7. Bafilomycin treatment of cells**

The transfected cells to be used for investigation of autophagy were divided into two groups. One received Bafilomycin A1 (BAFF) treatment while the other remained untreated. BAFF is a known inhibitor of the late phase of autophagy, and prevents maturation of autophagic vacuoles by inhibiting fusion between autophagosomes and lysosomes<sup>314</sup>.

A BAFF stock solution with a concentration of 1mM was made using DMSO and Bafilomycin A1. This was diluted to 1 $\mu$ M by adding 1mL PBS to 1 $\mu$ L of the stock. Each well in one 6-well plate seeded with HEK293T cells received 150 $\mu$ L of the 1 $\mu$ M BAFF stock and 1350 $\mu$ L growth media. These cells constituted the BAFF treatment group. Another plate received only 1500 $\mu$ L of growth media and constituted the control group not receiving any treatment. The plates were incubated for 16 hours at 37°C in a model 329 stationary CO<sub>2</sub> incubator. These experiments were performed in triplicate.

### **2.5.8. Protein isolation via cell lysis**

The HEK293T protein lysates that were used for Western blot analysis were prepared as follows: cells were trypsinized from the growth surface of each well of a six-well plate using 1mL 200mg/L trypsin/EDTA per well. The plates were incubated at 37°C for five minutes. Following this, 3mL growth media was added to each well. The 4mL media and cells were then removed from each well and added to the appropriately labelled 15mL polypropylene tubes. These were centrifuged at 15000rpm for five minutes and the pellets were re-suspended in 1mL PBS. This wash step was repeated two more times. The resulting pellets were dissolved in 80-200 $\mu$ L lysis buffer (Appendix V) depending on pellet size. The tubes were incubated on ice for 30 minutes after which they were centrifuged at 14000rpm for 45 minutes at 4°C. Lysates were stored at -80°C until further use.

### **2.5.9. Bradford protein concentration determination**

Bradford assays were performed to determine the protein concentrations of the extracted protein lysates (section 2.5.8). For preparation of a standard curve, 10 $\mu$ L serial diluted bovine serum albumin (BSA) ranging from 0 - 1000 $\mu$ g/ $\mu$ L was loaded into the wells of a 96-well luminometer plate, along with 1 $\mu$ L lysate samples, in duplicate. Two-hundred  $\mu$ L Bradford reagent was added to all standards and samples, and a Synergy HT luminometer (BioTek Instruments Inc., USA) running the KC4™ v 3.4 program (BioTek Instruments Inc., USA) was used to determine the protein concentration of each well at an absorbance of 595nm.

## **2.5.10. Western blotting**

### ***2.5.10.1. Sodium dodecyl sulphate polyacrylamide gel electrophoresis***

The lysates were electrophoresed in Mini-PROTEAN® TGX™ precast polyacrylamide gels (Bio-Rad Laboratories (Pty) Ltd., RSA) containing 1% Sodium dodecyl sulfate (SDS) ranging between 4% and 15%. First, a mixture of SDS loading dye (Appendix V) and 100µg whole cell lysate was prepared at a 1:2 ratio and incubated at 95°C for five minutes. The samples were then loaded onto the vertical gels and electrophoresed at 80V for 10 minutes, after which the voltage was increased to 120V for about 30-50 minutes, in 1× SDS-PAGE running buffer (Appendix V). Ten µL molecular weight marker (Spectra™ Multicolor Broad Range Protein Ladder, Thermo Scientific, USA) was co-electrophoresed with the protein products to use for size comparison.

### ***2.5.10.2. Membrane blocking***

Following SDS-PAGE, the proteins were transferred to a membrane using the iBlot® Dry Blotting system (Invitrogen, RSA). The membranes were removed from the transfer apparatus and washed in Tris-buffered saline, 0.1% Tween 20 (TBST) (Appendix V) for approximately two minutes, after which they were incubated at room temperature for two hours shaking on a Stuart® orbital shaker SSL1 (Barloworld Scientific Ltd., UK) in membrane blocking solution (Appendix V) to block all non-specific binding sites of the proteins.

### ***2.5.10.3. Addition of primary antibody***

The membranes were rinsed with TBST after the blocking step. Primary antibodies (Table 2.12) were diluted in membrane blocking solution and added to the membranes in a container to be incubated overnight shaking at 4°C on an Orbit 300 shaker (Labnet International Inc., USA).

### ***2.5.10.4. Addition of secondary antibody***

The membranes were washed in TBST, 5× 10 minutes, after which the membranes were placed in membrane blocking solution containing the appropriate horseradish peroxidase (HRP) conjugated secondary antibodies (Santa Cruz Biotechnology Inc., USA; Cell Signaling Technology Inc., USA) on a shaker for one hour at room temperature. Subsequently, the membranes were rinsed in TBST as described above. The concentrations at which the primary and secondary antibodies were used are indicated in Table 2.12.

### 2.5.10.5. Chemiluminescent visualization of membrane proteins

In a dark room, SuperSignal West Pico Luminol/Enhancer solution and SuperSignal West Pico Stable Peroxide solution, the two substrate components of the SuperSignal® West Pico Chemiluminescent Substrate kit (Thermo Scientific, USA), were mixed in a 1:1 ratio. After incubating the membranes for five minutes in this mixture, excess chemiluminescent reagent was removed and the membranes were placed in an autoradiography cassette and covered with a transparent plastic sheet. They were exposed to CL-Xposure™ autoradiography film (Thermo Scientific, USA) for 30 seconds to 30 minutes, depending on the strength of the signal. Following adequate exposure, the films were developed in a Hyperprocessor™ automatic autoradiography film processor (Amersham Pharmacia Biotech UK Ltd., UK) to visualize the protein bands.

**Table 2.12: Primary and secondary antibodies at their optimized concentrations used in Western blot assays.**

Primary AB	Dilution	Secondary AB	Dilution
NIK	1:1000	Donkey anti-rabbit	1:7500
IKK $\alpha$	1:1000	Donkey anti-rabbit	1:7500
Phospho-IKK $\alpha$	1:1000	Donkey anti-rabbit	1:7500
NF- $\kappa$ B p100/p52	1:750	Donkey anti-rabbit	1:7500
LC3-B	1:1000	Donkey anti-rabbit	1:7500
GAPDH	1:2000	Rabbit anti-mouse	1:5000

AB = Antibody; NIK = Nuclear factor-kappa B inducing kinase; IKK $\alpha$  = IkappaB kinase-alpha; NF- $\kappa$ B = Nuclear factor-kappa B; LC = Light chain; GAPDH = glyceraldehyde 3-phosphate dehydrogenase.

## 2.6. FLOW CYTOMETRY

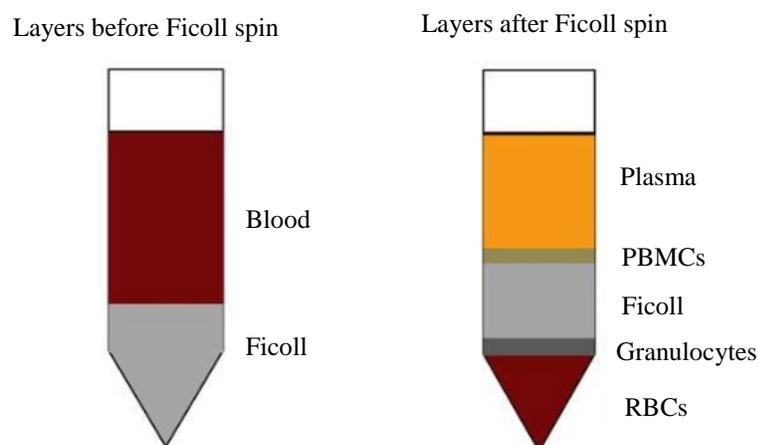
Since *TAP1* is known to play a role in CD8<sup>+</sup> T cell activation, flow cytometry was used to investigate the state of activation of CD8<sup>+</sup> T cells in patient PID\_012 compared to a control, using PBMCs isolated from the patient and an ethnically matched control.

### 2.6.1. Peripheral blood mononuclear cell isolation and growth

One Sodium Heparin/Ethylenediaminetetraacetic acid (NaHep/EDTA) vacutainer (Greiner Bio-One, Kremsmuenster, Austria) filled with 10mL of blood was required per patient and control to isolate a sufficient number of PBMCs. Each mL of healthy blood typically contains  $1 \times 10^6$  PBMCs, with that of PID patients depending on disease severity. Approximately  $5 \times 10^5$  PBMCs

were isolated per mL blood for PID\_012. Each vacutainer was slowly inverted five times and a 1:1 dilution of whole blood and 1× sterile PBS was prepared. For each blood/PBS mixture, 20µL Ficoll density media was pipetted into a 50mL conical tube. The blood/PBS dilutions were slowly pipetted over the density media without disturbing the Ficoll layers. The tubes were centrifuged at 400rcf for 25 minutes at room temperature. The centrifuge was allowed to stop without the aid of any brakes so that the density gradient was not disturbed. The tubes were removed carefully to minimize dispersion of the now clearly visible fractions. Red blood cells (RBCs) pelleted at the bottom of each tube, with the PBMCs forming fractions at the interfaces between plasma and density media ( $D < 1.077\text{g/mL}$ ) (Figure 2.5).

The plasma layers were aspirated with Pasteur pipettes after which the PBMC layers were carefully removed with clean Pasteur pipettes and dispensed into 50mL centrifuge tubes. In the first wash step, the PBMC-containing tubes were filled with PBS and centrifuged for 10 minutes at 400rcf with brakes and acceleration set to nine. The supernatants were removed and pellets re-suspended in 10mL PBS. The tubes were again centrifuged for 10 minutes under the same conditions, after which supernatants were discarded. Pellets were re-suspended in 2mL cryomedia (Appendix V) by adding the cryomedia one drop at a time to cryovials on ice.



**Figure 2.5: Position of cell types and density media after centrifugation of whole blood.** PBMCs formed a “halo” at the interface between the Ficoll and the plasma at the top, and were isolated<sup>315</sup>. RBCs = red blood cells; PBMCs = peripheral blood mononuclear cells.

### 2.6.2. Absolute cell counts

Flow cytometry was used to compare the amount and state of healthy control PBMCs to that of patient PID\_012. The service was provided to us by Stellenbosch University Flow Cytometry Unit (SUFCU) using the FACSCanto II. The FACSCanto II has three lasers, 405nm, 488nm and 633nm, allowing eight-color analyses. Its digital electronics capture up to 10 000 events per second. Six antibodies were used to compare the amount and state of CD4<sup>+</sup> and CD8<sup>+</sup> T cells of PID\_012 to that of a healthy control.

Frozen PBMCs were thawed in a water bath at 37°C until pea-sized pellets remained. PBMC growth media (Appendix V) was added to fill each cryovial. The contents were transferred to 15mL polypropylene tubes, which were filled with PBMC growth media and centrifuged at 400rcf at room temperature. Cells were washed three times with PBS as described above. The pellets were re-suspended in 2mL FACS buffer (Appendix V). Approximately 4x10<sup>6</sup> cells per 1mL FACS buffer were added to each of the two tubes (Table 2.13) for PID\_012 and the control. The tubes were centrifuged at 400rcf for five minutes and the supernatant discarded. Antibodies were added to each tube as indicated, which were vortexed and incubated for one hour at 4°C. The tubes each received 1mL FACS buffer and were again centrifuged as described, and the supernatant discarded. Two-hundred µL FACS buffer was added to each tube, which was then vortexed and the cells were examined using the FACSCanto II.

**Table 2.13: Amount of antibodies added to PBMCs of PID\_012 and control for flow cytometric analysis.**

Antibodies (in µL)	TUBE 1 Unstained	TUBE 2 Stained
APC PD-1	-	1.00
APC-Cy7 CD3	-	1.00
PerCP-Cy5.5 CD4	-	8.00
PE-Cy7 CD8	-	1.00
FITC CD38	-	0.50
BV510 HLA-DR	-	1.00
FACS Buffer	20.00	7.50

APC = allophycocyanin, PD-1 = programmed cell death-1; Cy = cyanin; CD = cluster of differentiation; PerCP = peridinin chlorophyll lprotein; PE = phycoerythrin; FITC = phycoerythrin; BV = brilliant violet; HLA-DR = human leukocyte antigen-D related; FACS = fluorescence-activated cell sorting.

## 2.7. CASE-CONTROL ASSOCIATION STUDIES

### 2.7.1. Genotyping

Samples from 382 pulmonary TB patients and 389 healthy controls were genotyped on the Illumina Multi-Ethnic Genotyping Array (MEGA) (Illumina San Diego, CA, USA) at the Hussmann Institute for Human Genomics (University of Miami, Florida, USA). Samples were diluted to 60ng/ $\mu$ L in 96 well plates and stored at -20°C during shipping. Quality control was performed on all samples to determine DNA quality and quantity. Samples were subjected to gel electrophoresis on a 0.8% agarose gel (Conditions: 100V for 1 hour,  $\lambda$ Hind III ladder), and any samples that displayed significant degradation were replaced with fresh aliquots. Known SNPs in the genes identified through exome sequencing were investigated for association with TB susceptibility in this cohort of 382 TB cases and 389 ethnically matched control individuals.

### 2.7.2. Data analysis

The genotyped SNPs were analyzed using PLINK v1.07<sup>316</sup>. Previous analyses using both linkage models and admixture showed that the SAC population is very admixed, with major genetic contributions from the Khoesan (32–43%), the Bantu-speaking Africans (20–36%), the Europeans (21–28%) and to a lesser extent, Asian populations (9–11%)<sup>317</sup>. Ancestry proportions were estimated using the ADMIXTURE software (version 1.3.0)<sup>318</sup>, and related to the following five reference populations: CEU (Utah Residents from North and West Europe), SAN (Upington, South Africa available through a collaboration with Dr. Brenna Henn), LWK (Luhya in Webuye, Kenya), SAS (GIH) (Gujarati Indian from Houston, Texas) and CHB (Han Chinese in Beijing China). Because complete separation was observed when using the data obtained from all five populations, the CHB, which contributed the smallest admixture component, was excluded. Statistical analysis was adjusted by adding the other four populations, together with sex and age, as covariates. Hardy-Weinberg equilibrium (HWE) testing was performed, and only SNPs with a  $p > 0.05$  were retained. The differences in genotype and allele distribution between cases and controls were assessed using logistic regression under an additive model and the six mentioned covariates. The p-values and Odds ratios (ORs) with 95% confidence intervals (CIs) were used to indicate the effect size of TB. Bonferroni corrections for multiple testing were done by considering the number of independent SNPs and each calculation was adjusted accordingly. A p-value of 0.05 was adopted as a threshold for significance.

PLINK v1.07 was used to carry out missingness assessments, both at a per-SNP and a per-sample basis. A cut-off of 0.1 was used, which corresponds to 10% missing information. Thus, SNPs covered in < 90% of study participants, as well as participants of which < 90% of their SNPs were genotyped, were removed. Power calculations were performed for each individual SNP using the online tool GAS Power Calculator ([http://csg.sph.umich.edu/abecasis/CaTS/gas\\_power\\_calculator/index.html](http://csg.sph.umich.edu/abecasis/CaTS/gas_power_calculator/index.html)). It took into account the size of both cases and control cohorts, significance level (0.05), prevalence of the disease (0.1), disease allele frequency (MAF of the cases) and genotype relative risk (1.5).

## CHAPTER 3: RESULTS

INDEX	PAGE
3.1. WHOLE EXOME SEQUENCING	68
3.1.1 Sanger sequencing	70
3.2. VARIANT FREQUENCY DETECTION IN CONTROLS	70
3.2.1. Allele-specific restriction enzyme digest	70
3.2.2 Database searches	72
3.3. <i>IN SILICO</i> VARIANT PREDICTIONS	72
3.3.1. <i>TAP1 I296M</i>	73
3.3.2. <i>MAP3K14 V345M</i>	74
3.3.3. <i>TAP1 P67S</i>	75
3.3.4. <i>ISG15 L54P</i>	75
3.3.5. <i>TYK2 C378F</i>	76
3.3.6. <i>IL12RB1 T432M</i>	77
3.3.7. <i>TYK2 A53T</i>	77
3.4. FUNCTIONAL STUDIES	79
3.4.1. Immunological workup of PID_040	79
3.4.2. Investigation of the effect of <i>MAP3K14 V345M</i>	79
3.4.2.1. <i>Site-directed mutagenesis</i>	79
3.4.2.2. <i>Plasmid transfection into human cells</i>	79
3.4.2.3. <i>Phosphorylation assay</i>	81
3.4.2.4. <i>Downstream functional effects of MAP3K14 V345M</i>	82
3.4.2.5. <i>Effect of MAP3K14 V345M on autophagy</i>	83
3.5. FLOW CYTOMETRY	85
3.6. CASE-CONTROL ASSOCIATION STUDIES	85

### 3.1. WHOLE EXOME SEQUENCING

WES was used to identify novel MSMD-causing variants in three MSMD patients. The total amount of variants identified in each patient ranged between 61 000 and 73 000 (Table 3.1). The average coverage of the target regions were 60-80×, with 95% of exomes covered at least 1× and 90% covered at least 10×. The most variants were observed in PID\_060 from the Bantu population, followed by PID\_012 from the SAC population, with the least amount of variants identified in the Caucasian PID\_060. This is consistent with what would be expected from these populations. The Bantu population is more ancient than both the European and SAC populations, and thus contains more cross-over events; it had more time to introduce more variants. The SAC population is highly admixed, with genetic contributions from several source populations causing high genetic variability. The European population has remain relatively intact over time, with less introduction of genetic heterogeneity. A summary of the quality control statistics, using FastQC, is given in Appendix II.

After filtration a total of 698 variants for PID\_012, 708 for PID\_040 and 1011 for PID\_060 were identified as possibly disease-causing. After removing all variants homozygous in either of the parents, only seven of the remaining variants were located in functionally relevant genes according to OMIM and NCBI (Table 3.2). Genes involved in any aspect of immune defense against invading pathogens were considered relevant, and only these were investigated further.

**Table 3.1: Variants identified in the MSMD patients after WES.**

	PID_012	PID_040	PID_060
Total variants	70 881	61 173	72 204
All synonymous and non-frameshifts removed	13 006	11 925	14 013
Remove all variants with a frequency >1% in 1KGP	2 772	2 495	3 474
Remove all variants with a frequency >1% in ESP6500	2 549	2 160	3 218
Remove all variants with negative GERP+++ scores	1 791	1 535	2 338
Remove all variants with positive FATHMM scores	698	708	1 011
Novel variants	105	114	153
Variants with rs numbers	593	594	858

1KGP = 1000 Genomes Project; ESP = Exome Sequencing Project; GERP = Genomic evolutionary rate profiling; FATHMM = Functional analysis through hidden Markov models; rs = Reference single nucleotide polymorphism.

**Table 3.2: Potentially disease-causing variants identified in functionally relevant genes in the MSMD patients.**

Patient	Gene	Homozygous variants	Heterozygous variants	rs ID's	Frequency
PID_012	<i>TSPAN33</i>	c.G331A, p.A111T		rs112480251	9.063x10 <sup>-5</sup>
	<i>JAK3</i>		c.G2753C:p.R918P		4.096x10 <sup>-5</sup>
	<i>TAP1</i>		c.A888G:p.I296M		
PID_040	<i>MAP3K14</i>	c.G1033A:p.V345M			
	<i>HLA-DRB5</i>	c.305_306del:p.A102fs c.308_309CCAG			
PID_060	<i>TAP1</i>	c.C199T:p.P67S		rs375389015	3.897x10 <sup>-5</sup>

Frequencies were determined by ExAC Browser Beta. rs ID's = Reference single nucleotide polymorphism cluster identifies; TSPAN = Tetraspanin; JAK = Janus kinase; TAP = Transporter associated with antigen processing; MAP3K14 Mitogen activated protein kinase kinase kinase 14; HLA-DRB = human leukocyte antigen-D related beta chain.

The two *HLA-DRB5* variants (*c.305\_306del* and *c.308\_309CCAG*) identified in patient PID\_040 were excluded from further analysis. The two variants were homozygous in the patient as well as in the healthy mother, indicating that they are not likely to cause disease. They are also absent from the father, making them very likely to be sequencing artefacts. The human leukocyte antigen (HLA) loci are encoded by a 4-Mb region on chromosome 6p21 that has been characterized as the most variable region in the human genome, and variants situated in such variable regions are very difficult to call.

Potentially disease-causing variants situated in MSMD-causing genes were identified in four of the investigated TBM patients (Table 3.3), all of which were heterozygous. A list of all the MSMD-associated genes investigated for potentially disease-causing variants are given in Appendix VII.

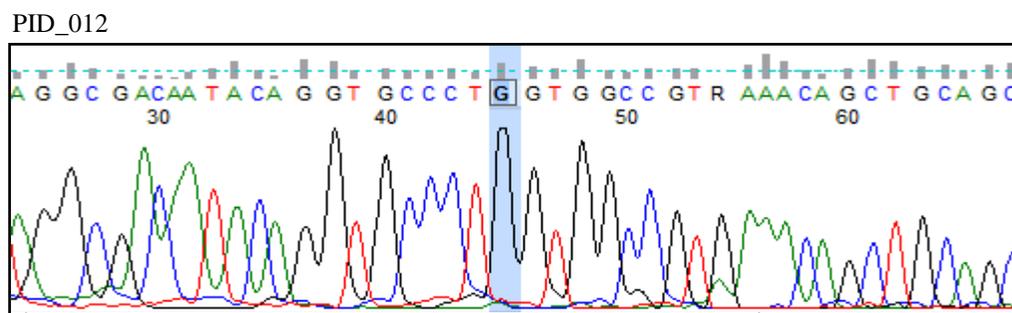
**Table 3.3: Potentially MSMD-causing variants identified in the TBM patients.**

Patient	Gene	Heterozygous variants	rs ID's	Frequency
D5371	<i>ISG15</i>	c.T161C:p.L54P	None	
	<i>TYK2</i>	c.G1133T:p.C378F	None	
D5374	<i>IL12RB1</i>	c.C1295T:p.T432M	rs200875188	2.901x10 <sup>-4</sup>
D5377	<i>TYK2</i>	c.G157A:p.A53T	rs55762744	8.002x10 <sup>-3</sup>
D5378	<i>ISG15</i>	c.T161C:p.L54P	None	

Frequencies were determined by ExAC Browser Beta. rs ID's = Reference single nucleotide polymorphism cluster identifies; ISG = Interferon-stimulated gene; TYK = Tyrosine kinase; IL12RB1 = Interleukin 12 receptor beta 1.

### 3.1.1. Sanger sequencing

Sanger sequencing was used to determine whether the variants identified in the MSMD patients were true variants or sequencing artefacts. All these results are shown in Appendix IX. Only the *JAK3 R918P* variant identified in PID\_012 (Table 3.2) was found to be a sequencing artefact and was therefore excluded from further analysis. Figure 3.1 shows the Sanger sequencing results of this variant in PID\_012. The wild-type guanine (G) highlighted in blue is present at position 2753 of *JAK3*, strongly suggesting that the variant identified by WES (a cytosine at this position) was a sequencing error.

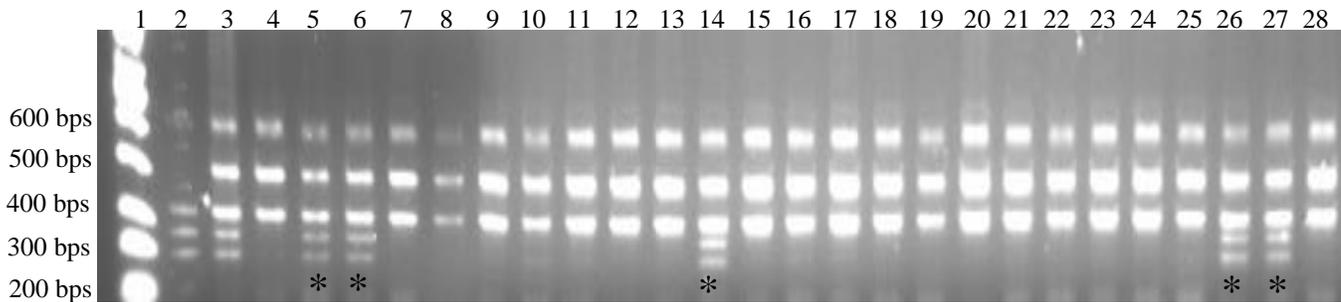


**Figure 3.1: Sanger sequencing results for *JAK3 R918P* in PID\_012.** The variant identified by WES is absent from the patient, indicating that this result was a sequencing error.

## 3.2. VARIANT FREQUENCY DETECTION IN CONTROLS

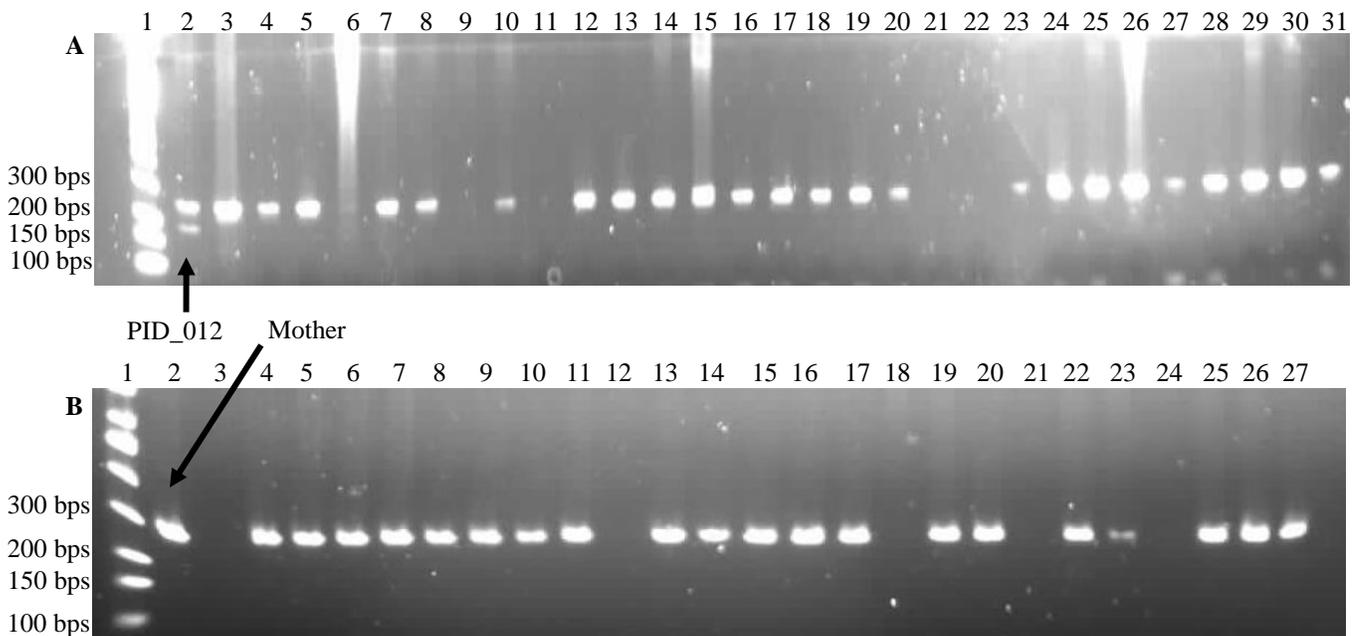
### 3.2.1. Allele-specific restriction enzyme analysis

To determine if the putative MSMD-causing sequence variants were common, non-disease-causing polymorphisms specific to a certain population, and not present in public databases, the MAF for each was determined in ethnically matched control individuals. The *TSPAN33 A111T* and *TAP1 I296M* variants in PID\_012 were genotyped by allele-specific restriction enzyme analysis (ASREA) using the *MscI* restriction enzyme. Following genotyping of 100 control individuals for *TSPAN33 A111T*, which included the patient's mother, a MAF of 4.3% was observed (Figure 3.2). It should be noted that the A allele of the variant was only observed in the homozygous state in PID\_012. However, given that this variant occurred at such a high frequency in a control group, it was categorized as a common polymorphism and not a rare potentially disease-causing mutation. For this reason, it was excluded from further analysis. All samples genotyped are shown in Appendix X.



**Figure 3.2: Enzyme digest results for *TSPAN33 A111T* identified in PID\_012.** Lane 1: 100bp DNA ladder; Lane 2: PID\_012; Lane 3: PID\_012's healthy mother; Lanes 4-28: ethnically matched control individuals. The mother and five other control individuals (lanes 5, 6, 14, 26, 27) are heterozygous for *TSPAN33 A111T*, while all other controls are homozygous wild-type. Only the patient is homozygous for the variant. \* = heterozygous controls.

Furthermore, the *G* allele of *TAP1 I296M* identified in PID\_012 was not detected in any of the control individuals, including the patient's mother (Figure 3.3.). The absence of this variant from 92 ethnically matched controls (Appendix XI) provided substantial evidence for its involvement in disease pathogenesis and was therefore included in further analysis.



**Figure 3.3: Representation of enzyme digest results for the *TAP1 I296M* variant identified in PID\_012.** Lane 1(A): 100bp DNA ladder; Lane 2(A): PID\_012; Lanes 3-31(A): ethnically matched control individuals; Lane 1(B): 100bp DNA ladder; Lane 2(B): PID\_012's healthy mother; Lanes 3-27(B): ethnically matched control individuals. Neither the mother nor any of the controls were carriers of the *TAP1 I296M* variant. Amplification was unsuccessful for lanes containing no product.

### 3.2.2. Database searches

We did not have access to suitable controls for genotyping of neither the *MAP3K14 V345M* nor the *TAPI P67S* variants identified in patients PID\_040 and PID\_060, respectively. The databases mentioned in section 2.3.2 were thus interrogated to determine the MAFs thereof. Neither of the variants were identified in these databases, lending credibility to the notion that these may be disease-causing variants.

However, according to ExAC Browser, the *T* allele of *TAPI P67S* (rs375389015) has been previously identified in heterozygous form in three South Asian individuals and one European individual (MAF =  $3.897 \times 10^{-5}$ , Table 3.4). PID\_060 is homozygous for this allele and it can be speculated that, in a homozygous state, this variant may be disease-causing.

**Table 3.4: Population frequencies of the *TAPI P67S* variant caused by a C/T nucleotide exchange.**

Population	Allele Count	Allele Number	Homozygotes	Allele Frequency
African	0	7054	0	0
East Asian	0	7638	0	0
European (Finnish)	0	5242	0	0
European (non-Finnish)	1	56258	0	$1.778 \times 10^{-5}$
Latino	0	10350	0	0
South Asian	3	15326	0	$1.957 \times 10^{-4}$
Other	0	764	0	0
Total	4	102632	0	$3.897 \times 10^{-5}$

Allele count = amount of mutant alleles observed; Allele number = total amount of alleles investigated; Homozygotes = amount of homozygote mutant individuals identified; Allele frequency = total frequency at which mutant allele was observed (i.e. allele count divided by allele number), as determined by ExAC Browser Beta.

### 3.3. *IN SILICO* VARIANT PREDICTIONS

SIFT, MutationTaster2 (MT2) and PolyPhen-2 (PP-2) were used to predict the probability that each identified sequence variant is disease-causing. Table 3.5 and Table 3.6 summarizes the results of these predictions for the MSMD and TBM patients, respectively. The prediction algorithms are different for each prediction tool and therefore the predictions made by each differs. HOPE was used to further explore the functional consequences of all variants, the results of which are shown in the succeeding sections.

**Table 3.5: *In silico* predictive results of the variants identified in the MSMD patients.**

	SIFT	MT2	PP-2: HumDiv	HumVar
PID_012: <i>TAP1 I296M</i>	T: 0.051	Dis: 0.993	Poss: 0.924	Poss: 0.839
PID_040: <i>MAP3K14 V345M</i>	D: 0.001	N/A	Prob: 1.000	Prob: 0.996
PID_060: <i>TAP1 P67S</i>	T: 0.108	P: 0.999	Poss: 0.596	B: 0.880

T = tolerated; D = damaging; Dis = disease-causing; N/A = not applicable; P = polymorphism, Poss = possibly disease-causing; Prob = probably disease-causing; B = benign. SIFT = Sorting intolerant from tolerant; MT2 = MutationTaster2; PP-2 = PolyPhen-2; TAP = Transporter associated with antigen processing; MAP3K14 = Mitogen activated protein kinase kinase kinase 14.

**Table 3.6: *In silico* predictive results of the variants identified in the TBM patients.**

	SIFT	MT2	PP-2: HumDiv	HumVar
D5371, D5378: <i>ISG15 L54P</i>	D: 0.003	Dis: 1.000	Prob: 1.000	Prob: 1.000
D5371: <i>TYK2 C378F</i>	D: 0.001	Dis: 1.000	Prob: 0.999	Prob: 0.971
D5374: <i>IL12RB1 T432M</i>	D: 0.000	P: 1.000	Prob: 0.999	Poss: 0.987
D5377: <i>TYK2 A53T</i>	D: 0.033	Dis: 0.998	Prob: 0.997	Prob: 0.922

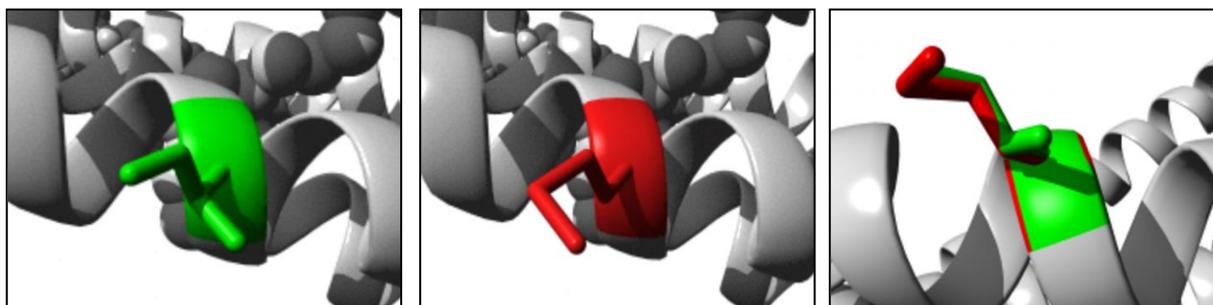
D = damaging; Dis = disease-causing; P = polymorphism, Poss = possibly disease-causing; Prob = probably disease-causing. SIFT = Sorting intolerant from tolerant; MT2 = MutationTaster2; PP-2 = PolyPhen-2; ISG = Interferon-stimulated gene; TYK = Tyrosine kinase; IL12RB1 = Interleukin 12 receptor beta 1.

### 3.3.1. *TAP1 I296M*

The *TAP1 I296M* variant identified in PID\_012 results in the substitution of an isoleucine (Ile) at amino acid position 296 with a methionine (Met). Figure 3.4 shows the schematic structures of the wild type and mutant amino acids generated by HOPE. This exchange occurs on the surface of the protein, as indicated in Figure 3.5. Three of the four prediction tools predicted this variant to be damaging and disease-causing. The Tolerated prediction by SIFT is only just above the threshold of 0.05. The variant is not present in 1KGP nor in ExAC Browser.



**Figure 3.4: Mutation of Ile to Met observed at position 296 in *TAP1* of PID\_012.** The backbone is indicated in red and the unique side chains in black. Figure drawn by HOPE.



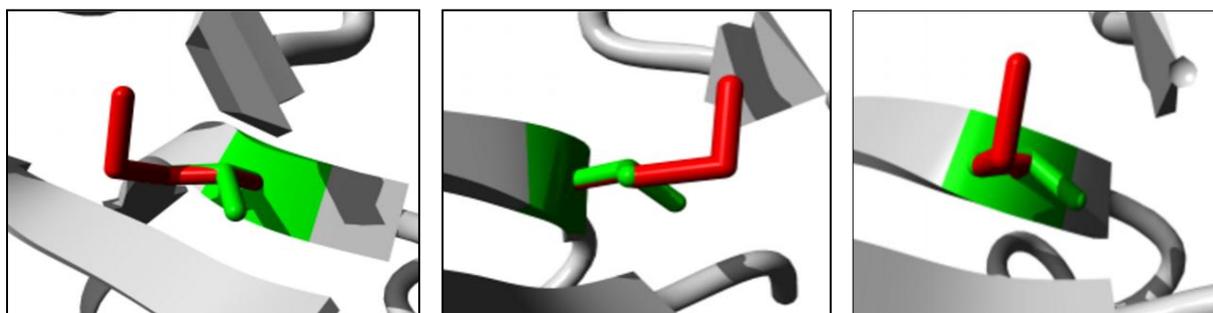
**Figure 3.5: Mutated and wild-type amino acids at position 296 of TAP1.** The mutated amino acid is indicated in red, with the wild-type in green. Figure drawn by HOPE.

### 3.3.2. *MAP3K14 V345M*

The *MAP3K14 V345M* variant identified in PID\_040 leads to the substitution of a valine (Val) with a Met at position 345 of NIK. The schematic representation of these two amino acids are shown in Figure 3.6. This variant is situated near a conserved site, buried in the core of the protein. Figure 3.7. gives a 3D representation of Val and Met.



**Figure 3.6: Mutation of Val to Met, as observed at position 345 in MAP3K14 of PID\_040.** The conserved backbone is indicated in red and the unique side chains in black. Figure drawn by HOPE.



**Figure 3.7: Mutated and wild-type amino acids at position 345 of MAP3K14.** The mutated amino acid is indicated in red, with the wild-type in green. Figure drawn by HOPE.

MT2 was not able to generate any results for this amino acid exchange, indicating the uncharacterized state of *MAP3K14*. All other scores hint towards the damaging, disease-causing nature of this variant, with great confidence.

### 3.3.3. *TAP1 P67S*

Patient PID\_060 presented with a substitution of a proline (Pro) with serine (Ser) at position 67. Figure 3.8 compares the wild-type amino acid to the mutated one. Only HumDiv of PP-2 predicted the variant to possibly be damaging, with less than 60% confidence. All other tools predict it to be a benign polymorphism not associated with disease. HOPE was not able to produce 3D structures for this variant.



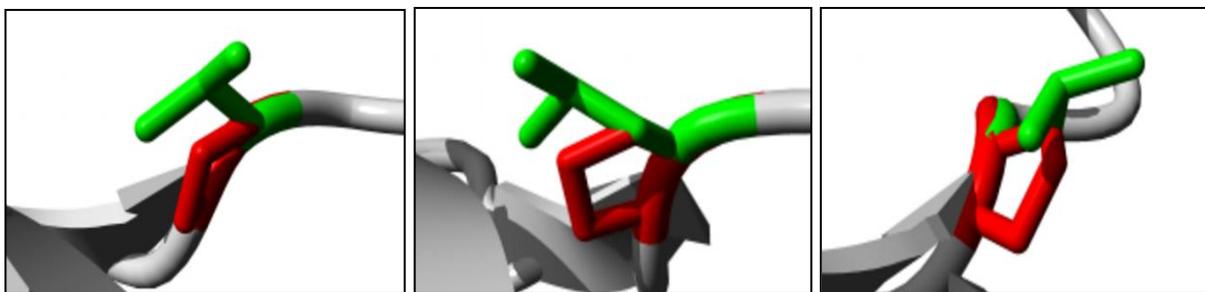
**Figure 3.8: Mutation of Pro to Ser, as observed at position 67 in *TAP1* of PID\_060.** The conserved backbone is indicated in red and the unique side chains in black. Figure drawn by HOPE.

### 3.3.4. *ISG15 L54P*

The exchange of a leucine (Leu) with a Pro was observed at position 54 of *ISG15* in TBM patients D5371 and D5378. Figure 3.9 shows the wild-type and mutant amino acids. Figure 3.10 compares the structures of Pro and Leu on a 3D level.



**Figure 3.9: Mutation of Leu to Pro, as observed at position 54 in *ISG15* of D5371 and D5378.** The conserved backbone is indicated in red and the unique side chains in black. Figure drawn by HOPE.



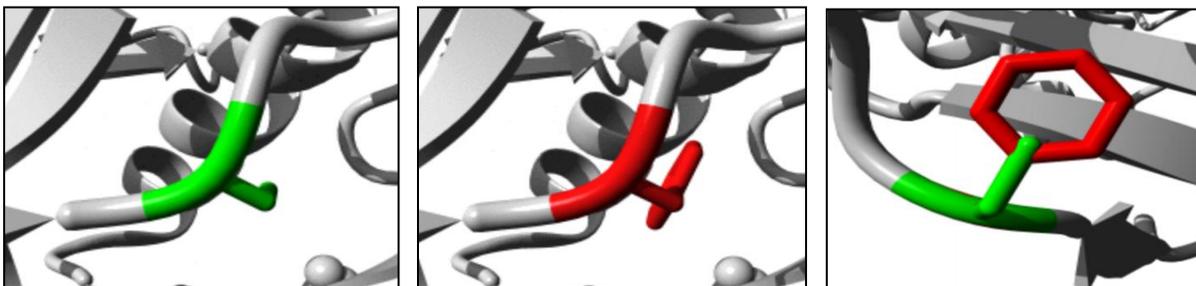
**Figure 3.10: Mutated and wild-type amino acids at position 54 of ISG15.** The mutated amino acid is indicated in red, with the wild-type in green. Figure drawn by HOPE.

### 3.3.5. *TYK2 C378F*

The mutation of a cysteine (Cys) at position 378 of TYK2 to a phenylalanine (Phe) was observed in TBM patient D5371. Schematic representations of the two amino acids are shown in Figure 3.11. This variation occurs in the core of the protein. Figure 3.12 compares the 3D structure of Phe to that of Cys.



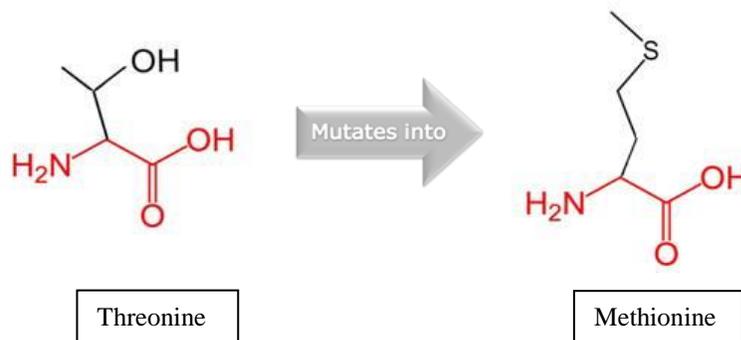
**Figure 3.11: Mutation of Cys into Phe, as observed at position 378 in TYK2 of D5371.** The conserved backbone is indicated in red and the unique side chains in black. Figure drawn by HOPE.



**Figure 3.12: Mutated and wild-type amino acids in TYK2 of D5371.** The mutated amino acid is indicated in red, with the wild-type in green. Figure drawn by HOPE.

### 3.3.6. *IL12RB1 T432M*

A mutation of threonine (Thr) to Met at position 432 was observed in TBM patient D5374. Figure 3.13 shows the structures of the original and mutant amino acids as generated by HOPE. ExAC Browser documented the mutant allele 35 times in five different populations, with a MAF of  $2.901 \times 10^{-4}$  (Table 3.7).



**Figure 3.13: Mutation of Thr into Met, as observed at position 432 in *IL12RB1* of D5374.** The conserved backbone is indicated in red and the unique side chains in black. Figure drawn by HOPE.

**Table 3.7: Population frequencies of the *IL12RB1 T432M* variant.**

Population	Allele Count	Allele Number	Homozygotes	Allele Frequency
African	23	9796	0	0.002348
East Asian	4	8626	0	0.0004637
European (Finnish)	0	6614	0	0
European (non-Finnish)	2	6668	0	0.00003
Latino	2	11560	0	0.000173
South Asian	4	16496	0	0.00002425
Other	0	898	0	0
Total	35	120658	0	$2.901 \times 10^{-4}$

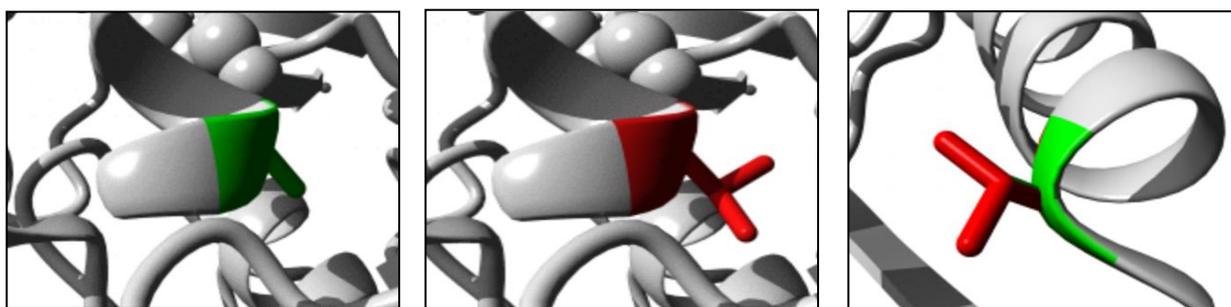
Allele count = amount of mutant alleles observed; Allele number = total amount of alleles investigated; Homozygotes = amount of homozygote mutant individuals identified; Allele frequency = total frequency at which mutant allele was observed (i.e. allele count divided by allele number).

### 3.3.7. *TYK2 A53T*

A mutation of alanine (Ala) to Thr at position 53 of *TYK2* was observed in TBM patient D5377 (Figure 3.14). The wildtype and mutant amino acids are compared in three dimensions in Figure 3.15. ExAC Browser observed the mutant allele 958 times across all investigated populations, with a MAF of  $8.002 \times 10^{-3}$  (Table 3.8).



**Figure 3.14: Mutation of Ala into Thr, as observed at position 53 in TYK2 of D5377.** The conserved backbone is indicated in red and the unique side chains in black. Figure drawn by HOPE.



**Figure 3.15: Mutated and wild-type amino acids in TYK2 of D5377.** The mutated amino acid is indicated in red, with the wild-type in green. Figure drawn by HOPE.

**Table 3.8: Population frequencies of the TYK2 A53T variant.**

Population	Allele Count	Allele Number	Homozygotes	Allele Frequency
African	24	10244	0	0.002343
East Asian	2	8600	0	0.0002326
European (Finnish)	33	6592	0	0.005006
European (non-Finnish)	704	65370	5	0.01077
Latino	44	11530	1	0.003816
South Asian	141	16506	2	0.008542
Other	10	882	0	0.01134
Total	958	119724	8	$8.002 \times 10^{-3}$

Allele count = amount of mutant alleles observed; Allele number = total amount of alleles investigated; Homozygotes = amount of homozygote mutant individuals identified; Allele frequency = total frequency at which mutant allele was observed (i.e. allele count divided by allele number), as determined by ExAC Browser Beta.

All four the variants identified in the MSMD genes of the TBM patients are predicted to be disease-causing and either probably or possibly damaging, except for the prediction by MT2 for *IL12RB1 T432M*.

### 3.4. FUNCTIONAL STUDIES

#### 3.4.1. Immunological workup of PID\_040

Basic routine immunological investigations of patient PID\_040 (as part of appropriate standard of care) over four years indicated a dramatic decrease of all investigated cell types over time, including the almost complete depletion of NK, B and CD8 cells (Table 3.9).

**Table 3.9: Total cell counts from 2010 until 2014, in units of cells per microliter.**

Cell counts of the patient at different ages						Reference
Subsets	2 Yr, 1 Mo	2 Yr, 3 Mo	4 Yr, 8Mo	5 Yr, 1 Mo	6 Yr	2-6 Yr
Lymphocytes	15774 ( <b>H</b> )	9522 ( <b>H</b> )	6720 ( <b>H</b> )	3999	2366	2340-5028
T cells	10719 ( <b>H</b> )	6785 ( <b>H</b> )	4682 ( <b>H</b> )	2655	1886	1578-3707
CD4+	8743 ( <b>H</b> )	5523 ( <b>H</b> )	3954 ( <b>H</b> )	2258 ( <b>H</b> )	1516	870-2144
CD8+	1584 ( <b>H</b> )	959	813	533	369 ( <b>L</b> )	472-1107
NK cells	2549 ( <b>H</b> )	1263 ( <b>H</b> )	214	778 ( <b>H</b> )	70 ( <b>L</b> )	155-565
B cells	2156 ( <b>H</b> )	1339 ( <b>H</b> )	1783 ( <b>H</b> )	398 ( <b>L</b> )	362 ( <b>L</b> )	434-1274

All subtypes decreased significantly over time. Normal reference counts: CD45 cells: 6000-9000; T cells: 1400-3700; CD4 cells: 700-2200; CD8 cells: 490-1300; B cells: 390-1400; NK cells: 130-720. CD = cluster of differentiation; NK = natural killer.

#### 3.4.2. Investigation of the effect of *MAP3K14 V345M*

To determine the functional involvement of *MAP3K14 V345M* in disease outcome, wild-type plasmid constructs expressing wild-type *MAP3K14* and *MAP3K14 V345M* were each co-transfected, along with plasmid constructs expressing wild-type *IKK $\alpha$* , into human HEK293T cells. The kinase activity of mutant *MAP3K14* was then compared to that of wild-type *MAP3K14* by measuring the degree of phosphorylation of *IKK $\alpha$* .

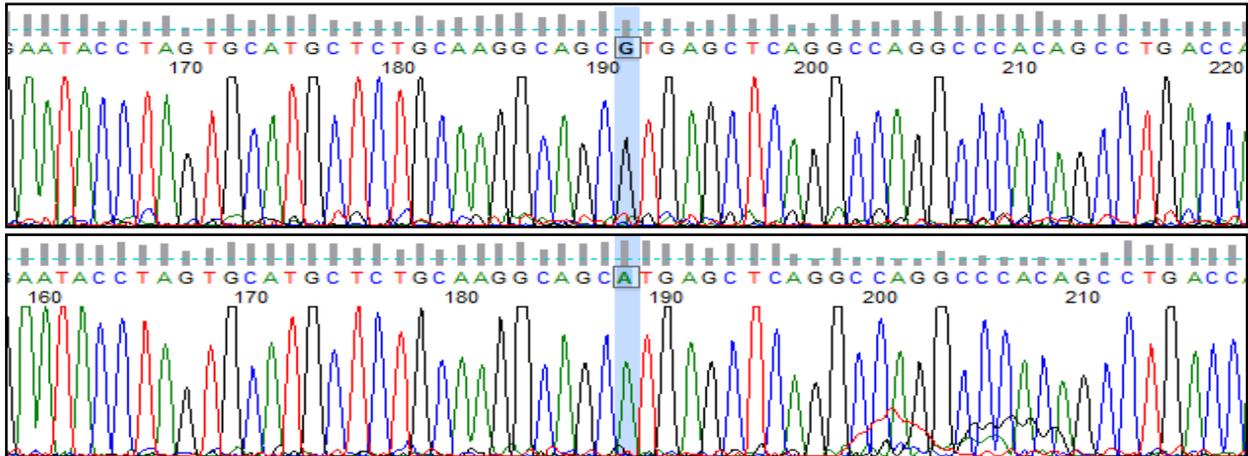
##### 3.4.2.1. Site-directed mutagenesis

Wild-type and mutated *MAP3K14* plasmids were generated using the Q5® site-directed mutagenesis kit. Sanger sequencing confirmed the successful introduction of the mutation (c.G1033A) (Figure 3.16).

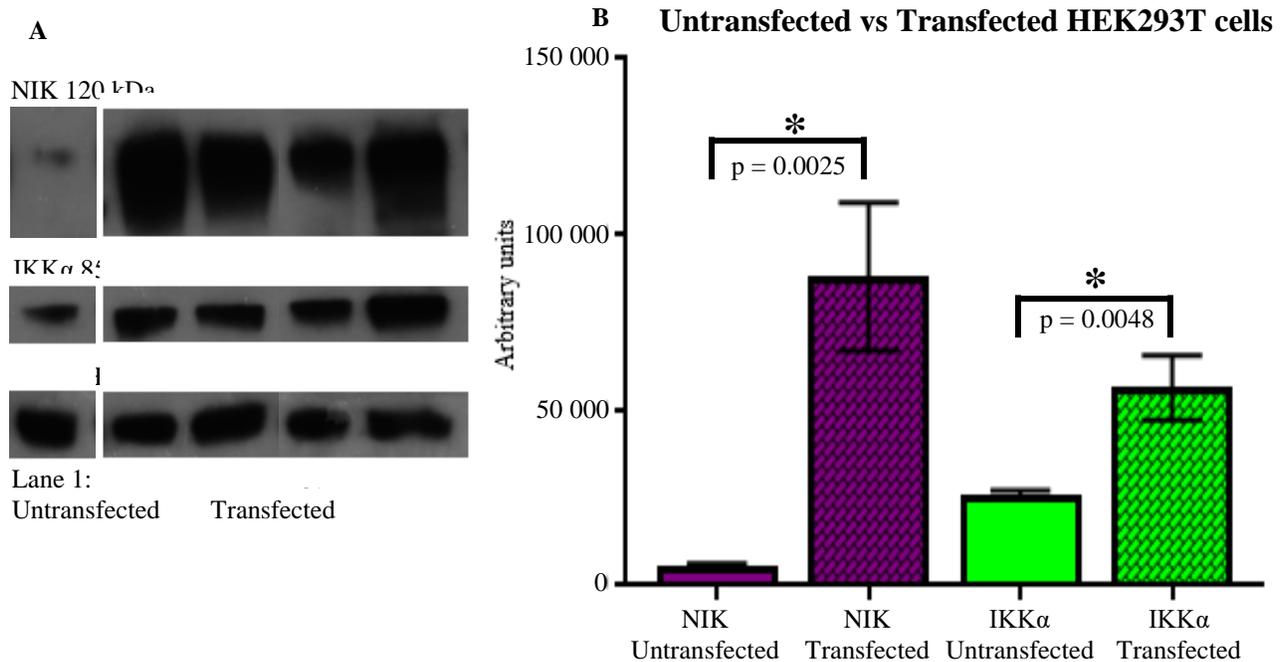
##### 3.4.2.2. Plasmid transfection into human cells

The plasmids were subsequently transfected to healthy human HEK293T cells. The successful transfection of these plasmids were proven using SDS-PAGE and Western blots, the results of

which are shown in Figure 3.17. No endogenous MAP3K14 (NIK) is present in HEK293T cells, with IKK $\alpha$  present at substantial levels. Figure 3.17 shows the significantly increased levels of both NIK ( $p = 0.0025$ ) and IKK $\alpha$  ( $p = 0.0048$ ) after transfection of these plasmids into HEK293T cells.



**Figure 3.16: Sanger sequencing of PWZL-NEO-FLAG-MAP3K14 (top) and PWZL-NEO-FLAG-MAP3K14-V345M (bottom) plasmids.** The c.G1033A mutation induced using Q5® site-directed mutagenesis is highlighted in blue, with no off-target variants induced within the surrounding nucleotides.

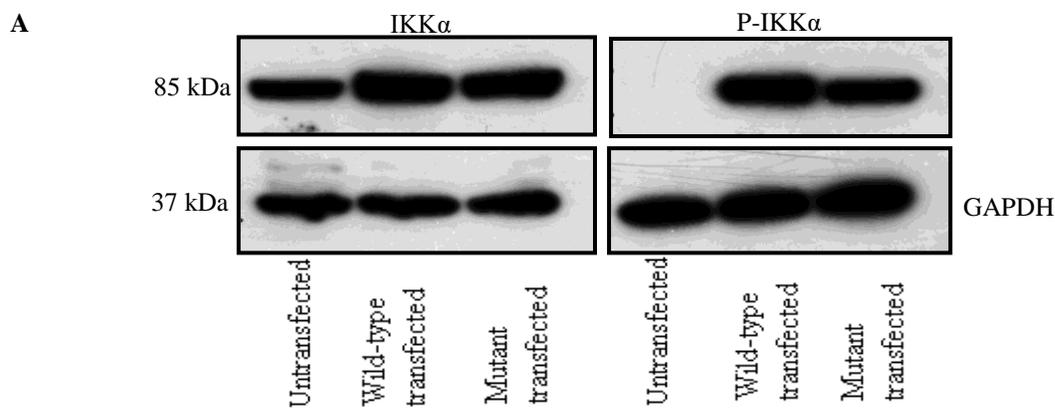


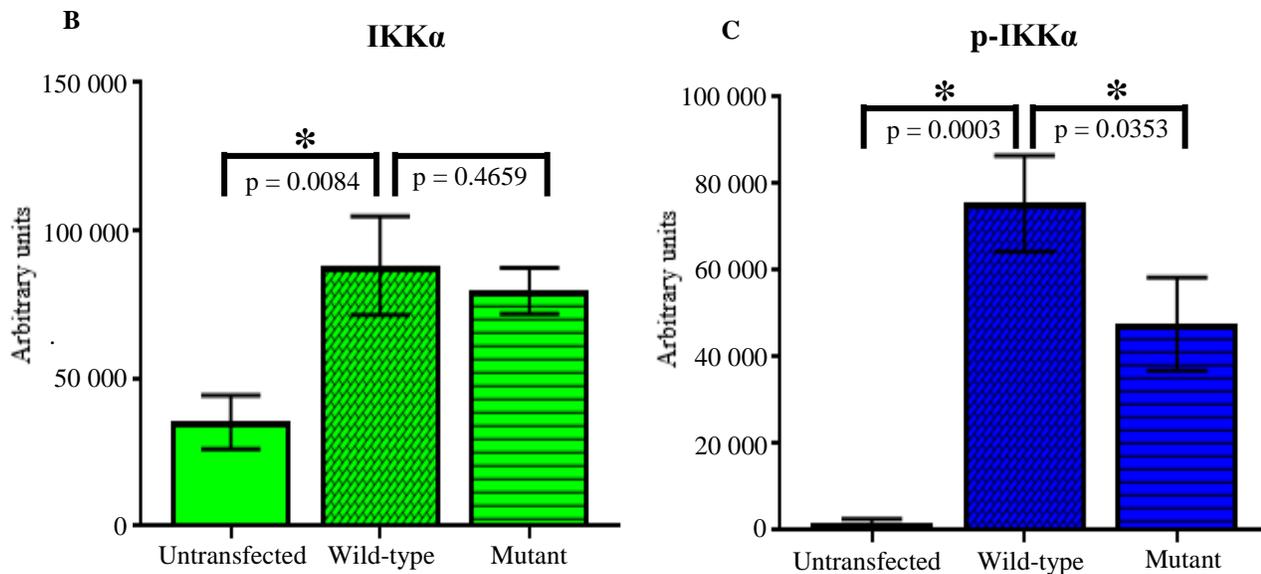
**Figure 3.17: Investigation of successful transfection of the *MAP3K14* and *IKK $\alpha$*  plasmids into HEK293T cells.** Both NIK (120kDa) and *IKK $\alpha$*  (85kDa) levels are significantly increased after transfection of the *MAP3K14* and *IKK $\alpha$*  plasmids into HEK293T cells, proving successful transfection. A) Lane 1: untransfected HEK293T cells; lanes 2-5: cells transfected with *IKK $\alpha$*  and wild-type *MAP3K14*. B) The corresponding quantified bar graphs. All experiments were conducted in triplicate. \* indicate significant association according to unpaired t-test. Three samples were used per group. HEK293T = Human embryonic kidney cells; kDa = kilo Dalton.

### 3.4.2.3. Phosphorylation assay

The difference in *IKK $\alpha$*  and phosphorylated *IKK $\alpha$*  (p-*IKK $\alpha$* ) between transfected and untransfected HEK293T cells were also investigated (Figure 3.18). Phosphorylated (p-*IKK $\alpha$* ) is not endogenously present in HEK293T cells. *IKK $\alpha$*  ( $p = 0.0084$ ) and p-*IKK $\alpha$*  ( $p = 0.0003$ ) increased significantly after plasmid transfections, further proving successful transfection.

Following successful transfections, the kinase activity of wild-type and mutant NIK was compared to one another by measuring the difference in phosphorylation of *IKK $\alpha$*  between these groups. The Western blot results are shown in Figure 3.18. The level of *IKK $\alpha$*  phosphorylation observed was significantly less in cells transfected with the mutant construct compared to the wildtype ( $p = 0.0353$ ), indicating that the mutation impairs kinase activity. No significant changes in the *IKK $\alpha$*  levels were observed between the groups, as expected, since the mutation only affects the kinase activity of NIK and not the production of *IKK $\alpha$* .



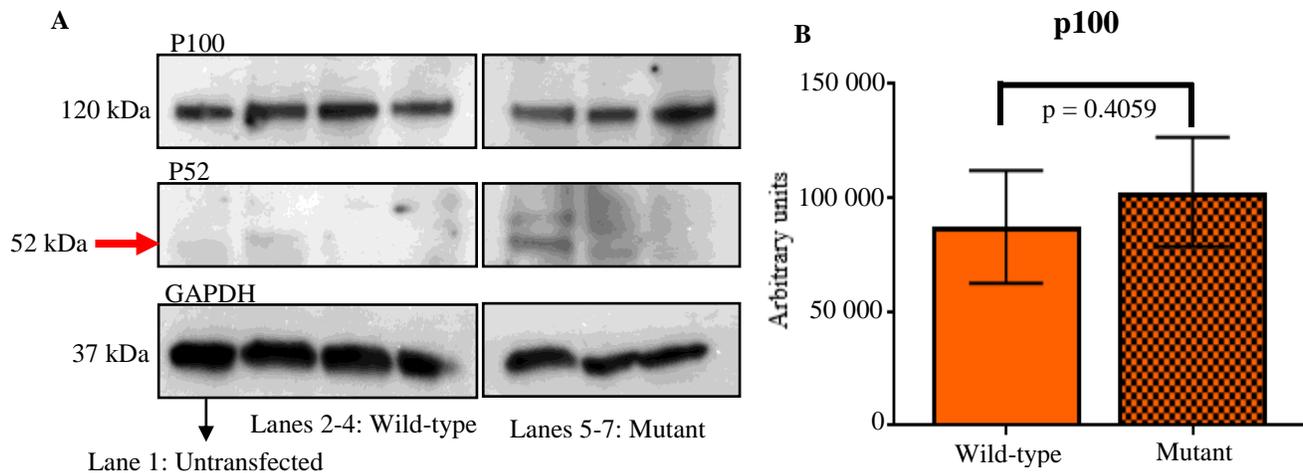


**Figure 3.18: Investigation of kinase activity of mutant versus wild-type NIK.** (A and B) Lower levels of IKK $\alpha$  in untransfected compared to transfected cells proved that the transfection was successful. Non-significant decrease in IKK $\alpha$  in cells transfected with *MAP3K14 V345M* compared to wild-type *MAP3K14*. (A and C) No p-IKK $\alpha$  detected in untransfected HEK293T cells, with significant amounts present in transfected cells prove efficient transfection. Significant decrease in p-IKK $\alpha$  levels in cells transfected with *MAP3K14 V345M* compared to wild-type *MAP3K14*. \* indicate significant association according to unpaired t-test. Three samples were used per group. kDa = kilo Dalton.

#### 3.4.2.4. Downstream functional effects of MAP3K14 V345M

Phosphorylated IKK $\alpha$  results in the ubiquitination and proteosomal degradation of p100, and p52 is produced. The difference in p100 as well as p52 levels between wild-type and mutant NIK was thus also investigated to determine whether *MAP3K14 V345M* had any downstream effects as well. However, because HEK293T cells were used instead of immune cells obtained from the patient, as would have been the ideal, we were not able to infect these cells to increase IKK $\alpha$  phosphorylation and thus activate p100 breakdown and p52 production.

Despite several attempts, no quantifiable levels of p52 were detected using Western blotting (Figure 3.19 A). Substantial levels of p100 was detected in the untransfected HEK293T cells as well as in three samples of cells transfected with IKK $\alpha$  and PWZL-NEO-FLAG-MAP3K14 (lanes 2-4) and three transfected with IKK $\alpha$  and PWZL-NEO-FLAG-MAP3K14-V345M (lanes 5-7). No differences in p100 levels between these groups were observed (Figure 3.19 B) ( $p = 0.4059$ ).



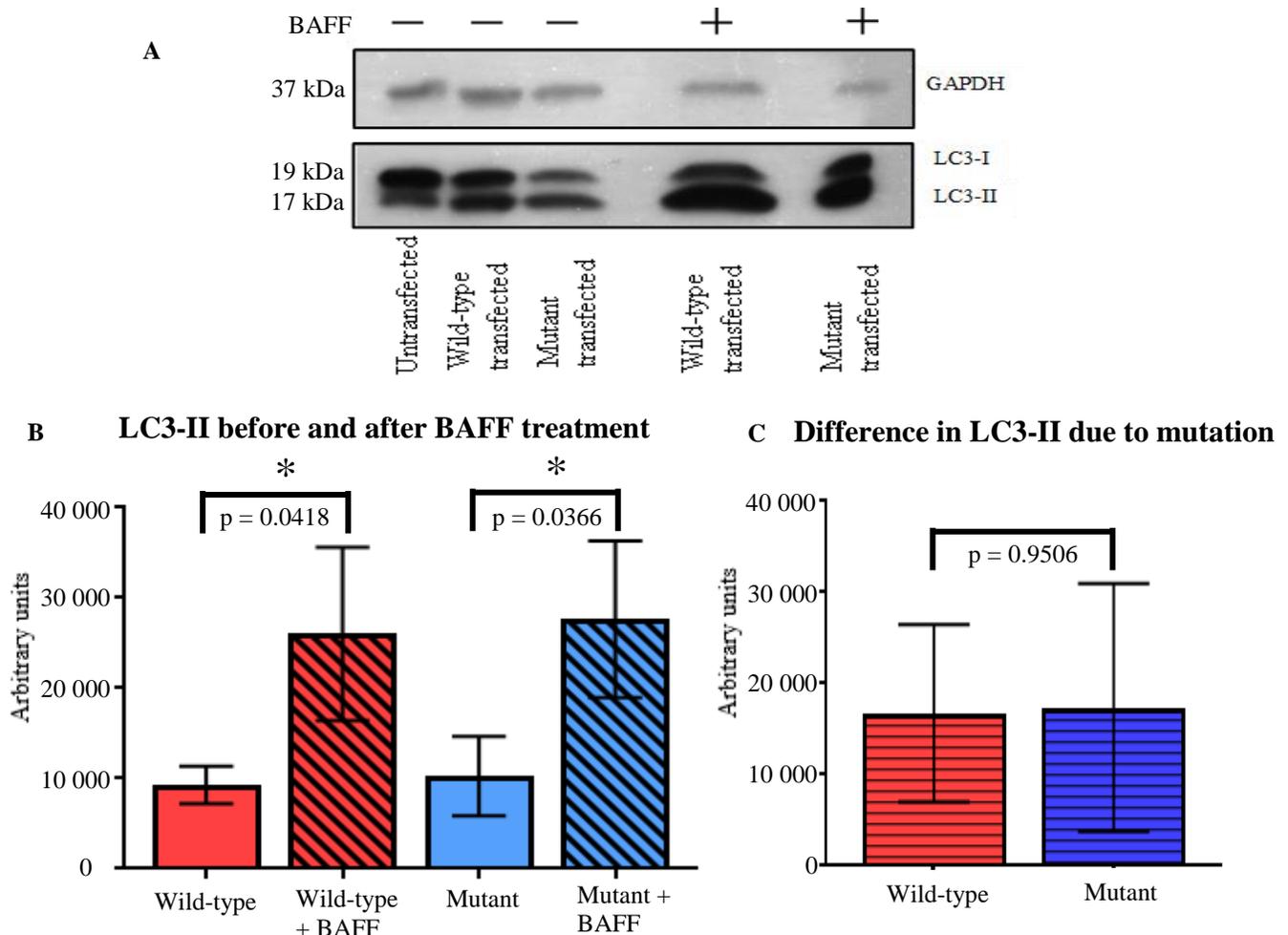
**Figure 3.19: Investigation of p52 and p100 levels of mutant versus wild-type NIK.** (A) No quantifiable p52 levels were observed in untransfected, wild-type-transfected or mutant-transfected samples. (A and B) p100 was detectable at significant amounts, but no differences between the wild-type and mutant groups were observed. p100 is 120 kDa in size and p52 is 52 kDa. P-value was determined using unpaired t-test. Nine samples were used per group. kDa = kilo Dalton.

#### 3.4.2.5. Effect of MAP3K14 V345M on autophagy

NIK and  $\text{IKK}\alpha$  are degraded by autophagy when the function of heat shock protein 90, required for the folding and maturation of certain signalling proteins, is inhibited<sup>319</sup>. The processing of p100 and NF- $\kappa$ B activity is thus inhibited<sup>320</sup>. However, when heat shock stress activates NF- $\kappa$ B, the autophagy pathway is in turn activated, indicating a close interaction and tight regulation between these two pathways<sup>321</sup>.

Autophagy is a very dynamic catabolic process during which double-membrane vesicles, known as autophagosomes, form by engulfing parts of the cytoplasm and subsequently fuse with lysosomes to degrade and recycle their contents<sup>322</sup>. Levels of LC3-II directly correlate with the number of autophagosomes<sup>323</sup>. LC3 immunoblotting is widely used to measure autophagic activity<sup>324–326</sup>. SDS-PAGE and immunoblotting detects two bands of endogenous LC3: LC3-I and LC3-II. LC3-I is present in the cytosol, while LC3-II is bound to PE and localised to autophagosomal membranes<sup>327,328</sup>. The difference between LC3-I and LC3-II is not a representation of LC3 processing, but rather of PE-binding. The majority of anti-LC3 antibodies detect LC3-II with much higher sensitivity than LC3-I<sup>329</sup>, meaning that the comparison between LC3-I and LC3-II, or the measurement of both together, is not sufficient.

The effect of *MAP3K14 V345M* on autophagy was thus also investigated in this study by measuring LC3-II levels. The Western blots and corresponding bar graphs are shown in Figure 3.20 (A and B). Significant differences in LC3-II were observed when comparing cells transfected with wild-type *MAP3K14* and *IKK $\alpha$*  before and after BAFF treatment, as well as those infected with mutant *MAP3K14* and *IKK $\alpha$*  before and after treatment. This indicates that the autophagic pathway functioned effectively. When comparing the difference in LC3-II levels between mutant and wild-type groups (Figure 3.20 C), no significant difference was observed, indicating that *MAP3K14 V345M* has no effect on autophagy.



**Figure 3.20: Measurement of LC3-II levels before and after BAFF treatment.** (A and B) Significant differences are seen when comparing cells transfected with wild-type *MAP3K14* and *IKK $\alpha$*  before and after BAFF treatment, as well as mutant *MAP3K14* and *IKK $\alpha$*  before and after treatment. (C) The mutation does not affect LC3-II levels. \* indicates significant association according to unpaired t-test. Three samples were used per group. BAFF = Bafilomycin A1; kDa = Kilo Dalton.

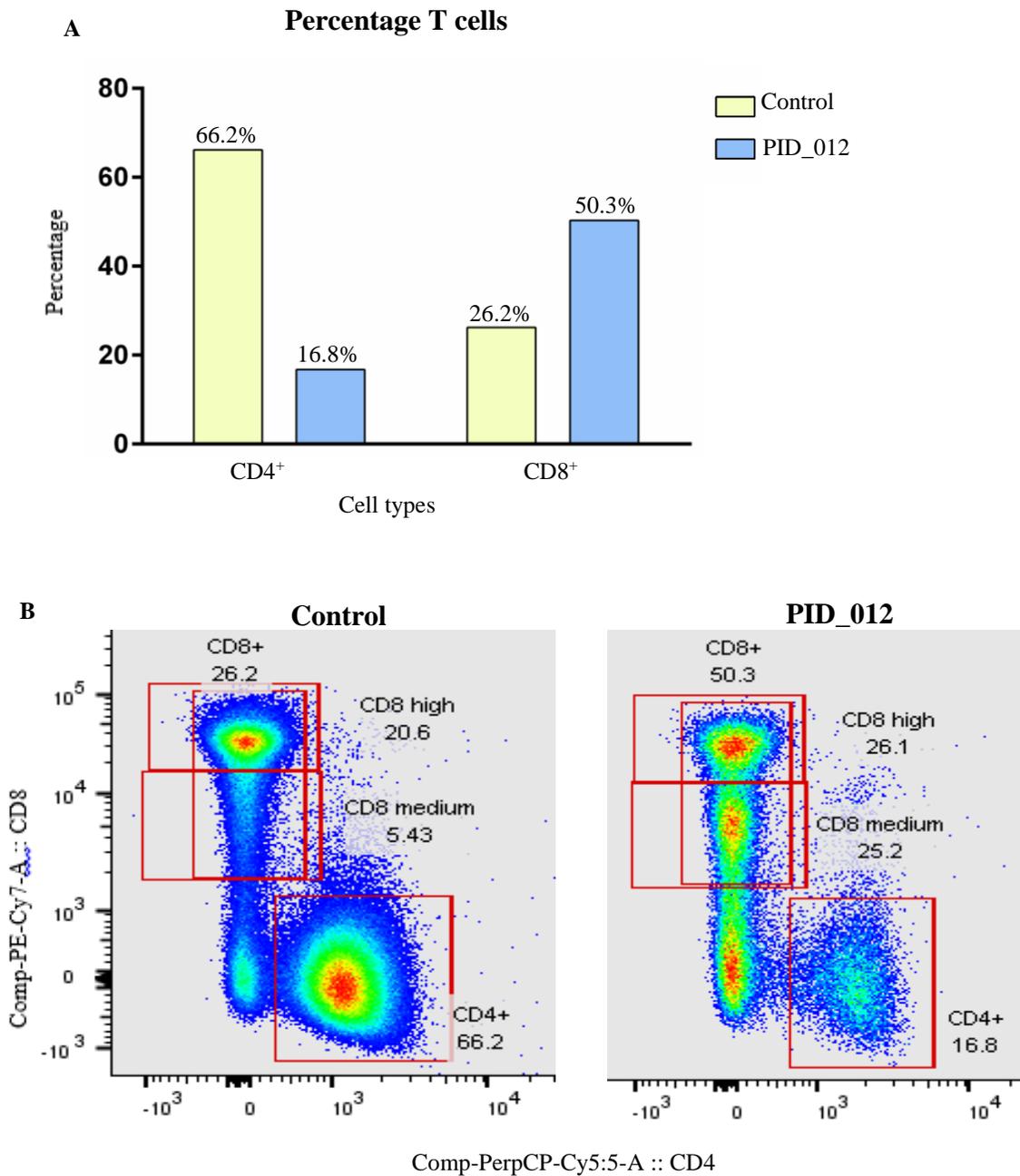
### 3.5. FLOW CYTOMETRY

Flow cytometry was performed for PID\_012 and an ethnically matched healthy control. The total amounts of CD4<sup>+</sup> and CD8<sup>+</sup> T cells were firstly determined. Figure 3.21 (A) shows a significant reduction in CD4<sup>+</sup> T cells in the patient compared to the control, accompanied by a significant increase in CD8<sup>+</sup> T cells. The flow cytometry plot in Figure 3.21 (B) gives a visual representation of the graph. It also shows two distinct groups of CD8<sup>+</sup> T cells (CD8<sup>high</sup> and CD8<sup>medium</sup>) of approximately the same size in the patient, compared to one dominant group in the control. The different groups of CD8<sup>+</sup> T cells in PID\_012 is indicative of different stages of activation of these cells. The degree of activation of CD4<sup>+</sup> and CD8<sup>+</sup> T cells were thus also investigated. Except for CD38<sup>+</sup>PD1<sup>+</sup> cells, all activated cell types were at least slightly increased in the patient compared to the control (Table 3.10; Figure 3.22). Especially in CD38<sup>+</sup> cell populations, these increases were quite severe, with the largest increase in activation seen in the CD38<sup>+</sup>PD1<sup>-</sup> population. Table 3.10 gives the percentages of each cell type indicated in Figure 3.22.

### 3.6. CASE-CONTROL ASSOCIATION STUDIES

We hypothesized that, by identifying disease-causing mutations in PID patients suffering from recurrent TB infections, we have identified two novel candidate genes (*MAP3K14* and *TAP1*) potentially associated with increased TB susceptibility. We thus screened a cohort of 382 TB cases and 389 controls from the SAC population for known SNPs in these two genes to determine whether they are associated with TB.

No association with PTB was observed for the 32 SNPs investigated in *TAP1* and *MAP3K14*. However, using Genetic Association Study (GAS) Power Calculator ([http://csg.sph.umich.edu/abecasis/cats/gas\\_power\\_calculator/index.html](http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html)), it was determined that the power of this study was only large enough (i.e. >80%) for 11 of these SNPs. The cohort investigated was thus not large enough to definitively prove or disprove association between *TAP1* and *MAP4K14* and TB susceptibility in the SAC population. The lack of power explains the very high uncorrected as well as Bonferroni-corrected p-values. All investigated SNPs are shown in Appendix XII. The  $r^2$  and D' values of each SNP in *TAP1* and *MAP3K14* are also shown, combined with the linkage disequilibrium (LD) plots of these two genes.

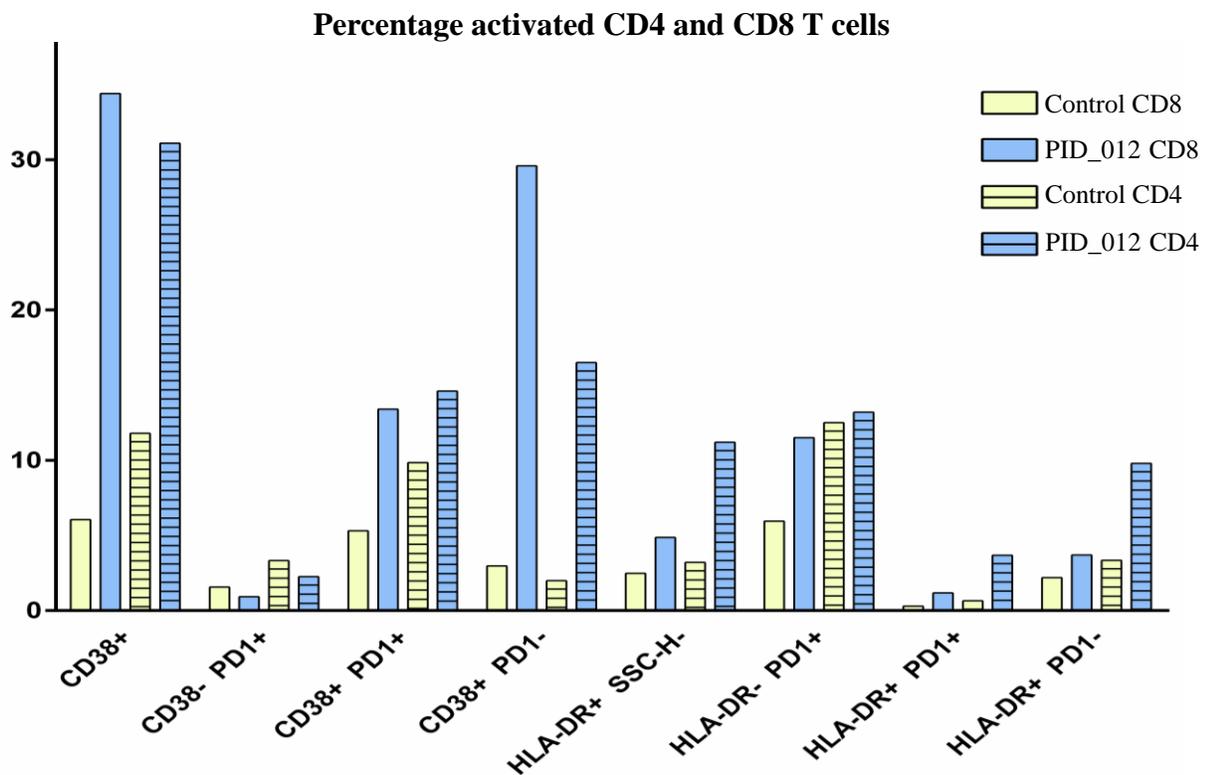


**Figure 3.21: Percentage of CD4<sup>+</sup> and CD8<sup>+</sup> T cells in PID\_012 and a healthy control.** A) A severe decrease in CD4<sup>+</sup> T cells and increase in CD8<sup>+</sup> T cells is observed in PID\_012. B) Three distinct groups of activated CD8<sup>+</sup> T cells are seen. CD = Cluster of differentiation; PerCP = peridinin chlorophyll lprotein, Cy = cyanin; PE = phycoerythrin.

**Table 3.10. Percentages of activated CD4+ and CD8+ T cells.**

Cell populations	Control CD8+	PID_012 CD8+	Control CD4+	PID_012 CD4+
CD38+	6.05%	34.40%	11.80%	31.10%
CD38- PD1+	1.57%	0.92%	3.32%	2.26%
CD38+ PD1+	5.31%	13.40%	9.84%	14.60%
CD38+ PD1-	2.98%	29.60%	1.99%	16.50%
HLA-DR+ SSC-H-	2.48%	4.87%	3.21%	11.20%
HLA-DR- PD1+	5.96%	11.50%	12.50%	13.20%
HLA-DR+ PD1+	0.29%	1.17%	0.65%	3.68%
HLA-DR+ PD1-	2.19%	3.70%	3.35%	9.80%

Eight different cell populations were investigated to determine the difference in activation of PBMCs of PID\_012 and the control. With the exception of CD38<sup>-</sup>PD1<sup>+</sup>, all cell types were at least slightly more activated in the patient. CD = Cluster of differentiation; PD-1 = programmed cell death-1; HLA-DR = human leukocyte antigen-D related. SSC-H = Side scatter pulse height.



**Figure 3.22: Percentages of activated CD4+ and CD8+ T cells.** CD = Cluster of differentiation; PD-1 = programmed cell death-1; HLA-DR = human leukocyte antigen-D related. SSC-H = Side scatter pulse height.

**CHAPTER 4: DISCUSSION**

<b>INDEX</b>	<b>PAGE</b>
4.1. OVERVIEW OF STUDY FINDINGS	90
4.2. WHOLE EXOME SEQUENCING	91
4.2.1. MSMD patients	91
4.2.1.1. <i>MAP3K14: The NF-<math>\kappa</math>B signaling pathway</i>	92
4.2.1.2. <i>TAP1: MHC class I activation</i>	93
4.2.2. TBM cohort	94
4.2.2.1. IL12RB1	95
4.2.2.2. ISG15	96
4.2.2.3. TYK2	97
4.3. <i>IN SILICO</i> VARIANT PREDICTIONS	99
4.3.1. MSMD patients	100
4.3.1.1. TAP1 I296M	100
4.3.1.2. MAP3K14 V345M	104
4.3.1.3. TAP1 P67S	106
4.3.2. TBM patients	108
4.3.2.1. ISG15 L54P	108
4.3.2.2. TYK2 C378F	109
4.3.2.3. TYK2 A53T	112
4.3.2.4. IL12RB1 T432M	113
4.4. FUNCTIONAL STUDIES	115
4.4.1. Effect of <i>MAP3K14 V345M</i> on the function of NIK	115
4.4.2. Downstream effects of <i>MAP3K14 V345M</i>	117
4.4.3. Effect of <i>MAP3K14 V345M</i> on autophagy	118
4.5. FLOW CYTOMETRY	120
4.6. CASE-CONTROL ASSOCIATION STUDIES	123
4.6.1. Population genetics	125
4.6.2. Complex genetic effects	125
4.6.3. Categorizing cases and controls	126

4.6.4.	<i>M.tb</i> strain diversity	127
4.7.	STUDY LIMITATIONS	128
4.7.1.	Whole exome sequencing	128
4.7.2.	Patients	130
4.7.3.	Functional studies	131
4.7.4.	Case-control association studies	132
4.7.4.1.	<i>Number of markers tested</i>	132
4.7.4.2.	<i>Phenotypic resolution</i>	133
4.7.4.3.	<i>Population stratification</i>	133
4.7.4.4.	<i>Statistical power and sample size</i>	134
4.8.	FUTURE STUDIES	135
4.9.	CONCLUDING REMARKS	136

#### 4.1. OVERVIEW OF STUDY FINDINGS

The present study aimed to identify novel TB susceptibility genes by finding MSMD-causing genes through whole exome sequencing. We hypothesized that, since one of the primary features of MSMD is an increased susceptibility to mycobacterial disease, MSMD-causative genes are good candidates to investigate for increased TB susceptibility in the general population. It is also speculated that a number of children presenting with severe TB, such as TBM, may have a monogenic predisposition to this disease<sup>281</sup>. For this reason, 10 genes previously associated with MSMD, as well as the putative MSMD-causing variants identified in the present study, were screened as potential TBM-causing genes in 10 TBM patients.

WES of three South African MSMD patients from different ethnic backgrounds identified three novel putative MSMD-causing mutations in two genes, namely *TAP1* and *MAP3K14*. Four variants situated in three genes (*IL12RB1*, *ISG15*, *TYK2*) were also identified in the TBM cohort as possible TBM-associated genes. *In silico* functional analysis was used to investigate the potential pathophysiological role of each variant. Additionally, functional studies were performed to determine the biological consequences of the *MAP3K14 V345M* and the *TAP1 I296M* mutations. *MAP3K14 V345M* significantly reduced the kinase activity of NIK, while it appeared to have no or limited effects on p100 breakdown, p52 production and autophagy.

*TAP1 I296M* appeared to cause reduced CD4<sup>+</sup> T cell counts, while increasing the CD8<sup>medium</sup> cell population, and thus the overall CD8<sup>+</sup> T cell count. Unfortunately, it was not possible to obtain PBMCs from patient PID\_060 and therefore, the functional significance of the *TAP1 P67S* mutation was not assessed in the present study.

Finally, a case-control association study was conducted to determine whether SNPs in *MAP3K14* and *TAP1* were associated with TB in the SAC population. However, using data from SAC TB patients and control individuals genotyped on the Illumina MEGA® array, no association between TB and these genes was found.

## 4.2. WHOLE EXOME SEQUENCING

Approximately one third of all patients displaying clear clinical signs of primary immunodeficiency present with groups of symptoms that have not been described previously<sup>330</sup>. It is thus difficult to predict the type and outcome of disease in these patients and provide them with appropriate treatment. A powerful tool for identifying the underlying genetic defects in at least some of these patients is high throughput next generation sequencing, particularly WES<sup>330</sup>.

WES is a cost-effective sequencing technique that has become the first-line approach to identify disease-causing mutations for monogenic disorders<sup>331,332</sup>. It is centred around the fact that protein coding gene variants are more likely to cause monogenic disorders than those situated outside the exome<sup>333,334</sup>. Advances in next generation sequencing techniques have led to the identification of more than 34 new gene defects associated with PIDs in the past 2 years, and have highlighted the remarkable variation in phenotypes of patients with mutations in the same gene<sup>335</sup>.

The next section describes how WES was used to identify novel MSMD-causing mutations in the present study.

### 4.2.1. MSMD patients

Three patients diagnosed with MSMD, as well as five of their healthy parents, were selected for WES to identify novel, rare, disease-causing mutations. After quality control, VCF file generation and variant calling, a total of 2 417 sequence variants were selected for further analysis.

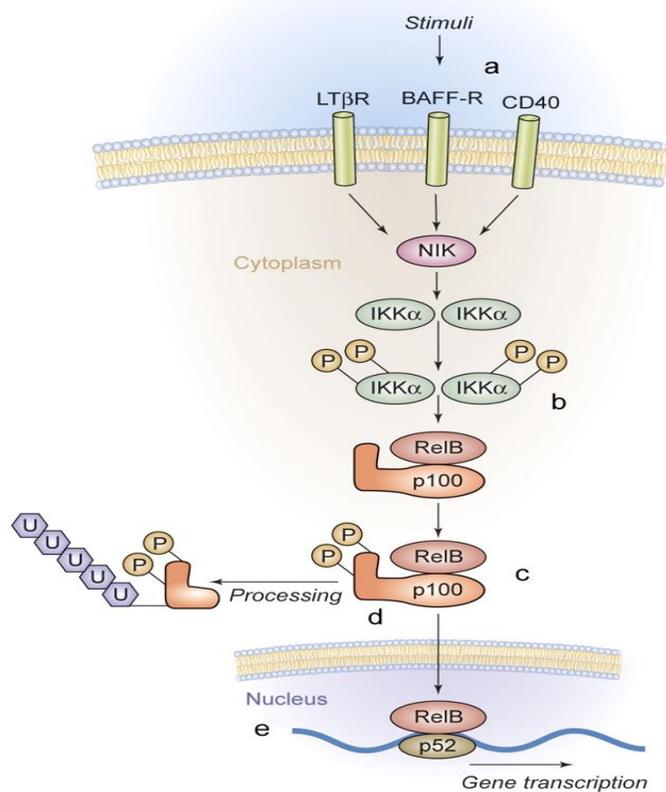
More than 600 bioinformatics tools exist for the analysis and interpretation of sequencing data, including those to assess the quality of the data and those to align the sequences to a reference genome<sup>336</sup>. The process of obtaining a single processed file containing all the potential variants of interest for further investigation has also been relatively standardized over time. Bioinformatics analysis of the sequencing data of the three MSMD patients allowed the identification of putative disease-causing variants. Only seven of these variants were located in functionally relevant genes according to OMIM and NCBI. Three of these were sequencing artefacts, while one was observed in control individuals and thus excluded from further analysis.

Three variants, namely *MAP3K14 V345M*, *TAP1 I296M* and *TAP1 P67S*, were selected for further analysis.

#### **4.2.1.1. MAP3K14: The NF- $\kappa$ B signaling pathway**

*MAP3K14* encodes the nuclear-factor-kappa-B inducing kinase (NIK) protein, which plays a crucial role in regulation of the non-canonical NF- $\kappa$ B signalling pathway involved in the transcription of several immunological genes<sup>337</sup>. It was therefore considered to be an excellent candidate gene for PID. The stimulation of tumour necrosis factor receptors (TNFR) such as lymphotoxin beta receptors (LT $\beta$ R), CD40 or B cell-activating factor belonging to the TNF family (BAFFR) activates NIK to phosphorylate and thus activate IKK $\alpha$ <sup>338</sup>. This leads to the phosphorylation of serine residues 866 and 870 of p100, which then undergoes ubiquitination and proteosomal degradation to form p52. RelB and p52 form transcriptionally active complexes which are translocated to the nucleus to activate the transcription of several downstream genes, which include a number of important immune system genes<sup>339</sup>. This pathway (Figure 4.1) is associated with generation of lymphoid organs, B-cell maturation and adaptive immunity. Defective signalling has been associated with multifaceted aberrant lymphoid immunity and thus primary immunodeficiency<sup>340</sup>.

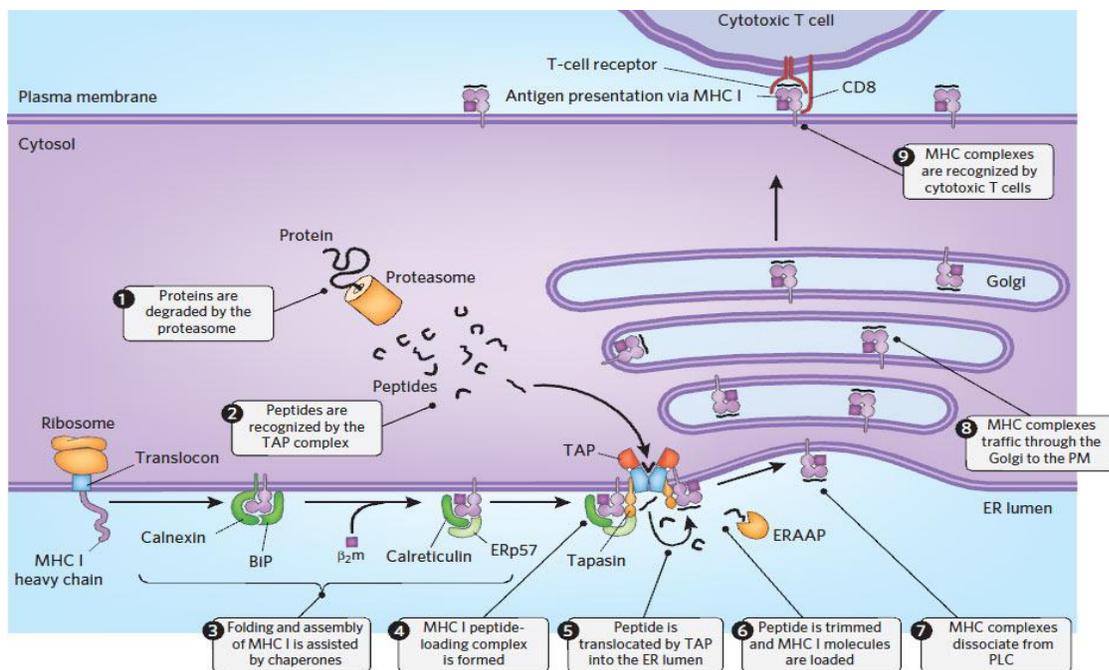
The mutation identified in *MAP3K14* in this study reduces the phosphorylation of IKK $\alpha$ , resulting in a decreased rate of p100 breakdown and subsequent p52 production. Less p52 is thus available to form transcriptionally active complexes with RelB, and the transcription rate of cytokine genes will therefore decrease.



**Figure 4.1: The non-canonical NF- $\kappa$ B pathway.** (a) This pathway is activated by a limited group of cell-surface receptors belonging to the TNF receptor superfamily. (b) Its main goal is to activate IKK $\alpha$ , (c) which directly phosphorylates NF- $\kappa$ B2/p100 to (d) induce proteolytic processing of p100 to p52. (e) p52 dimerizes with RelB to translocate to the nucleus and activate gene transcription<sup>341</sup>. LT $\beta$ R = lymphotoxin beta receptors; BAFF-R = B cell-activating factor belonging to the TNF family; CD = Cluster of differentiation; NIK = Nuclear factor kappa B-inducing kinase; IKK $\alpha$  = IkappaB kinase-alpha; P = Phosphate; U = Ubiquitin.

#### 4.2.1.2. TAP1: MHC class I activation

Transporter associated with antigen processing (TAP) is part of the ATP-binding cassette (ABC) transporters superfamily, which is responsible for the translocation of a broad range of solutes over biological membranes through the binding of ATP<sup>342</sup>. TAP is a heterodimer consisting of TAP1 and TAP2, and forms part of the major histocompatibility complex class I (MHC class I) peptide-loading complex<sup>343</sup>. TAP is important for effective activation of adaptive immune responses by transporting antigenic peptides from the cytosol to the lumen of the endoplasmic reticulum (ER). Here, these peptides form stable complexes with MHC class I molecules, which move to the surface of the cell to present the antigens to CD8<sup>+</sup>-cytotoxic T lymphocytes (Figure 4.2). These cells then recognize and eliminate the infection<sup>344</sup>.



**Figure 4.2: Function of TAP1 in antigen processing via MHC class I molecules.** CD = Cluster of differentiation; MHC = Major histocompatibility complex; PM = Plasma membrane; ER = Endoplasmic reticulum; TAP = Transporter associated with antigen processing; BiP = Immunoglobulin-binding protein;  $\beta_2m$  =  $\beta$ 2-microglobulin; ERAAP = Endoplasmic reticulum aminopeptidase associated with antigen processing; PLC = Peptide loading complex.

Mutations in *TAP1* lead to ineffective activation or functioning of the TAP heterodimer<sup>345</sup>. Antigenic peptides will thus not be transported effectively to the ER lumen to lead to the activation of MHC class I molecules. This will lead to a disruption in T cell activation and T cell responses. In the present study, patient PID\_012 presented with significantly increased CD8<sup>+</sup> T cell populations combined with decreased CD4<sup>+</sup> T cell populations compared to a healthy control. The states of CD8<sup>+</sup> activation observed in this patient also differed greatly from that observed in the control. One can speculate that these changes are caused by the identified mutation. However, this needs to be functionally validated in future studies.

#### 4.2.2. TBM cohort

WES identified four potentially disease-causing variants situated in three (*IL12RB1*, *ISG15*, *TYK2*) of the twelve genes investigated in the TBM cohort. Young children suffering from primary TB are especially at risk of developing TBM<sup>346</sup>. Any genes involved in increased

susceptibility to primary TB could thus potentially also be involved in the development of TBM. We hypothesized that variants present in MSMD-associated genes may be the cause of TBM.

#### 4.2.2.1. IL12RB1

Active TB in humans and mice leads to increased transcription of *IL12RB1*, which is crucial in protecting against nontuberculous as well as tuberculous mycobacterial disease<sup>347,348</sup>. A three-fold upregulation of *IL12RB1* transcript has been observed in post-mortem brain samples obtained from human TBM patients<sup>349</sup>. When *M.tb* infection causes active disease, pleural fluid mononuclear cells are activated to produce IL12Rβ1 – a receptor chain common to IL12 and IL23 receptors.

A lack of IL12Rβ1 leads to defective IL12 and IL23 signalling<sup>350–352</sup>. Diminished IL12 responses result in defective IFN-γ production<sup>353,354</sup>. IL12Rβ1-deficiency is associated with impaired Th17 cell development<sup>355</sup>. Several inflammatory and autoimmune diseases are associated with defective IL17 and IL23 production, such as Crohn's disease, rheumatoid arthritis, psoriasis, B cell malignancies<sup>356–359</sup>, and these cytokines are crucial for adaptive immunity against extracellular bacteria<sup>360</sup>. In individuals lacking *IL12RB1*, susceptibility to disseminated disease associated with nontuberculous mycobacteria<sup>353,354,361</sup> and the MTBC<sup>362–364</sup> is observed.

A secondary isoform of IL12Rβ1, known as IL12Rβ1ΔTM, was found to be produced by the mouse homolog *il12rb1*. This isoform is expressed after *in vitro* *M.tb* infection and increases immune responses associated with IL12p40<sup>365</sup>. It was speculated that IL12Rβ1ΔTM may be involved in restricting *M.tb* to the lungs, due to the fact that expression of IL12Rβ1ΔTM in the lungs corresponds with active replication and dissemination of *M.tb* to lymphoid organs<sup>365</sup>. A recent study found that IL12Rβ1ΔTM enhances control of TB in extra-pulmonary organs *in vitro*<sup>366</sup>. Mice incapable of producing IL12Rβ1ΔTM presented with increased *M.tb* burden in the mediastinal lymph nodes, liver and spleen compared to controls. This coincided with decreased production of TNF-α and IFN-γ by T<sub>H</sub> cells<sup>367</sup>. Should the *IL12RB1* T432M variant identified in TBM patient D5374 in this study affect the functioning or production of IL12Rβ1ΔTM, it could thus be responsible for the dissemination of *M.tb* to the patient's brain to cause TBM.

#### 4.2.2.2. ISG15

The role interferon stimulated gene 15 (*ISG15*) plays in *M.tb* infection is still poorly understood. However, a lack of ISG15 is associated with increased susceptibility to mycobacterial infections due to impaired IFN- $\gamma$  production by T lymphocytes and NK cells<sup>368</sup>. WES identified two bi-allelic mutations in *ISG15* leading to complete ISG15 deficiency and BCG disease in two distinct families from Turkish and Iranian decent<sup>330,368</sup>. Knockout of *isg15* in mice led to increased mortality rates following *M.tb* infection compared to *M.tb*-infected wild-type control mice, providing evidence for its crucial role in anti-mycobacterial disease<sup>369</sup>.

ISG15 is primarily released by steady-state neutrophils during bacterial and viral infections and acts in concert with IL12 to induce IFN expression<sup>370,371</sup>. Specifically type I IFN, namely IFN- $\alpha$  and IFN- $\beta$ , are strong up-regulators of ISG15, although IFN- $\gamma$  has also been observed to cause upregulation<sup>372</sup>. ISG15 is only expressed in vertebrates and is unconserved across different species, potentially indicating that *ISG15* adapted in response to specific conditions rather than being an essential housekeeping gene<sup>373</sup>.

ISG15 can conjugate to other proteins through iso-peptide bonds during a process known as ISGylation, which is similar to ubiquitination<sup>374</sup>, and has been observed in free as well as conjugated forms both intra- and extracellularly. A combination of unique enzymes are involved in ISGylation, namely USP18, an ISG15-specific protease; UBE1L (E1), an ISG15 activating enzyme; UBCH8 (E2), a conjugating enzyme; and protein ligases (E3 ligases)<sup>369</sup>. Both forms of ISG15 is increased during *M.tb* infection, hinting that both have a significant function in immunity against TB infection<sup>372</sup>. Cell-mediated killing, NK and T cell proliferation, as well as IFN- $\gamma$  secretion is stimulated by unbound ISG15<sup>370</sup>, which is also involved in the inhibition of IFN- $\alpha/\beta$ -dependent auto-inflammation<sup>375</sup>.

A recent study also observed increased type I IFN-inducible protein levels, including ISG15, in micro-particles (MPs) obtained from *M.tb*-infected human macrophages compared to uninfected macrophages<sup>372</sup>. This further promotes the theory that ISG15 plays a prominent role in the human host's innate immune response to *M.tb* infection, most probably influenced by the type I IFN signaling pathway.

MPs are small membranous particles released under normal conditions and are speculated to form a communication platform between host cells<sup>376</sup>. *M.tb* infection leads to an increase in the release of MPs to induce a pro-inflammatory response. Several signal transduction pathways have been associated with ISGylation of proteins<sup>374</sup>, and more than 300 proteins involved in a variety of functions are known ISGylation candidates. The influence of ISGylation could thus be extensive<sup>377–379</sup>. Although the crucial role of ISGylation and ISG15 has been shown in immune responses against viral infections<sup>380,381</sup>, its function in immune regulation in response to bacterial infections is less well characterized. The modulation of ISGylation has recently been proven to be critical in controlling *M.tb* and *Salmonella typhimurium* infections, although it is not involved in resistance against *Mycobacterium bovis* BCG<sup>382</sup>.

Although little is known about the host genetic factors associated with TBM, it can be speculated that loss-of-function mutations in *ISG15* can lead to diminished ISGylation of proteins involved in TBM, such as TLR2, TIRAP and LTA4H, and thus increased susceptibility to this disease. It will be interesting to further investigate the impact of defective ISGylation on these proteins and thus on TBM susceptibility.

#### 4.2.2.3. TYK2

Tyrosine kinase 2 (*TYK2*) encodes a non-receptor tyrosine kinase, which is part of the Janus kinase (JAK) family. JAKs become activated upon interaction of cytokines with their specific cell surface receptors<sup>383</sup>. *TYK2* promotes IL10 reactivation in an IFN- $\gamma$ -dependent fashion and facilitates IL10 signaling, both of which negatively control Th1 adaptive immunity<sup>383</sup>. Upon *M.tb* infection of the human host, DCs and macrophages control the infection by initiating an immune response. These functions can be suppressed by IL10, implying that *TYK2* could be involved in TB pathogenesis<sup>384</sup>. Mutations in *TYK2* are typically associated with hyper-IgE syndrome (HIES) accompanied by susceptibility to fungi, mycobacteria and viruses<sup>385–387</sup>.

A recent study used extensive bioinformatics methods to identify various pairs of genes linked to PTB, one of these pairs being Microtubule Associated Monooxygenase, Calponin and LIM Domain Containing 1 (*MICAL1*) and *TYK2*<sup>388</sup>. These two genes showed very high positive correlation, and were down-regulated in PTB cases versus controls. *MICAL1* is a cytoskeletal

regulator and is crucial in the processes of phagocytosis and cell migration<sup>389,390</sup>. Resistance to PTB depends on the correct functioning of these two processes, thus implicating *MICAL1* downregulation in PTB susceptibility<sup>388,391</sup>. The first patient identified with AR complete *TYK2*-deficiency was Japanese and presented with symptoms of HIES accompanied by several viral and intracellular bacterial infections, such as lymphadenitis caused by BCG vaccination<sup>387,392</sup>. Susceptibility to bacterial and viral infections were caused by defective *IFN- $\alpha/\beta$*  and *IL12* signaling, respectively. Impaired responses to *IL6* and *IL10* were also documented.

Moreover, seven additional, ethnically heterogeneous, *TYK2*-deficient patients were described, all of which were homozygous for one of five different null mutations<sup>392,393</sup>. All eight *TYK2*-deficient patients had diminished cellular responses to *IL23*, *IFN- $\alpha/\beta$* , *IL12* and *IL10*<sup>385</sup>. Regular levels of *IL17*<sup>+</sup> T cells observed in these patients explained the absence of mucocutaneous candidiasis. Reduction of *IL12* and *IFN* type I levels were responsible for increased susceptibility to viral and bacterial infections, respectively. Patients with a deficiency of *IL10* or its receptors (*IL10R1* and *IL10R2*) typically present clinically with acute intestinal phenotypes, like inflammatory bowel disease<sup>394-396</sup>. However, this was not observed in these *TYK2*-deficient patients with *IL10* deficiency, indicating the lack of clinical implications of *IL10*. Normal responses to *IFN- $\lambda$*  (*IL28/IL29*), *IFN- $\gamma$* , *IL21*, *IL27* and LIF (leukemia inhibitory factor) were observed in all patients.

However, in contrast to the first Japanese patient, the latter seven patients presented with bacterial and viral infections, but with none of the characteristic features of HIES<sup>386,397,398</sup>. This could potentially be due to normal fibroblast and leukocyte responses to *IL6* observed in these seven patients only<sup>385</sup>. Neither the silencing of defective *TYK2* nor the introduction of wild-type *TYK2* corrected the lack of sensitivity to *IL6*. This suggests that only increased susceptibility to viral and mycobacterial infections due to diminished *IFN- $\alpha/\beta$*  and *IL12* responses, and not HIES and impaired *IL6* responses, are inherent characteristics of *TYK2* deficiency<sup>385</sup>. One of these seven patients (P2) presented with *Brucella* meningitis and BCG disease<sup>393</sup>. Another (P3) died due to disseminated TB, while her brother (P4) presented with meningitis of which the cause was unknown. Two siblings, P5 and P6, suffered from BCG disease and P7 was diagnosed with miliary TB<sup>385</sup>.

It is evident that *TYK2* is involved in increased susceptibility to different forms of disseminated TB, BCG disease and possibly meningitis. It is thus very plausible that it could also be involved in TBM.

### 4.3. *IN SILICO* VARIANT PREDICTIONS

The majority of disease-causing mutations identified in humans are situated in the protein-coding region of the genome<sup>399</sup>. In order to design effective downstream experiments to aid in the development of more accurate diagnostic tools and more effective treatment strategies, information about these functional and structural effects are crucial<sup>304</sup>. Information about the function of proteins in biological systems, 3D protein structures, and information resulting from mutagenesis experiments predicting functional motifs, are especially important<sup>304</sup>.

Various online tools that use information from a variety of sources have been created to aid researchers in the challenging and time-consuming process of drawing conclusions with regards to pathogenicity of identified mutations<sup>303,306,399–410</sup>. Four of these servers/tools (HOPE, SIFT, MT2 and PP-2) were chosen for the analysis of the seven variants identified in the MSMD and TBM patients in the present study, based on their effectiveness in previous studies<sup>292,304,310,412</sup>. Three of these tools (SIFT, MT2, PP-2) made predictions with regards to the effect of each amino acid change, while the fourth (HOPE) further explored the functional consequences of the variants.

When comparing SIFT, MT2 and PP-2, MT2 was the most sensitive (88.7%) as well as the most accurate (88%), while the HumVar classifier of PP-2 had the highest specificity (88.7%)<sup>306</sup>. As seen in Table 3.5 and Table 3.6 in section 3.3, these three methods sometimes give conflicting scores for a single variant. Each tool has its own strengths and weaknesses, which is why predictability is greatly improved when using a combination of these tools<sup>292</sup> – as was done in the current study.

Some of the strengths of these three, and other, tools were used to produce HOPE: a next-generation online tool to analyze mutants automatically<sup>304</sup>. Similar to the previous tools, HOPE aims to describe the molecular source of disease phenotypes caused by mutated proteins. HOPE

gathers data from several online databases to predict the effects an amino acid change may have on protein function and its 3D structure. The results are explained with easily-interpretable 3D illustrations, animations and text in reports that can aid in downstream experimental designs<sup>304</sup>.

### 4.3.1. MSMD patients

#### 4.3.1.1. TAP1 I296M

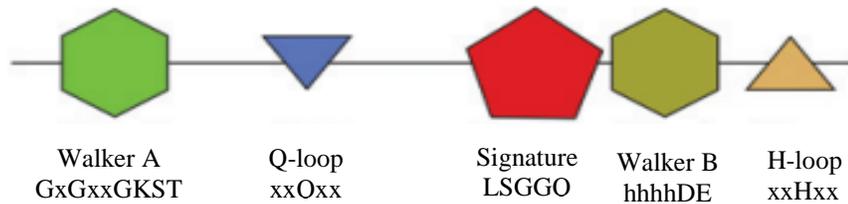
HOPE was employed to investigate the effect of *TAP1 I296M*, in order to determine whether it has the potential to be damaging to the protein. The exact 3D-structure of TAP1 is unknown, therefore HOPE modelled this based on a homologous structure. The PDB entry “4AYW” was identified as an effective modelling template, which represents the protein structure of the human mitochondrial ABC transporter known as ABCB10, situated in the mitochondrial inner membrane<sup>413</sup>.

With regards to contacts made by the variant, this residue was found to be strongly involved in disturbing multimeric contacts. This is due to the larger size of methionine (Met) when compared to the wild-type isoleucine (Ile).

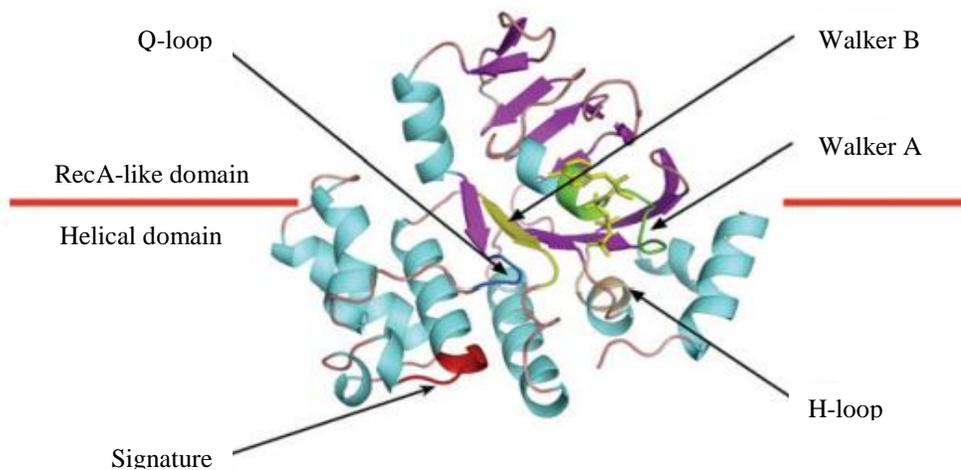
ABC systems are one of the biggest protein families. The ABC module, with the function of binding and hydrolyzing ATP, is highly conserved among these proteins. In order to be characterized as an ABC module, five short sequence motifs must be present and situated in a specific order (Figure 4.3): (1) a phosphate-binding loop known as the Walker A motif; (2) a highly conserved glutamine residue in the Q-loop; (3) the highly conserved signature motif (LSGGQ) unique to ABC ATPases; (4) a magnesium binding site known as Walker B; and (5) a highly conserved histidine residue within the H-loop, also known as the Switch region<sup>414,415</sup>.

Two structural domains, RecA-like domain and a helical domain, make up an ABC module (Figure 4.4). The Walker A and B motifs are situated in the RecA-like domain, while the helical domain contains the signature motif and is thus unique to ABC ATPases<sup>416</sup>. Two bendable loops, the Q-loop and the H-loop, join the domains. The highly conserved glutamine residue in the Q-loop facilitates interaction between the transmembrane subunits of ABC transporters and the ABC subunits<sup>417</sup>. ABC transporters typically have two integral membrane (IM) modules

consisting of four to eight transmembrane helices (TMHs) and two ABC modules. ABC modules are divided into Classes 1, 2 and 3 based on their main functions. Class I, which TAP belongs to, contains most of the exporters and is made up of systems with IM modules fused to ABC<sup>418</sup>.



**Figure 4.3: Conserved motifs in the ABC.** (A) Linear representation of the main motifs found in all ABC ATPases. Large polygonal symbols represent the Walker A and B motifs, which form the nucleotide-binding fold of the RecA-like ATPase family, and the signature motif, which is unique to ABC proteins and also interacts with ATP. Other motifs, including the Q-loop and the H-loop (a.k.a. the switch region), contain only one highly conserved residue and are represented by triangles. These residues make contacts with the  $\gamma$ -phosphate of ATP. The one-letter code of amino acids is used, x and h represents any and hydrophobic amino acids respectively.



**Figure 4.4: Domains of the ABC.** Localization of the conserved motif on the structure of an ABC monomer. The same color code as for the linear representation in Figure 4.3 is used<sup>418</sup>.

The ABC transporter associated with antigen processing (TAP) is a protein pump situated in the endoplasmic reticulum (ER) responsible for the transport of cytosolic peptides into the ER, where they are loaded onto MHC class I molecules for immune surveillance<sup>419</sup>. TAP is an exporter and consists of two homologous subunits, TAP1 and TAP2, each contributing one C-terminal nucleotide binding domain (NBD), one six-alpha-helix transmembrane domain (TMD)



exchange of an alanine (Ala) at position 95 with an aspartic acid (Asn) (A95D), was identified to have deleterious effects on adrenoleukodystrophy protein structure and function to cause X-linked adrenoleukodystrophy<sup>423</sup>. Mutations in transmembrane domains of ABC transporters can thus affect protein function and structure.

HOPE also found that only the wild-type Ile residue was observed at the position of interest, making it a conserved residue. The mutation of a 100% conserved residue is usually damaging for the protein. However, mutant Met has some properties in common with wild-type Ile: both are aliphatic, hydrophobic, neutral and nonpolar. This mutation might occur in some rare cases with little effect, but it is more likely to be damaging to the protein.

Even though both amino acids are hydrophobic, the mutated residue is situated on the surface of the protein. As mentioned, the replacement of Ile with the bigger Met could potentially diminish contacts made with other amino acids/proteins. This means that the activities associated with the TMD it is situated in, such as ATP and substrate binding, ATPase activity and transmembrane movement of substances, could be affected. The mutation could disturb interaction between different domains as well, affect signal transduction and thus influence the interaction of this protein with others, or interactions within the protein itself. This site is 100% conserved across mammalian species, which may suggest that a mutation at this position is not well tolerated.

Previous studies provided evidence that the substitution of an isoleucine with a methionine, either in rare missense mutations or in common polymorphisms, can be associated with disease. The novel missense mutation I26M in *DUOXA2* was shown to cause congenital hypo-thyroidism (CH) with goiter. The mutation affects the generation of hydrogen peroxide ( $H_2O_2$ ) without altering protein expression levels<sup>424</sup>. The SNP rs738409 C/G situated in the patatin-like phospholipase-3 (PNPLA3)/adiponutrin gene, which encodes the I148M protein variant, has also been associated with non-alcoholic fatty liver disease<sup>425-428</sup>. Three mutations (I312M, G545S, V316T) in mosaic penicillin-binding protein 2 (*PBP2*) were found to be responsible for reduced susceptibility to several antibiotic treatments in patients infected with *Neisseria gonorrhoeae*<sup>429</sup>.

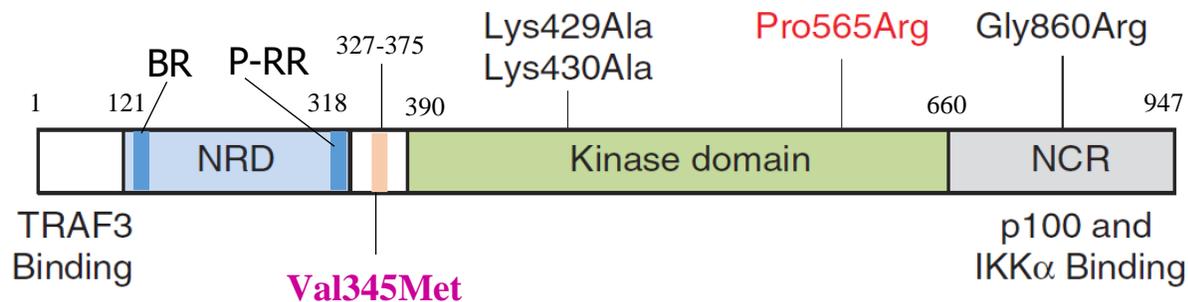
TAP is vital for normal immune functioning, and loss of TAP function/expression has been associated with some forms of cancer, as well as several viral and bacterial infections<sup>419,430-433</sup>. Variants located in the exons of TAP1 can influence antigen presentation specificity, and thus disturb the translocation of peptides<sup>434,435</sup>. Due to its involvement in immune responses, it is plausible that *TAP* gene polymorphisms can contribute to increased susceptibility to TB, although only a few studies have investigated this<sup>436-440</sup>. Associations have been observed between several *TAP1* SNPs and TB, and will be discussed in section 4.6.

#### **4.3.1.2. MAP3K14 V345M**

HOPE investigated the effect of *MAP3K14 V345M* as discussed above, to determine whether it has the potential to damage the protein. The 3D-structure of MAP3K14 is known, and its PDB entry “4G3D” was used for analysis. Protein annotations were obtained from the UniProt entry “M3K14\_HUMAN”.

The residue of interest is in contact with residues from other domains, and the introduction of a larger Met at this position, compared to the wild-type valine (Val), could potentially disrupt these interactions. The residue is situated at the core of MAP3K14, and the bigger Met may cause steric hindrance. This increases its chances of diminishing interactions between the domain it is situated in and others, as well as within the domain itself. The interactions might be important for the correct functioning of NIK.

Based on its amino acid sequence, four main domains were identified in NIK (Figure 4.6), namely the TNF receptor-associated factor 3 (TRAF3) binding domain, the negative regulatory domain (NRD), the serine/threonine kinase domain and the non-catalytic region (NCR)<sup>340</sup>. The NRD contains a basic region (BR), spanning amino acids 127-146, as well as a proline-rich repeat (P-RR) motif, situated between amino acids 250 and 317<sup>441</sup>. The NRD binds to the C-terminal of NIK to form a stable complex, and inhibits its binding to IKK $\alpha$ . When NIK becomes activated, this intramolecular interaction is lost and the binding sites on NCR are open to bind to IKK $\alpha$ . The NCR is responsible for interaction with important molecules of the signaling cascade, such as IKK $\alpha$ , p100, TRAF1, TRAF2, TRAF5, and TRAF6<sup>441-444</sup>.



**Figure 4.6: Schematic representation of NIK protein domain structures.** The NRD (light blue) contains the BR and the P-RR domains (darker blue). The conserved domain containing V345M (purple label) is indicated in pink and situated before the kinase domain (green). This is followed by the NCR (grey). The P565R amino acid change observed in P1 and P2 by Willmann et al. is situated in the kinase domain. Black labels indicate the mutations previously observed in the murine studies described (Lys429Ala; Lys430Ala; Gly860Arg)<sup>340</sup>. TRAF3 = Tumor necrosis factor receptor-associated factor 3; NRD = Negative regulatory domain; BR = basic region; P-RR = Proline-rich repeat; NCR = Non-catalytic region; Lys = Lysine; Ala = Alanine; Pro = Proline; Arg = Arginine; Gly = Glycine; Val = Valine; Met = Methionine; IKK $\alpha$  = IkappaB kinase-alpha.

The HOPE analysis indicated that V345M is located just before the interpro domain known as Mitogen-Activated Protein Kinase Kinase Kinase 14 (IPR017425), as seen in Figure 4.6, which has been associated with protein kinase as well as transferase activity. Because of its close proximity to the kinase domain, it is tempting to speculate that V345M may have an effect on the ability on NIK to phosphorylate IKK $\alpha$ . This is discussed in greater detail in section 4.4.1.

Val is a nonpolar, hydrophobic and aliphatic amino acid. The main difference between Val and Met is the presence of a C-beta branch in Val: two non-hydrogen substituents attached to its C-beta carbon instead of only one as for Met. Val is thus bulkier near the protein backbone and more restrictive to main-chain conformations. The Val side chain is very non-reactive, and rarely directly involved in protein function, although it can play a role in substrate recognition.

The substitution of a valine with a methionine, either in rare missense mutations or in common polymorphisms, have been associated with several diseased states. The rs4680 polymorphism, resulting in the exchange of a Val with a Met at position 158 of catechol-O-methyltransferase (*COMT V158M*), was associated with the modulation of functional connectivity as well as activity of brain regions in the working memory networks of especially older individuals. Methionine at this position caused a four-times decrease in dopamine-degrading activity

compared to valine<sup>445</sup>. A Dutch population-based study also found an association between a Val to Met mutation at position 973 of the human insulin receptor (*HIR V973M*) with hyperglycemia as well as the increased occurrence of non-insulin-dependent diabetes mellitus<sup>446</sup>. The single nucleotide missense substitution of Val at position 865 of human androgen receptor with Met have also been associated with complete androgen insensitivity<sup>447</sup>. The polymorphism V66M observed in the brain-derived neurotrophic factor (BDNF) gene has also been associated with changes in brain anatomy and memory<sup>448</sup>.

#### **4.3.1.3. TAP1 P67S**

The structure and function of *TAP1* have already been discussed in section 4.3.1.1. This gene is also further discussed in section 4.6.

Wild-type proline (Pro) is a cyclic, non-polar, medium-sized hydrophobic and aliphatic amino acid, the only one with a secondary amine group. It is unique in that the alpha-amino group is attached directly to the side chain, making the alpha carbon a direct substituent of the side chain. Mutant serine (Ser) is polar, more water-soluble and smaller than Pro. HOPE concluded that the rigid structure of proline induces a unique backbone conformation that might be required at this specific position, which is lost upon the introduction of Ser. The mutation can also result in loss of interactions due to the size reduction. Hydrophobic interactions either on the surface of the protein or at its core will also be lost when the hydrophobic Pro is replaced with water-soluble Ser.

This *TAP1* variant identified in our study (rs375389015) has previously been observed in a heterozygous state in non-Finnish European, South Asian and Zimbabwean individuals. It is predicted to be a benign polymorphism and has a MAF of  $3.897 \times 10^{-5}$  according to ExAC Browser. However, this variant is observed in a homozygous state in PID\_060 and could thus be associated with more severe effects.

Mutations in *NOTCH3* are associated with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), which is a hereditary small vessel disease of the brain. A patient presenting with the homozygous *NOTCH3 R544C* mutation had a

more severe phenotype compared to her heterozygous *NOTCH3 P544C* sister<sup>449</sup>. The auto-inflammatory disorder Familial Mediterranean Fever (FMF) caused by mutations in the pyrin/marenostrin (*MEFV*) gene is associated with over-activation of neutrophils. The pro-inflammatory danger signal S100A12, derived from neutrophils, is highly increased in FMF. Increased S100A12 serum levels were observed in patients with the *M694V MEFV* mutation, with the highest level observed in homozygous *M694V MEFV* patients, followed by compound heterozygous and then heterozygous *M694V* patients<sup>450</sup>.

Homozygous forms of mutations, where both alleles are mutated, can have more severe effects on heterozygous patients with only one mutated allele. The type IIa form of Usher syndrome, an inherited disorder with the main feature of hearing impairment (HI), results from mutations in *USH2A*. The most common mutation in this gene leads to a frameshift at codon 767 to cause a premature stop-codon (Glu767Serfs\*21). Patients homozygous for this mutation, with two truncating mutations, presented with more severe HI than individuals with one truncating and one non-truncating, or those with no truncating mutations<sup>451</sup>.

The amino acid exchange of a proline with a serine can be associated with increased susceptibility to several diseases. A common polymorphism in Hypoxia-inducible factor 1 $\alpha$  (*HIF-1 $\alpha$* ), namely 1772C>T (rs 11549465), results in the substitution of proline at position 582 with a serine (P582S). The TT genotype is significantly increased in patients with breast cancer compared to controls, serving as a marker for the development of breast cancer<sup>452</sup>. The missense variation C > T at the first base of codon 143 of transforming growth  $\beta$ -induced factor 1 (*TGIF1*), rs4468717, leads to the substituting of Pro with Ser. This variant has been associated with the complex disorder high myopia, which has severe effects on ocular health, in a northern Indian population. *In silico* variant predictions also showed the negative impact this variant may have on protein function and structure<sup>453</sup>. A functional polymorphism in NAD(P)H dehydrogenase, quinone 1 (*NQO1*) has been associated with increased risk of developing leukemia. The 609C>T substitution also leads to a Pro/Ser exchange, and the homozygous TT genotype correlated with leukemia risk in a meta-analysis comparing 3 634 cases and 4 827 controls<sup>454</sup>.

### 4.3.2. TBM patients

#### 4.3.2.1. ISG15 L54P

HOPE investigated the effect of *ISG15 L54P*, as discussed above, to determine its effect on protein structure and function. The 3D-structure of ISG15 is known and its PDB entry “3SDL” was used for analysis. The UniProt entry “ISG15\_HUMAN” was used for protein annotations.

ISG15 is about 15kDa in size and consists of 158 amino acids<sup>373</sup>. It belongs to the ubiquitin-like (UBL) protein family and catalyzes proteins during ISGylation by conjugating to lysine residues of substrates through isopeptide bonds. This requires E1-activating and E2-conjugating enzymes combined with E3 ligases<sup>374</sup>. Two UBL domains (N-terminal and C-terminal) connected with a short linker (amino acids 76 – 79) make up ISG15, and both are needed for effective binding of ISG15 to its targets. The N-terminal domain does not take part in ISG activation by the E1 enzyme or ISG15 linkage to the E2 enzyme, but appears to regulate the transfer of ISG15 to its substrates, which is mediated by E3 ligases<sup>455</sup>. It is thus presumed that ISG15 E3 ligases are in some way associated with the N-terminal UBL domain’s regulatory role. A conserved motif (LRLRGG) is situated at the C-terminal UBL and covalently connects to target proteins. Highly conserved, species-specific residues have also been documented in the linker region<sup>456</sup>.

HOPE concluded that *ISG15 L54P* is situated within an ubiquitin-like 1 domain with the function of protein binding. This domain is important for interaction with other molecules or proteins, and the residue of interest is in contact with residues from other domains. It is situated in the core of the protein and, because the mutant Pro is smaller than the wild-type Leu, the variant is likely to influence protein folding. Contacts with this residue will thus be disturbed or lost completely, ultimately affecting the protein’s function. Both amino acids are non-polar and aliphatic. The rigidity of Pro might influence the protein’s binding and interacting capacity.

Single amino acid exchanges in the UBL domains of ISG15 and related proteins can be associated with various diseases. The mutation L10A in the N-terminal UBL domain of ISG15 have been observed to significantly decrease the protein’s binding affinity to Influenza B Virus NS1B. The single mutations D76Q and K77N situated in the short linker between the two UBL domains had the same effect, but to a lesser extent<sup>457</sup>. The E3 ligase Parkin, associated with Parkinson’s disease (PD), also holds an UBL domain in its N-terminal, and an R1 domain in its

C-terminal. By interacting with the R1 domain, UBL suppresses Parkin auto-ubiquitination, mitophagy, mitochondria translocation and substrate ubiquitination. Mutations in the UBL (F4A, R6A, P14A and E16A) abolishes interactions with the R1 domain and increases PD pathogenesis<sup>458</sup>. The R22Q mutation situated in the UBL of Parkin was also associated with a 40% decrease in ISGylation of this gene<sup>459</sup>.

Substitutions of leucine with proline can result in disease in humans. Transmembrane protein coding genes are often related to diseases in humans. Most mutations in these genes that are associated with disease result either in a leucine/proline or a glycine/arginine exchange<sup>460</sup>. Myeloid differentiation primary response gene 88 (MYD88) is a common adaptor protein forming part of the innate immune system. The L265P amino acid substitution in MYD88 causes constitutive activation of the protein to result in amplified pro-inflammatory signals intracellularly. Vitreoretinal lymphoma (VRL) has been associated with this mutation<sup>461</sup>. The hereditary peripheral neuropathy CMT1A, the most common form of Charcot-Marie-Tooth (CMT) disease, is caused by mutations in the peripheral myelin protein 22 (PMP22) gene. Its features include muscle weakness, peripheral nerve demyelination and nerve conduction velocity reduction. The *PMP22 L16P* mutation have been shown to underlie CMT1A<sup>462</sup>.

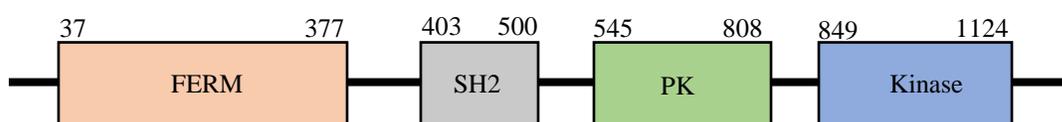
#### 4.3.2.2. TYK2 C378F

The effect of *TYK2 C378F* was also determined by HOPE, as described above. The 3D structure of TYK2 is known and the PDB entry for it, “4PO6”, was used for analysis. The UniProt entry “TYK2\_HUMAN” was used for protein annotations.

Cys is a Sulphur-containing, polar, hydrophobic amino acid that plays a key role in stabilizing extracellular proteins. The hydrophobic, inert nature of the benzyl side chain makes Phe neutral and nonpolar. Phe is bigger than Cys and is situated near a highly conserved position. Cys is buried in the core of the protein and the larger Phe might not fit, thus altering the structure and potentially also the function of TYK2.

TYK2 belongs to the JAK-family tyrosine kinases, which is involved in transmitting extracellular signals from various cytokine receptors and regulating diverse cellular processes.

These include hematopoiesis, differentiation, growth and immune responses<sup>463</sup>. JAK family members bind to cytokine receptors to activate JAK kinases, resulting in phosphorylation and activation of STAT transcription factors<sup>464</sup>. Immune functions can become compromised when the JAK pathway is inhibited, while constant activation of JAK kinases may cause cancer and myeloproliferative disease<sup>465</sup>. Four structural domains are contained in JAK proteins: an N-terminal Band 4.1, Ezrin, Radixin, Moesin-homology (FERM) domain, followed by an SH2 domain and a pseudo-kinase (PK) domain, and ending with a C-terminal catalytic (kinase) domain. These are shown in Figure 4.7.



**Figure 4.7: The domain structure of Jak-family kinases.** N-terminal FERM domain is indicated in pink, followed by the SH2 domain in grey, the PK domain in green and the C-terminal Kinase domain in blue. FERM = Band 4.1, Ezrin, Radixin, Moesin-homology; PK = Pseudo-kinase.

Interaction between JAK kinases and the cytoplasmic tails of their receptors require the FERM and SH2 domains, and the SH2 domain is also involved in interaction with receptors and activation of JAK<sup>466</sup>. The PK domain mainly serves a regulatory role through interactions with the neighboring tyrosine kinase domain<sup>467,468</sup>.

The FERM and SH2 domains of TYK2 form a tightly joined structural unit referred to as the FERM-SH2 domain<sup>469</sup>, in which *TYK2 C378F* is situated. TYK2 FERM consists of three subdomains: F1 (residues 26–107), F2 (residues 145–271) and F3 (residues 272–430). F1 contains a UBL domain trailed by an insertion of 37 residues (L1) that interacts with the C-terminal of the SH2 domain. F2 is completely helical and has a basic patch on the  $\alpha 2'$ ,  $\alpha 2''$  and  $\alpha 3$  helices. Fifteen surface-exposed residues are contained in these helices. Two big, disordered loops are contained in F3 (residues 296–308 and 330–370)<sup>469</sup>. SH2- flanking linkers (L2 and L3) as well as L1 keep SH2 in place next to F1 of the FERM domain. The core of SH2 folds into a three-stranded antiparallel  $\beta$ -sheet ( $\beta B$ ,  $\beta C$ , and  $\beta D$ ) with two intermittent short  $\alpha$ -helices ( $\alpha A$  and  $\alpha B$ ). The FERM, SH2 and linker sequences are overall conserved across the JAK family members<sup>469</sup>.

FERM-SH2 structure in TYK2 closely resembles that of JAK2<sup>463</sup>. Mutations documented to alter receptor binding in members of the JAK family almost all fall within the hydrophobic core of the FERM domain. Most of them are buried hydrophobic or charged residues forming hidden salt bridge interactions, and would be expected to cause loss of receptor binding and loss of function due to compromised structure of the FERM domain. These mutations include Y102A and F103A in *TYK2*<sup>470</sup>, and the *JAK1* mutants L158A/F159A<sup>471,472</sup>, L166A/ I167A<sup>473,474</sup>, Y270A<sup>475</sup> and Y281A<sup>476</sup> to name but a few. Although *TYK2 C378F* is not situated within the FERM domain, it is directly adjacent to it, and will most likely also influence its structure and function, due to the change in size and charge introduced by the mutation.

Ouabain is a cardiac glycoside that binds to the alpha subunit of Na/K-ATPase to inhibit its activity. The C113F mutation of this alpha subunit increased ouabain's inhibition constant more than 1000 times<sup>477</sup>. Retinoblastoma (*RB*) encodes the RB protein that is present in most vertebrate cells<sup>478</sup>. Cells that are completely differentiated or resting express the under-phosphorylated form of this protein, while cells that are proliferating express the hyper-phosphorylated form<sup>479</sup>. Just under-phosphorylated RB are capable of binding the T antigen SV40, which suggests that *RB* contains a mutation in these resting/differentiated cells. The point mutation G>T was identified in exon 21 of this gene to cause the Cys to Phe substitution at position 176 (C176F)<sup>480</sup>.

The substitution of cysteine with a phenylalanine can have structural and functional consequences for proteins. Hormonal responses to androgens during the different stages of development are mediated by the ligand-transcription factor known as the androgen receptor (AnR)<sup>481</sup>. The testicular feminization syndrome, characterized by genotypic males developing the sex characteristics of females, is caused by abnormal AnR<sup>482</sup>. The substitution of non-conserved Cys at position 560 in the binding domain of human AnR have critical effects on the stability of DNA-receptor complexes. Substitution with Phe increased trans-activating efficiency by 150%, while three other amino acid substitutions led to a 40-70% decrease<sup>483</sup>.

#### 4.3.2.3. TYK2 A53T

The effect of *TYK2 A53T* on protein function and structure was also determined by HOPE, as described above.

Thr is larger and less hydrophobic than Ala. Ala is buried in the core of the protein and since Thr is much larger than Ala, this substitution is likely to cause steric hindrance resulting in a conformational change. This mutation may thus cause loss of hydrophobic interactions in the core of the protein, and probably also affect the structure of the protein. The structure and function of *TYK2* have been discussed in section 4.3.2.2.

*TYK2 A53T* is situated within the FERM domain, which is highly conserved. This domain has the main functions of transferase activity and substrate binding. The structure and function of this domain, as well as the effect of mutations identified in it, have previously been discussed.

A homozygous deletion of GCTT nucleotides at positions 550-553 in *TYK2* resulted in a frame-shift mutation to cause a premature stop codon at position 90 of *TYK2* in a single patient. Effects of this mutation included multiple defective cytokine signaling pathways, impairment of Th1 differentiation and acceleration of Th2 differentiation. Patient symptoms included unusual susceptibility to mycobacteria, fungi and virus and atopic dermatitis with increased IgE serum levels<sup>387</sup>. Mutations in the FERM domain is associated with diminished capacity of binding to cytokine receptors and improper catalytic regulation<sup>484</sup>. Most of the residues identified to be mutated in the FERM domain is buried in the core of the protein, which means that they compromise the structure of and thus destabilize the FERM domain. Mutations identified in the FERM domains of other JAK family members to compromise FERM structure and interfere with cytokine binding, include *JAK1 L80A/Y81A*<sup>464</sup>, *JAK3 W81A*<sup>465</sup>, *JAK3 P83A/P84A/S85A/H86A*<sup>484</sup>, and *JAK3 L98A/L99A/I102A*<sup>485</sup> to name a few<sup>463</sup>.

Comparison of two different outbreaks of the H7 subtype influenza virus capable of infecting humans provide a good example of an alanine to threonine variant causing disease. The H7N7 outbreak took place in the Netherlands, while the H7N9 outbreak occurred in China. Murine monoclonal antibodies (mAbs) recognize the A antigenic site of H7 hemagglutinin (HA). These

mAbs were shown to deactivate pseudo-viruses containing HA from the H7N9 outbreak, but not that of the H7N7 outbreak. These differences are caused by the presence of a Thr at position 143 in the HA of H7N7 compared to the Ala observed at this position of the H7N9<sup>486</sup>.

The main features of the degenerative disorder Parkinson's disease are impaired movement and resting tremor. Alpha-synuclein (ASN) A53T transgenic rats that expresses the abnormal human mutant ASNgene showed abnormal degradation of striatal neurons, increased ASN aggregation as well as a drastic impairment in general motor activities correlating with increased age<sup>487</sup>.

The A54T polymorphism identified in the intestinal fatty acid binding protein (FABP2) has been shown to correlate with dyslipidemia and increased blood glucose and insulin levels in Mexican-American Type 2 diabetes patients<sup>488</sup>. The substitution of an alanine with a threonine has thus been associated with several diseased states.

#### **4.3.2.4. IL12RB1 T432M**

HOPE was used to investigate the effect of *IL12RB1 T432M* to determine whether it has the potential to be damaging to the protein. No structural information was available for this protein and no appropriate 3D structure or modelling template could be found. HOPE thus used UniProt annotations of "I12R1\_HUMAN" and predictions for analysis of the mutation.

The glycosylated type 1 transmembrane protein IL12Rβ1 is encoded by Interleukin-12 receptor β1 (IL12RB1) and binds to the IL12p40-domain of IL12 and IL23 to regulate their sensitivity<sup>489</sup>. *IL12RB1* polymorphisms are associated with a wide variety of diseases regulated by IL12 and IL23, such as nontuberculous mycobacterial infections, pediatric asthma, cancer, atopic dermatitis and TB<sup>362,489,490</sup>. Human IL12RB1 is 70.5 kDa in size and consists of five extracellular, one transmembrane and one cytoplasmic domains. It has two extracellular cytokine binding domains (CBD), D2 and D3, and three fibronectin-type III domains<sup>491</sup>. The cytokine-binding region is important for binding to IL12 and IL23, while the cytoplasmic region works with IL23R and IL12Rβ2 to transmit signals through the JAK2 and TYK2 kinases<sup>492</sup>. HOPE concluded that T432M is situated in the Fibronectin type-III 4 domain of IL12RB1 at an exon-intron border.

Fibronectin (Fn) is a bulky protein present in the extracellular matrix with the function of cell adhesion. It consists of repeated modules types I, II and III, each containing a hydrophobic core surrounded by two anti-parallel  $\beta$ -sheets<sup>493</sup>. The Fn-III domain is a common extracellular protein structural motif and usually acts as a structural spacer. Fn-I and -II domains contain disulphide bonds, which are absent in Fn-III domains, making them more elastic. This elasticity may be crucial for adhesion mechanisms and may allow for larger exploring areas to search for binding substrates<sup>494</sup>.

Met is bigger than Thr and also more hydrophobic. This mutation can thus result in loss of hydrogen bonds, and/or disturb correct protein folding. If the variant is situated at the hydrophobic core of one of the repeated modules, the protein will not be able to fold correctly, which may affect its proper functioning. This might result in loss of IL12 and IL23 sensitivity.

The dominantly inherited muscle disease known as hereditary myopathy with early respiratory failure (HMERF) is characterized in its early phase by respiratory weakness and muscle failure. It has been associated with miss-folding of the fibronectin III (FN3) 119 subdomain in titin<sup>495</sup>. Titin is the largest known protein and consists mostly of immunoglobulin-like and repeating fibronectin III modules. Five amino acid alterations have been shown to cause disease (P30068R, C30071R, W30088C, W30088R, P30091L), all of which diminished the solubility of the FN3 119 domain. *In silico* analysis showed that all of the disease-associated mutations resulted in inefficient folding of the FN3 119 domain<sup>495</sup>.

Although not much is known about the structure of *IL12RB1*, it is possible, based on the mentioned studies, that the T432M variant situated in the fibronectin III domain of this gene can alter the structure and function of this domain to interfere with IL12RB1 functions. If this mutation alters the receptor's sensitivity for IL12 and/or IL23, either one or both of these signaling pathways can be diminished. This will result in increased susceptibility to several diseases, of which TB is one.

The protein encoded by X-ray cross-complementing group 3 (*XRCC3*) is involved in homologous recombination repair (HRR) of DNA double-strand break (DSB). A recent meta-

analysis<sup>496</sup>, as well as a few previous studies<sup>497,498</sup>, highlighted a significant association between the risk of developing thyroid cancer and the polymorphism *XRCC3 T241M*. The great majority of lung cancer cases result from non-small-cell lung cancer (NSCLC), with mutations affecting the activation of epidermal growth factor receptor (EGFR), a common feature in this disease. Molecular targeting with the tyrosine kinase inhibitor for this receptor, known as Gefitinib, can be an effective initial treatment. Mutations in the gene that encodes the EGF receptor, such as T790M, however, cause resistance to this drug<sup>499</sup>. Although the substitution of threonine with methionine is not well documented, the mentioned studies prove that it can be disease-causing.

#### **4.4. FUNCTIONAL STUDIES**

##### **4.4.1. Effect of *MAP3K14 V345M* on the function of NIK**

Functional studies aimed to confirm the involvement of *MAP3K14 V345M* identified in PID\_040 in disease-causality. This patient was initially diagnosed with humoral immunodeficiency due to severe hypogammaglobinemia, which refers to a predisposition to infections normally protected against by antibody responses<sup>500</sup>. An MSMD diagnosis was later considered when PID\_040 developed a BCG abscess in the leg and subsequently BCG meningitis due to dissemination after vaccination. In 2014, the patient presented with severely depleted levels of CD45, CD8, B- and NK cells. T cell and CD4 cell counts were only just within normal ranges. Persistently low levels of class-switch memory B-cells (CD27+IgD-) led to chronically low Ig serum levels.

Of all the variants prioritized as potential disease-causing for this patient, the *MAP3K14 V345M* was the most likely. In *Map3k14* mutant mouse studies (where *Map3k14* [Nik] was either mutated or knocked-out) splenic and thymic structures, lymph nodes and Peyer's patches were completely disordered, which led to B-cell deficiency<sup>501-503</sup>. In these mice, Ig serum and B-cell levels were severely decreased due to compromised class-switch recombination (CSR) and somatic hypermutation (SHM) to cause humoral immunodeficiency<sup>504</sup>. These observations in mice closely resemble the phenotype of PID\_040.

Furthermore, Willmann and colleagues recently described a *MAP3K14 P565R* mutation in two patients, P1 and P2, who were first-degree cousins both from consanguineous parents<sup>340</sup>. P1 and P2 closely resembled the phenotypes described in the *MAP3K4* mutant mouse model which

included decreased marginal zone and memory B cells, B-cell lymphopenia, hypogammaglobulinemia and compromised SHM and CSR. P1 suffered from granulomatous hepatitis and tuberculosis osteomyelitis due to dissemination after BCG vaccination. The phenotype of PID\_040 thus closely resembles that of the two above-mentioned patients, as well as what was observed in murine studies, making it highly likely that *MAP3K14 V345M* is the disease-causing mutation in PID\_040. For this reason, functional investigation of the mutation was undertaken.

Due to logistical complications, no PBMCs or blood samples from PID\_040 were available, which limited the amount of functional studies conducted. Therefore, an overexpression model using wild-type and mutant plasmid constructs each co-transfected separately with an IKK $\alpha$ -expressing plasmid into the immortalized HEK293T cell line was used to investigate the functional effect of the mutation (sections 3.4.2.1 and 3.4.2.2). Using this model, the ability of the mutant construct to phosphorylate IKK $\alpha$  was compared to that of the wild-type construct as was done in the study by Willmann and colleagues.

While the P565R bi-allelic mutation identified in the study by Willman and co-workers is situated within the kinase domain of NIK<sup>340</sup> (amino acids 390-660), the V345M mutation identified in the present study is located just before the kinase domain. Because of its close proximity to the kinase domain, it was speculated that V345M may still have an effect on the ability of NIK to phosphorylate IKK $\alpha$ . This was indeed the case as levels of phosphorylated IKK $\alpha$  was significantly reduced in the mutant overexpression cells compared to the cells expressing the wild-type NIK construct. It should be noted that in the study by Willman and colleagues, phosphorylation of IKK $\alpha$  was completely absent in the mutant expression cell line. Thus, *MAP3K14 V345M* is also a loss-of-function mutation, but with less severe effects compared to *MAP3K14 P565R*.

Constructs that encode NIK starting at residue 377 do not produce soluble protein, while fragments starting at 327 produce soluble enzymes<sup>505</sup>. Although not part of the kinase domain, residues 327 to 375 in NIK are conserved across species. This means that mutations within this region are likely to be associated with disease. It also implies that the V345M mutation identified in this conserved area could alter the production of enzymatically active NIK.

Mutations outside the kinase domain of various proteins have been shown to affect their kinase activity. Mutations located in the C1B subdomain of the neuronal-specific protein kinase C gamma (PKC $\gamma$ ) gene was associated with spinocerebellar ataxia type 14 (SCA14). This subdomain is part of the regulatory domain of PKC $\gamma$  and the mutations reduced the kinase activity of this gene<sup>506</sup>. The protein tyrosine kinase Csk contains an N-terminal regulatory region that is made up of SH2 and SH3 domains. Site-specific mutations or deletions were induced in this region to determine its effect on the ability of Csk to phosphorylate its substrate, Src. Kinase activity was reduced by 70% and 96%, respectively, upon deletion of the SH3 and the SH2 domains<sup>507</sup>. Another study showed that a point mutation (I145P) in the first coiled-coil domain of human *c-fes*, which encodes the non-receptor protein tyrosine kinase Fes, drastically increased Fes tyrosine kinase activity, while combining this mutation with another related point mutation (L334P) in the second coiled-coil motif reduced the kinase activity by about 50%<sup>508</sup>. Fes is strongly expressed in myeloid hematopoietic cells and is involved in signal transduction pathways for various hematopoietic cytokines (such as IL3, IL4, IL6)<sup>509</sup>. These studies provide evidence that mutations outside of the kinase domain of various proteins can affect their kinase activity. It is thus plausible that *MAP3K14 V345M* can affect the kinase activity of NIK without being situated in the kinase domain.

#### **4.4.2. Downstream effects of *MAP3K14 V345M***

As mentioned in section 4.2.1.1, activated NIK phosphorylates IKK $\alpha$  to in turn phosphorylate p100. Subsequent proteosomal degradation of p100 forms p52, which combines with RelB in transcriptionally active complexes and move to the nucleus to activate the transcription of several downstream genes. Because p-IKK $\alpha$  is needed for proteolytic processing of p100 to form p52, the present study thus also investigated the effect of *MAP3K14 V345M* on the levels of p100 and p52.

The study by Willmann and others investigated patient-derived Epstein-Barr virus-immortalized lymphoblastoid cell lines (B-LCL) and observed normal NIK levels, while IKK $\alpha$  and p100 levels were increased. P52 levels were severely decreased and led to reduced RelB and p52 levels in the nucleus of patient cells<sup>340</sup>. The increased IKK $\alpha$  and p100 levels could indicate NF- $\kappa$ B pathway activation by viral proteins<sup>510</sup>. The *MAP3K14 P565R* mutation identified completely

inhibited NIK from phosphorylating IKK $\alpha$ , thus resulting in inhibition of p100 phosphorylation and proteosomal degradation to produce p52. Following transferal of wild-type *MAP3K14* into patient fibroblasts, the levels of p52 in the nucleus reached wild-type levels, suggesting that functional NIK is essential for p100 activation<sup>340</sup>.

As mentioned previously, we were unable to obtain any patient-derived cellular material to conduct functional analysis of the *MAP3K14 V345M* mutation. For this reason, the levels of p52 could not be assessed, as was done in the study by Willman and colleagues. An attempt was made to use the over expression model to determine whether p100 and p52 levels were affected by *MAP3K14 V345M*, however, no quantifiable levels of p52 in either pWZL-Neo-Myr-Flag-MAP3K14 or pWZL-Neo-Myr-Flag-MAP3K14-V345M transfected HEK293 cells (Figure 3.19 A; section 3.4.2.4) were detected.

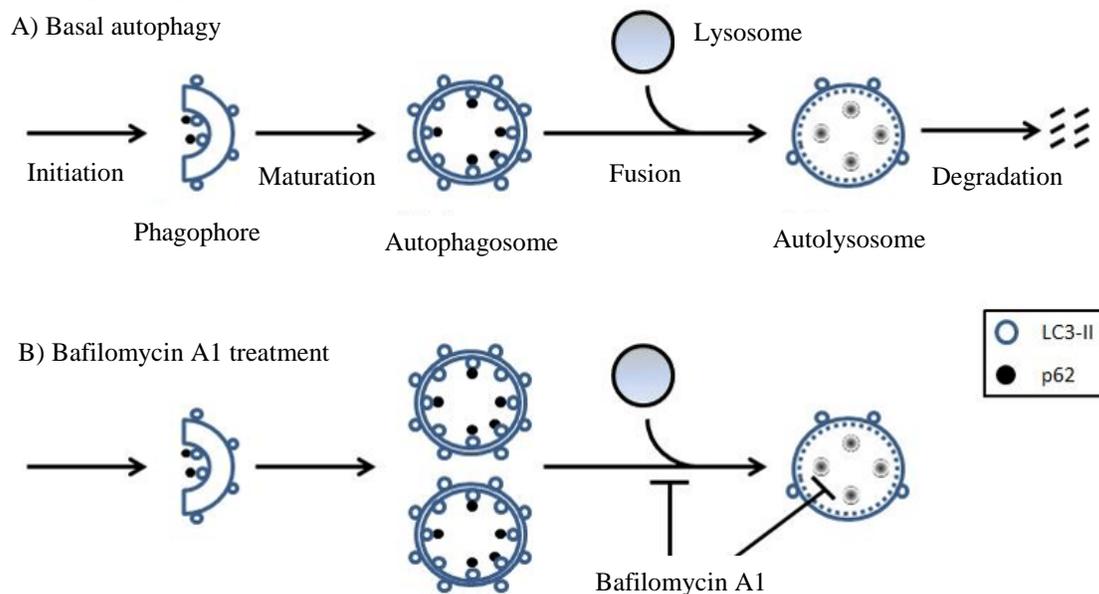
Significant p100 levels were detected in untransfected HEK293T cells as well as in cells transfected with pWZL-Neo-Myr-Flag-MAP3K14 and pWZL-Neo-Myr-Flag-MAP3K14-V345M (Figure 3.19 A and B; section 3.4.2.4). However, there was no discernible differences in p100 levels between these groups. This is most probably due to the fact that HEK293 cells were used instead of patient-derived immune cells. HEK293 cells are not capable of eliciting an immune response upon activation, which means that only baseline p52 and p100 levels under unstimulated conditions could be measured. Ideally, the difference in P100 and p52 in wildtype versus mutant cells should have been measured using patient and control derived cells.

#### **4.4.3. Effect of *MAP3K14 V345M* on autophagy**

The interpretation of LC3 immunoblots can be problematic and misleading due to the degradation of LC3-II by autophagy<sup>329</sup>. Treatment of cells with lysosomal protease inhibitors such as BAFB partly inhibits the breakdown of LC3-II due to its inhibition of autophagosome-lysosome fusion, while LC3-I levels remain unaffected<sup>511,512</sup>. The measurement of autophagic degradation activity is defined as autophagic flux, and is characterized by the steady-state volume of autophagosomes, their turnover as well as their half-life<sup>322</sup>. The lower/higher the degradation activity is, the lower/higher the respective rate of degradation, and thus autophagic flux, will be. Autophagic flux is most accurately measured by combining several techniques, like

western blot analysis of major proteins, fluorescence microscopy and transmission electron microscopy<sup>323</sup>. Each of these approaches either directly or indirectly measure the amount of autophagosomes, in the presence or absence of protease inhibitors. Autophagic flux is thus presumed to occur effectively when LC3-II levels increase in the presence of a lysosomal degradation inhibitor (like BAF), because LC3-II breakdown is inhibited<sup>513</sup>. Autophagic flux is thus not indicated by the amount of LC3-II at a given time. Autophagy induction leads to increased formation of autophagosomes, and thus an increase in total LC3-II levels. An increase in LC3-II levels thus indicate induction of autophagy.

The most effective method currently used to measure autophagy is the comparison of LC3-II levels before and after BAF treatment between different samples<sup>327</sup>. The process of autophagy and the influence of BAF treatment on it is shown in Figure 4.8. Autophagy is associated with innate and adaptive immune responses as well as inflammation regulation<sup>514,515</sup>. The dysregulation of autophagy has been associated with a variety of diseases, which include neurodegenerative, metabolic and infectious diseases, as well as cancer<sup>516-522</sup>.



**Figure 4.8: Basal autophagy and the effect of BAF treatment.** Autophagic flux/turnover is represented by the difference between LC3-II marker levels before and after BAF treatment. It can be used to assess alterations in autophagy. LC3 = Light-chain 3.

Activation of autophagy decreases the amount of intracellular bacilli after *M.tb* infection through inhibition of NF- $\kappa$ B<sup>523,524</sup>. NIK and IKK $\alpha$  are degraded by autophagy when the function of heat shock protein 90, required for the folding and maturation of certain signaling proteins, is inhibited<sup>319</sup>. The processing of p100 and NF- $\kappa$ B activity is thus inhibited<sup>320</sup>. However, when heat shock stress activates NF- $\kappa$ B, the autophagy pathway is activated, which suggests that there is a close interaction and tight regulation between these two pathways<sup>321</sup>. The relationship between autophagy and NF- $\kappa$ B is still, however, very poorly understood.

The effect of *MAP3K14 V345M* on autophagy was investigated in the present study by measuring the amount of LC3-II in mutant and wild-type samples before and after BAFF treatment. The control protein GAPDH was used to normalize each of the values to enable comparison between samples. Although the western blot image in Figure 3.20 in section 3.4.2.5 illustrates a decrease in LC3-II levels upon introduction of *MAP3K14 V345M* into HEK293 cells, this decrease is not significant after normalization. Differences in both pWZL-Neo-Myr-Flag-MAP3K14 and pWZL-Neo-Myr-Flag-MAP3K14-V345M transfected cells before and after BAFF treatment show that the autophagy pathway function correctly. To determine whether the induced mutation affected autophagy, the difference between pWZL-Neo-Myr-Flag-MAP3K14 cells with and without BAFF treatment was compared to that of *MAP3K14 V345M* cells with and without BAFF treatment. This difference was, however, not statistically significant ( $p = 0.178$ ), indicating that the induced mutation did not influence autophagy.

It is possible that the phenotypes observed in PID\_040 could be caused by decreased IKK $\alpha$  phosphorylation. These functional studies would need to be repeated with patient-derived immune cells to confirm our findings and test this synergism hypothesis.

#### **4.5. FLOW CYTOMETRY**

The outcome of TB infection mostly depends on the efficiency of the host immune system<sup>525</sup>. Protection against TB in human hosts depends on immune responses mediated by CD4<sup>+</sup> and CD8<sup>+</sup> T cells<sup>526</sup>. In the present study flow cytometry was used to measure the amount of CD4<sup>+</sup> and CD8<sup>+</sup> T cells in patient PID\_012 and a healthy control, both of which were HIV-negative at the time. This was done because *TAP1*, in which the I296M variant was identified in this patient,

is associated with CD8 T cell activation. We observed significant reduction in CD4<sup>+</sup> T cells in the patient compared to the control, accompanied by an almost two-fold increase in CD8<sup>+</sup> T cells attributed to a dramatic increase in the CD8<sup>medium</sup> cell population.

The crucial influence of CD4<sup>+</sup> T cells on TB disease outcome has been demonstrated in LTBI patients infected with HIV. In these patients, the depletion of CD4<sup>+</sup> T cells results in LTBI reactivation. Moreover, in TB patients without HIV, CD4 depletion have also been observed<sup>527</sup>. Similar to what we observed, several previous studies documented CD4 lymphopenia in HIV-negative TB patients<sup>528-531</sup>. One study observed this CD4 reduction in both smear negative and smear positive PTB patients<sup>532</sup>. Significantly decreased CD4 counts were also observed in HIV-seronegative extra-pulmonary and miliary TB patients, suggesting that TB should be considered a disease related to “idiopathic CD4<sup>+</sup> T-lymphocytopenia”<sup>533</sup>.

The role of CD8<sup>+</sup> T cells in TB is less well understood than that of CD4<sup>+</sup> T cells<sup>534,535</sup>. CD8<sup>+</sup> T-cells recognize antigens in the context of MHC class I molecules<sup>536</sup>. These cell types recognize and lyse cells infected with *M.tb* and produce IFN- $\gamma$  and TNF- $\alpha$  to activate macrophages<sup>102</sup>. In this study we observed a total CD8<sup>+</sup> T cell subset in a healthy control individual of 26.2%, consisting of 20.6% CD8<sup>high</sup> (also referred to as CD8<sup>bright</sup>) and 5.43% CD8<sup>medium</sup> (or CD8<sup>dim</sup>) T cells. In contrast, PID\_012 presented with a total CD8<sup>+</sup> T cell population of 50.3% with as much as half of that being CD8<sup>medium</sup> T cells.

CD8 antigen is expressed on some T cells, known as CD8<sup>+</sup> T cells or CD8<sup>high</sup> T cells, as well as on a subset of natural killer (NK) cells, known as CD8<sup>medium</sup> T cells. These “high” and “medium” distinctions are based on the lower amount of CD8 antigen present on NK cells compared to the number of antigen on CD8<sup>+</sup> T cells<sup>537,538</sup>.

Investigation of the relationship between the presence of class I human leukocyte antigen (HLA) molecules and co-expression of their receptors resulted in the observation that TAP-deficient individuals present with significantly increased CD8<sup>+</sup> NK cell percentages compared to controls<sup>539</sup>. CD8<sup>+</sup> NK cells have increased cytotoxicity and are less prone to target cell-induced apoptosis compared to CD8<sup>-</sup> NK cells, with most NK cells deficient of functional TAP being

CD8<sup>+</sup> cells<sup>540</sup>. CD8 is thought to influence the growth of the killer Ig-like receptors (KIR) collection in TAP deficient individuals<sup>539</sup>. Changes in the compartments of TCR $\alpha\beta$ 1 CD8 T-cells and NK cells have been associated with decreased expression of MHC class I molecules. These molecules play a role in NK cell maturation and “licensing”, and are responsible for the selection of cytotoxic T cells in the thymus<sup>541,542</sup>.

The existence of a spectrum of immune response in TB has been demonstrated<sup>543</sup>. TB patients were classified into reactive (RR), unreactive (UU) or intermediate groups. RR patients present with good cell-mediated immunity combined with limited or no antibody production, while UU individuals have poor cellular responses combined with extensive antibody production<sup>544</sup>. Intermediate forms present with features of both groups. Intermediate and UU TB forms were associated with CD8<sup>+</sup> and B cell lymphocytosis, reduced CD4/CD8 ratios, and unaltered CD4<sup>+</sup> subsets and T lymphocytes, while RR types typically showed CD4<sup>+</sup> lymphopenia and reduced CD4/CD8 ratios<sup>544</sup>. Healthy individuals typically present with a CD4/CD8 ratio in the range of 1.1 to 2.5 (1.8 average), which is in agreement with the ratio of 2.5 observed in the control individual used in our study.

Smear-positive and smear-negative TB patients typically show reduced CD4<sup>+</sup> and total T cell counts with increased amounts of CD8<sup>+</sup> cells and thus a reduced CD4/CD8 ratio (0.82 – 0.88)<sup>532</sup>. A heterozygous splice mutation in class IA phosphatidylinositol 3-kinases (*PI3K*) has been shown to cause immunodeficiency associated with antibody defects, lymphoproliferation, and terminal differentiation and senescence of CD8 T cells<sup>545</sup>. Most patients presenting with this mutation also displayed decreased CD4/CD8 ratios<sup>545,546</sup>, indicating that immunodeficiency and decreased CD4/CD8 ratios could be linked. Another study, however, noted significantly decreased CD4<sup>+</sup> counts, which normalized upon sufficient treatment, but normal CD8<sup>+</sup> counts in PTB patients compared to healthy controls. CD4/CD8 ratios were significantly reduced<sup>529</sup>.

Based on previous evidence that clinical manifestations, such as autoimmune disease and HIV progression to AIDS, are associated with irregular depletion or expansion of specific lymphocyte subsets, the genetic predictors of lymphocyte levels and their potential role in immune-related disorders were investigated<sup>547</sup>. Multiple SNPs in a region situated in the MHC was strongly

associated with CD4/CD8 ratios, and CD4 and CD8 levels. This result was subsequently replicated in 988 independent individuals. Two main independent quantitative trait loci (QTL) were identified in the MHC to regulate CD4/CD8 ratios. The most prominent (rs2524054) regulates CD8 levels and is situated in the class I cluster, while the second (rs9270986) regulates CD4 levels and is situated in the class II cluster. This suggests that MHC variants, partially via T cell dysregulation, might predispose individuals to immune-related disease<sup>547</sup>.

This study provides proof that mutations associated with the MHC, such as those in *TAP*, can influence CD4 and/or CD8 counts and thus CD4/CD8 ratios. The significantly decreased CD4/CD8 ratio of 0.3, the increased CD8<sup>+</sup> cell population and the decreased CD4<sup>+</sup> cell population observed in PID\_012 in the present study could thus well be caused by the *TAP1 I296M* variant identified.

#### 4.6. CASE-CONTROL ASSOCIATION STUDIES

A case-control association study was conducted to determine whether SNPs in *MAP3K14* and *TAP1* were associated with TB in the SAC population, using data from SAC TB patients and control individuals genotyped on the Illumina MEGA® array. No association was observed between neither *MAP3K14* nor *TAP1* and TB in this population.

To our knowledge, no previous case-control association studies have been conducted for *MAP3K14* and TB susceptibility.

A significant association between PTB and *TAP1* rs1135216 A/G was observed in an Iranian population, where the G allele increased PTB risk<sup>439</sup>. The rs1135216 A/G and rs1057141 A/G *TAP1* variants were also associated with increased TB risk in a Chinese population<sup>440</sup>. The rs1057141 G/G and rs1135216 G/A genotypes of *TAP1* positively correlated with HIV-TB co-infection, and may serve as risk factors for TB development in HIV-positive individuals<sup>548</sup>. However, no association between TB and *TAP1* were found in North-western Colombian, Korean or North Indian populations<sup>181,436,437</sup>.

In mice, TB10.4-specific T cells are dominant in CD8<sup>+</sup> T cell responses to *M.tb* infection. This protection depends on IFN $\gamma$  production, and antigen presentation by TAP is essential<sup>549</sup>. Antimicrobial activity of CD8<sup>+</sup> T cells can only be initiated when *M.tb*-infected cells are recognized by T cells receptors, for which TAP is needed. In mice where *TAPI* was completely knocked-out, susceptibility to TB, bacterial burden, mortality rates as well as the severity of tissue pathology were all increased<sup>549</sup>.

The frequency of *TAPI* polymorphisms was investigated in an indigenous Zimbabwean population by direct DNA sequencing and single-stranded conformation polymorphism procedures. Eleven nucleotide variations were identified in the complete *TAPI* exome, eight of which were non-conservative substitutions in functionally important regions of the protein<sup>550</sup>. In addition to this, five novel polymorphic sites, one of which was the P67S variant identified in PID\_060, were also identified. This and several previous studies observed substantial differences in *TAPI* allelic distributions between different populations<sup>551-554</sup>.

The potential association between TB and *TAPI* observed in the present study is thus in accordance with the critical function of this gene, and correlates with the findings of some previous studies. Variants located in the exons of *TAPI* can influence antigen presentation specificity, and thus disturb the translocation of mycobacterial peptides across biological membranes<sup>434,435</sup>.

Several factors can account for the inconsistencies observed in TB literature, as evident from the mentioned studies, such as differences in the investigated populations (other than just geography), the definition of phenotypes of cases and controls, complex genetic effects and possible variations in the infecting *M.tb* strains<sup>440,555</sup>. It is also possible that the presentation of antigens are affected by more than one polymorphism situated either in the same or in different genes. Polymorphisms can thus either intensify or counteract the effects of other polymorphisms<sup>556</sup>.

#### 4.6.1. Population genetics

One of the major causes of the discrepancies observed between genetic association studies is genetic heterogeneity, or inestimable polygenic effects, between different study populations<sup>557,558</sup>. The effect of various rare variants combined can result in polygenic effects<sup>559</sup>. Unfortunately, rare variants in TB disease have received little attention to date. Resequencing of *TLR* associated this gene with TB, and have uncovered several novel polymorphisms in populations from South Africa and Uganda<sup>560,561</sup>. CNVs have been associated with autoimmune diseases and HIV, and might also be involved in TB development<sup>562</sup>. The low frequencies at which these variants are present require them to be treated as rare variants during analysis, implying the need for novel analytical methods for association detection, very large sample sizes or large pedigrees<sup>563</sup>.

LD differs greatly between different populations and is another genetic component that may attribute to study inconsistencies. It is possible for trait-associated polymorphisms to be in LD with genotyped markers and be un-typed, which means that the ability of detecting associations between markers and traits can be confounded by variations in LD<sup>564</sup>. For example, no association was observed between four previously examined *NRAMP1* polymorphisms and TB in an African American population group, although association was found between other SNPs in this gene and TB<sup>565,566</sup>. *NRAMP1* association with TB in this population would have been missed, had these additional SNPs not been analyzed. Because of the complex nature of TB, genetic heterogeneity is very likely; thus, disease susceptibility is influenced by different variants across different populations. This highlights the importance of comprehensive genotyping rather than focusing on specific SNPs or promoters when undertaking genetic epidemiological studies<sup>567</sup>.

#### 4.6.2. Complex genetic effects

The effects of gene-environment and gene-gene interactions on TB is not well studied, despite the fact that these interactions often characterize complex traits. Several important genes may work in unison to influence TB susceptibility, but due to the fact that their separate effects do not adhere to certain standards for statistical significance, these interactions may be missed<sup>567</sup>. Examples include identified interactions between *NOS2A* and *TLR4*, *NOS2A* and *IFNGR1*, and

*NRAMP1* and *TLR2*, although analysis found no significant main effects of neither *TLR2*, *TLR4* nor *IFNGRI*<sup>565,566</sup>.

Gene effects may further be altered by epidemiological factors, such as the interactions between the human and a mycobacterial genomes. Additionally, a strong confounder of TB immune responses is HIV. Since most TB studies exclude individuals that are HIV-positive, information with regards to its effect on the host genetics of TB is limited<sup>567</sup>. Furthermore, HIV have been shown to alter *TNFR1*'s effect on the expression of  $TNF\alpha$ <sup>568</sup>, possibly implying that some genetic factors can predispose only HIV-infected or -uninfected individuals to TB.

#### **4.6.3. Categorizing cases and controls**

TB pathogenesis occurs in two stages: LTBI, diagnosed by positive TST and/or positive IGRA combined with the lack of symptoms or clinical signs of active disease; and active TB disease, characterized by *M.tb* growth in sputum/culture or positive acid-fast bacilli (AFB) smear, accompanied by characteristic chest X-rays and active disease symptoms (fever, weight loss, persistent cough)<sup>569,570</sup>. Different genetic factors have been associated with these two stages. Distinct regions have been associated with LTBI phenotypes<sup>571-573</sup> while other studies related genetic association to progression of active disease from LTBI<sup>568,574,575</sup>. These studies, however, are minimal.

In order to investigate an association between a disease and specific risk factors, case-control studies require controls to be exactly similar to cases, with the exception of being diseased. Controls used in TB studies should thus have been exposed to infectious TB patients and have had the chance to develop active disease. However, no classification of exposure in controls currently exist. In TB-endemic settings it is often assumed that individuals without active disease will be latently infected, since all have been exposed to *M.tb*<sup>576,577</sup>. In contrast, LTBI is never diagnosed in some individuals persistently exposed to *M.tb*<sup>570,572</sup>.

Population controls are also used in many studies, although the danger of misclassification is very possible<sup>578</sup>. In these cases where the LTBI statuses of controls are unknown, the lack of associations observed between TB and polymorphisms does not exclude the possibility of

association between specific variants and LTBI. Furthermore, individuals carrying risk variants might never be exposed to active TB cases and thus never develop TB<sup>579</sup>.

TB diagnosis also varies greatly among studies. *M.tb* growth in culture remains the best way of effectively diagnosing TB<sup>570</sup>. However, due to the lack of facilities required for culturing *M.tb*, many researchers base their diagnosis on less-sensitive AFB smear positivity. Differences in the severity of disease could be indicated by the grade of AFB smears. Also, because of HIV co-infection, culture-positive, smear-negative TB is problematic in some areas<sup>567</sup>. Fundamental differences in severity of disease could be reflected by the variety of TB diagnostic criteria among studies, which can become very important if host genes affect TB severity. Further bias could be introduced into genetic studies by the differences in specificity and sensitivity between AFB smears and culture leading to misclassifications<sup>567</sup>.

#### **4.6.4. *M.tb* strain diversity**

Six major strain lineages of *M.tb* have been associated with specific clinical presentations, geographical regions and rate of progression to active disease<sup>193,580,581</sup>. Disease severity can thus also be influenced by infecting *M.tb* strains. An interaction between *M.tb* genotype and host genotype has been observed, where a *TLR2* genotype associated with the Beijing strain<sup>582</sup>.

In conclusion, the definition of phenotype is extremely important in TB studies, and the duration, presentation as well as severity of TB is exceptionally heterogeneous<sup>583</sup>. Although most cases of TB present as pulmonary disease, almost all organ systems can be affected and even in PTB, presentation of the disease can differ greatly among individuals. Due to the heterogeneous nature of TB, defining a consistent phenotype for genetic analysis is tricky. TB diagnostic criteria differ significantly between studies based on resource availability. Furthermore, TB can develop any time post infection and it is possible for active disease to only develop after data collection.

All PTB cases used in the current study were smear/culture positive, and all controls thus smear/culture negative. However, the possibility exist that controls were latently infected and developed PTB at a later stage in life. Disease severity and dissemination was also not investigated or classified in the cases, which could potentially skew the data.

## 4.7. STUDY LIMITATIONS

### 4.7.1. Whole exome sequencing

The foremost limitation of using WES to detect potentially disease-causing mutations is that only exonic variants are identified. Variants in intronic regions could not be investigated, although intronic variants can have significant effects and can affect and interact with epigenetic factors and influence gene and protein expression. However, the exome contains 85% of disease-causing mutations in Mendelian disorders and many disease-predisposing SNPs throughout the genome. It was thus not feasible to conduct the much more expensive and data-intensive whole genome sequencing in order to investigate the rest of the genome that only has a 15% chance of causing disease.

The use of WES to detect novel MSMD-causing mutations in South African patients has several limitations. Technical and analytical limitations, such as assembly and variant calling, are challenges when identifying novel variants. Other causal variants can exist in patients that have been sequenced but not accurately called<sup>584</sup>. In some cases allele swaps occur, when variants identified as being heterozygous recognize the major allele as the minor allele. A large amount of false positive calls could thus be generated, as well as several sequencing artefacts. The causal variant could also be missed.

Improvements are constantly being made to WES technologies. The current sequence information available from databases like Refseq and consensus coding sequence (CCDS) are used to design sequence capture methods, which means that it is not possible to capture yet-to-be annotated or still unknown exons. Entire genes, or parts of genes, can thus be outside the target definition of a specific exon and be missed completely. Another limitation is that the function of variants outside the exome in Mendelian disease has not yet been described<sup>585</sup>. There is also a great need for a well-defined analytical pipeline to use for the filtration of potentially pathogenic variants. Additional putative variants could be present in patients without having been identified, due to stringent filtration parameters employed.

Probably the most substantial analytical limitation in conducting WES on SA patients is that no population-specific databases exist. This makes it difficult to determine whether sequence

variants identified by WES and subsequent filtering pipelines are good candidates, since it is possible for these variants to be present at high frequencies in these populations. It is thus of utmost importance for the frequency of any detected variants to be screened in representative control groups. More than 300 controls from the SAC population were available for genotyping. However, we believe that the amount of controls genotyped in each case was sufficient, since a MAF cut-off of 1% was used to indicate rare mutation.

The human SNP database, dbSNP 132 (<http://www.ncbi.nlm.nih.gov/projects/SNP/>), currently contains more than 17 million SNPs. However, these have been identified at an error rate of 15-17%<sup>586,587</sup>, implying that at least some of these SNPs are sequencing artefacts instead of true variants. The frequency at which the variants identified in the current study have been observed in previous studies (such as *TAP1 P67S* rs375389015, *IL12RB1 T432M* rs200875188 and *TYK2 A53T* rs55762744) could thus potentially be much lower than reported, increasing the likelihood that these are involved in rare diseases.

The human reference genome offers a common framework for annotating genes and other functional units as well as regulatory elements. However, it does not provide any information as to how DNA sequences differ among individuals<sup>588</sup>. Characterizing rare variants in disease causality can be very challenging. On average, rare alleles appeared relatively recently, thus showing higher geographic clustering than common variants<sup>589,590</sup>. Large samples from each of the South African populations investigated in this study should thus be studied in order to build a catalogue of rare variants in the genomes of these populations<sup>591</sup>. The Genome of the Netherlands (GoNL) Project characterized the variations in DNA sequences of 769 Dutch individuals from 11 of the 12 Dutch provinces. The participants included 231 trios, 11 quartets with monozygotic twins and 8 quartets with dizygotic twins<sup>591</sup>. This data can potentially provide a good reference for the analysis of Afrikaner individuals, but no such large reference database characterizing Southern African diversity, which includes the diverse KhoeSan populations, is publically available.

In the decade since the Human Genome Project was completed, astonishingly little has been done with regards to genomic composition of African populations. Most genomic studies have

focused on the genomes of European, American and Asian populations. A new partnership have emerged between the Wellcome Trust and the US National Institutes of Health (NIH) to focus on funding population-based genetic studies in Africa. The Human Heredity and Health in Africa (H3Africa) Project will also build capacity for African scientists to conduct genomic studies<sup>592</sup>. The African Genome Variation Project (AGVP) also aims to provide a more efficient resource to aid in the design, implementation and interpretation of genomic studies in sub-Saharan Africa and worldwide<sup>593</sup>.

The 1KGP, HGMD and HapMap data used for variant identification and *in silico* prediction do also not represent the ethnic groups of SA. To at least some extent, the white South African population can be compared to European and American populations, while the Bantu population can also be compared to the African data present in these databases. However, the SAC population is highly admixed and have received genetic contributions from a large variety of populations<sup>594</sup>. Some potentially disease-causing variants in these three MSMD patients could thus potentially have been missed due to population differences. A large SA dataset is needed to control for population structure, distinguishing SAC population-specific SNVs that are unlikely present in any current public database from potentially deleterious variants.

#### **4.7.2. Patients**

Identifying MSMD patients in the TB- and HIV-endemic setting of South Africa is challenging, as any infections observed in patients are usually considered to be caused by untreated infectious epidemics rather than inborn errors of immunity<sup>225</sup>. MSMD manifests in many forms and can vary from local recurring to devastating disseminated disease<sup>226,347</sup>. Patients are susceptible to a wide variety of mycobacterial and Salmonella infections<sup>196,263,362,595,596</sup>. MSMD patients are easily be missed in a high burden TB setting where the population as a whole is constantly being exposed to TB in their environment. These patients may often be diagnosed as having recurrent TB or reactivation of a previous infection, rather than suffering from a genetic predisposition<sup>225</sup>. It is crucial to improve MSMD awareness in areas rampaged by TB in order to identify patients with novel or known MSMD-causing mutations.

The TBM cohort included in this study was very small. Ten individuals is not sufficient to definitively prove a relationship between MSMD and TBM. WES is expensive and data-intensive, and our resources did not allow for the exome sequencing of more individuals. However, our preliminary data indicate that there might well exist a link between MSMD-causing genes and the development of TBM in children from the SAC population. It is definitely worth investigating this hypothesis further in larger study cohorts from the same as well as other populations.

#### **4.7.3. Functional studies**

To accurately determine whether the variants detected in the MSMD and TBM patients could cause disease, functional studies designed specifically according to the function of each gene of interest, had to be conducted. This would have been very resource- and time-intensive and could not be achieved within the scope of this study. A great limitation to this study is that we did not validate the variants identified in the TBM patients using Sanger sequencing, nor was its frequencies investigated in ethnically matched controls. This will be done in future studies. For the scope of this study, the aim of this part of the study was purely to investigate a potential link between MSMD and TBM. Another limitation is that variants identified by exome sequencing, but not observed when using Sanger sequencing, was presumed to be sequencing artefacts and removed from further analysis. However, allele dropout is another well-known cause of missed mutations, and occurs due to unequal amplification of heterozygote alleles. Sanger sequencing could have been repeated using larger amplicons to eliminate this possibility.

Functional studies were conducted for the *MAP3K14 V345M* mutation identified in PID\_040 using an over-expression model. Although the ideal would have been to conduct these functional studies on patient-derived immune cells, this was not logistically possible. However, the results obtained in the current study give a good indication that the disease-causing mutation was successfully identified in PID\_040. Should we be able to obtain PBMCs from this patient in the future, many more studies can be conducted to prove the involvement of this mutation in disease.

The activity of the non-canonical NF- $\kappa$ B pathway can be investigated in patient- and control-derived cells by activating the BAFF and/or LT $\beta$  receptors with appropriate ligands and

subsequently measuring NIK, IKK $\alpha$ , p100 and p52 levels. If any differences are observed between the case and control, wild-type *MAP3K14* can be transferred into patient fibroblasts using retroviral genes to determine if functional NIK protein is the determining factor. Patient PBMCs can be stimulated with a range of stimuli to determine whether B cell survival is affected, by focusing on B cell activation and class-switch recombination capacity. The effect of the mutation of T cells can be investigated more extensively by flow cytometric analysis of different cell populations. The capability of patient NK cells to secrete cytokines (IFN $\gamma$  and TNF $\alpha$ ) and exert cytolytic function can also be investigated by activating these cells with phorbol 12-myristate 13-acetate and ionomycin.

#### **4.7.4. Case-control association studies**

By nature, case-control association studies have numerous limitations, such as the amount of markers tested, phenotypic heterogeneity, population stratification and statistical power and sample size. The following sections discuss the relevance of these short-comings in the present study.

##### **4.7.4.1. Number of markers tested**

The present study employed case-control analysis to compare differences in allele frequencies and genotype distribution of known tagSNPs in *TAP1* and *MAP3K14* present on the Illumina MEGA® between a case and a control cohorts.

The MEGA® contains data from Phase three of the 1KGP, Population Architecture in Genomics and Epidemiology (PAGE), the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA), and OMIM/Clinvar to create a true multi-ethnic array of approximately 700 000 propriety SNPs. A total of 19 SNPs in *TAP1* and 13 in *MAP3K14* are present on this array, all of which were screened for association with PTB in a cohort of controls and healthy cases. Because 32 markers were tested for TB association, Bonferroni correction was used to correct for multiple testing. The uncorrected p-values obtained for each SNP were thus multiplied by 32. The more SNPs investigated, the more stringent the correction applied, potentially leading to false negative association. However, the uncorrected p-values in the

current study were all very high, indicating the lack of association even before multiple testing correction.

#### **4.7.4.2. Phenotypic resolution**

TB and even PTB in itself is clinically very heterogeneous, with several environmental and host genetic factors influencing its expression. LTBI and active disease in the PTB cohort were not distinguished, and genes may significantly differ between these statuses<sup>568,573</sup>. The power of detecting small effect sizes may be increased, and “background noise” created by clinically heterogeneous samples may be reduced, when PTB cases are classified into specific subtypes, such as latency and active disease as well as different stages of active disease after infection. However, it should be noted that the traditional diagnosis of PTB (applied in this study) might still identify genes contributing to overall PTB pathology.

The three MSMD cases investigated in this study presented with clinical phenotypes indicative of some form of disseminated TB. It is thus possible that no association was found between SNPs in *TAP1* and *MAP3K14* and TB susceptibility because a cohort of only PTB cases were investigated, but these genes are involved in susceptibility to disseminated forms of TB. However, the objective here was only to identify genes associated with TB susceptibility in the SAC population, and not with different forms of the heterogeneous disease.

#### **4.7.4.3. Population stratification**

Underlying genetic differences among different populations account for the struggle to confirm association between diseases and certain candidate genes in different studies. Several environmental and genetic factors play a role in phenotypic presentation of complex diseases, such as TB, in different patients<sup>567</sup>. Given the environmental and genetic differences between populations, disease pathogenesis in different populations could be caused by unique combinations of genetic and environmental mechanisms. Moreover, at least some degree of cryptic population structure have been identified in most major ethnic groups<sup>597</sup>. Thus, if cryptic subpopulations exist within the population investigated for association, in which the allele frequencies for the chosen candidate genes and disease risk differ, false association might be detected. Likewise, if disease is more prevalent in the subpopulation with lower allele

frequencies, association might also be missed due to population stratifications<sup>598</sup>. In this study, all PTB cases and controls self-identified as part of the South African Colored population. No genetic information of their parents or grandparents were obtained, which creates room for great genetic admixture between individuals and creates genetically heterogeneous cohorts. However, population stratification was adjusted for to compensate for this.

#### **4.7.4.4. *Statistical power and sample size***

A number of related factors influence the size of a sample needed to obtain statistical significance when conducting genetic case-control association studies, such as the power of the study and the specific effect size<sup>599</sup>. The power of a study refers to its capability of detecting true effect if it exists, and to distinguish between an inconclusive result and a significant negative association. Specific effect size refers to the degree to which the investigated allelic variation contribute to disease pathogenesis<sup>600</sup>.

When no other basis for choosing the power of a study exist, a value of 80% is used, indicating that the study has an 80% probability of detecting true effects<sup>601</sup>. In the current study, however, a power of 80% was only achieved for 11 of the 32 SNPs investigated. As much as 2500 cases and 2500 controls would have been needed to reach an 80% power for the SNP with the lowest MAF of 0.0313. The lack of association observed in this study could thus be caused by the lack of sufficient power, and this case-control association study needs to be repeated in cohorts of sufficient size. Using a study population large enough to generate adequate statistical power is crucial for successfully detecting disease-causing genes in complex diseases via genetic association studies.

Meta-analyses can be used to combine association statistics from individual studies, in order to compensate for small sample sizes. This strategy uses a combination of statistical methods to collect correlational and experimental results from various autonomous studies aimed at addressing similar hypotheses. This method could be considered for future work to evaluate association between *TAPI* and *MAP3K14* and rare disseminated forms of TB. It can also be effective for further investigating the potential link between TBM and MSMD-causing genes.

Since several SNPs were investigated for disease association across different genes, multiple testing had to be accounted for. The major complication of conducting multiple comparisons is that the possibility of erroneously concluding a significant association when there is in fact none increases with each added test<sup>602</sup>. The present study used the Bonferroni correction to correct for multiple testing, which alters the p-value used to evaluate significance according to the total number of tests conducted.

#### 4.8. FUTURE STUDIES

We are in the process of recruiting more MSMD patients to be whole exome sequenced, and are especially aiming to sequence trios. The same sequencing and bioinformatics techniques, variant verification, variant frequency detection and *in silico* variant prediction methods described (2.2, 2.3, 2.4) will be implemented to identify potentially disease-causing variants. Appropriate functional studies will be designed according to the function of each gene in which a variant is identified. Flow cytometry and case-control association studies will again be performed. More TBM patients will also be sequenced to further investigate the potential link between MSMD-genes and TBM.

Should we be able to obtain PBMCs from PID\_040, all the functional studies conducted by Willmann and colleagues<sup>340</sup> as well as the functional work described in this thesis (Section 2.5 and Section 4.7.3) will be repeated using these cells. We predict that much more significant and reliable results will be obtain in this way to definitively prove *MAP3K14 V345M* to be the disease-causing mutation in PID\_040. Functional studies will also be designed to investigate the antigen presenting capacity of TAP1 in PBMCs from PID\_012, which can also be conducted for PID\_060 should we be able to obtain blood or PBMC samples from him. These would probably include investigation of the TAP transporter's capacity to recognize and bind peptides, the effective formation of MHC class I peptide-loading complexes, and recruitment and activation of cytotoxic T cells via CD8+ T cells.

The variants identified in the TBM patients should also be verified using Sanger sequencing and their frequencies investigated in ethnically-matched controls. These variants should then also be functionally validated in order to determine a definite link between MSMD and TBM.

#### 4.9. CONCLUDING REMARKS

The first case of MSMD was clinically described in 1951<sup>603</sup>. Since then, 18 genetic etiologies have been described<sup>604</sup>. However, much greater genetic heterogeneity underlies MSMD, as evident from the fact that about 50% of known MSMD patients do not suffer from these characterized defects. IFN- $\gamma$ -mediated immunity is crucial in controlling mycobacterial infections, and also seem to control immunity to other pathogens that infiltrate macrophages. A narrow phenotype have been associated with inborn errors of IFN- $\gamma$  immunity in humans. Phenocopies of MSMD that imitate innate errors of IFN- $\gamma$  immunity caused by autoantibodies against IFN- $\gamma$  have also been discovered<sup>605-608</sup>. Important immunological implications have been derived from the genetic examination of MSMD<sup>8,609-611</sup>. These studies have provided significant clinical, immunological and genetic characterization of infectious diseases affecting otherwise healthy individuals, and have proven that single-gene innate errors of immunity are the most probable cause of primary lethal infections in otherwise healthy children<sup>3,610</sup>. These types of studies aid in the effective treatment of patients by comprehending mycobacterial disease pathogenesis, and enable genetic counselling of patients and family members.

The genetic examination of MSMD has also enabled the effective genetic analysis of TB, especially severe childhood forms<sup>196,228</sup>. The identification of new genetic disorders underlying MSMD and TB is further enhanced by next generation sequencing technologies, with WES and whole genome sequencing (WGS) at the frontier<sup>330,375</sup>. These tools will most probably lead to the identification of additional pathways involved in MSMD and TB susceptibility through its ability to conduct detailed genetic investigations. It will thus also contribute to the development of more effective patient-based treatment strategies<sup>612,613</sup>.

The present study successfully identified two novel putative MSMD-causing genes and identified a possible association between MSMD-genes and the development of TBM in children. The era of personalized medical treatment based on unique genetic features has arrived, and can now be extended to mycobacterial diseases. It can be applied to the different levels of clinical disease management, namely prevention, diagnosis, treatment and prognosis, and genetic predisposition combined with protein dysfunction studies are used to tailor-make patient-specific approaches in mycobacterial disease<sup>614</sup>.

Identifying individuals at increased risk of developing TB at an early stage of life can save the lives of millions of patients, as well as identify at-risk relatives, by implementing patient-specific vaccine strategies and treatment regimes.

## REFERENCE LIST

Addgene:	<a href="https://www.addgene.org">https://www.addgene.org</a>
CCDS:	<a href="https://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi">https://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi</a>
dbSNP:	<a href="http://www.ncbi.nlm.nih.gov/projects/SNP/">http://www.ncbi.nlm.nih.gov/projects/SNP/</a>
Encode:	<a href="https://genome.ucsc.edu/ENCODE">https://genome.ucsc.edu/ENCODE</a>
ExAC Browser Beta:	<a href="http://exac.broadinstitute.org/">http://exac.broadinstitute.org/</a>
Gencode:	<a href="http://www.gencodegenes.org/">http://www.gencodegenes.org/</a>
HapMap:	<a href="http://hapmap.ncbi.nlm.nih.gov/">http://hapmap.ncbi.nlm.nih.gov/</a>
HOPE:	<a href="http://www.cmbi.ru.nl/hope/">http://www.cmbi.ru.nl/hope/</a>
RefSeq:	<a href="http://www.ncbi.nlm.nih.gov/refseq/">http://www.ncbi.nlm.nih.gov/refseq/</a>
MT2:	<a href="http://www.mutationtaster.org/">http://www.mutationtaster.org/</a>
NCBI GenBank:	<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>
NEBcutter v2.0:	<a href="http://nc2.neb.com/NEBcutter2/">http://nc2.neb.com/NEBcutter2/</a>
PDB:	<a href="http://www.rcsb.org/pdb/home/home.do">http://www.rcsb.org/pdb/home/home.do</a>
PP-2:	<a href="http://genetics.bwh.harvard.edu/pph2/">http://genetics.bwh.harvard.edu/pph2/</a>
Primer3Plus:	<a href="http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/">http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/</a>
wANNOVAR:	<a href="http://wannovar.usc.edu/index.php">http://wannovar.usc.edu/index.php</a>
SIFT:	<a href="http://sift.jcvi.org/">http://sift.jcvi.org/</a>
UniProt:	<a href="http://www.uniprot.org/">http://www.uniprot.org/</a>

1. Nicolle, C. Les infections inapparentes. *Scientia* **33**, 181–271 (1933).
2. Casanova, J. L. & Abel, L. The human model: a genetic dissection of immunity to infection in natural conditions. *Nat. Rev.* **4**, 55–66 (2004).
3. Alcais, A., Abel, L. & Casanova, J. L. Human genetics of infectious diseases: between proof of principle and paradigm. *J. Clin. Invest.* **119**, 2506–2514 (2009).
4. Casanova, J.-L. Human genetic basis of interindividual variability in the course of infection. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E7118-7127 (2015).
5. Sorensen, T. I., Nielsen, G. G., Andersen, P. K. & Teasdale, T. W. Genetic and environmental influences on premature death in adult adoptees. *N.Engl.J.Med.* **318**, 727–732 (1988).
6. Kwiatkowski, D. Science, medicine, and the future: susceptibility to infection. *BMJ* **321**, 1061–1065 (2000).
7. Alcais, A. & Abel, L. [Identification of a new leprosy susceptibility gene with positional cloning]. *Médecine Sci. MS* **20**, 729–732 (2004).

8. Casanova, J. L. & Abel, L. Primary immunodeficiencies: a field in its infancy. *Science* **317**, 617–619 (2007).
9. Allison, A. C. Observational, hypothesis-driven and genomics research strategies for analyzing inherited differences in responses to infectious diseases. *Public Health Genomics* **12**, 41–52 (2009).
10. Fischer, A. Human primary immunodeficiency diseases. *Immunity* **27**, 835–845 (2007).
11. Notarangelo, L. D. & Casanova, J.-L. Primary immunodeficiencies: increasing market share. *Curr. Opin. Immunol.* **21**, 461–465 (2009).
12. Moutsopoulos, N. M., Lionakis, M. S. & Hajishengallis, G. Inborn errors in immunity: unique natural models to dissect oral immunity. *J. Dent. Res.* **94**, 753–758 (2015).
13. Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends Genet. TIG* **17**, 502–510 (2001).
14. Quintana-Murci, L., Alcais, A., Abel, L. & Casanova, J. L. Immunology in natura: clinical, epidemiological and evolutionary genetics of infectious diseases. *Nat. Immunol.* **8**, 1165–1171 (2007).
15. Casanova, J.-L., Conley, M. E., Seligman, S. J., Abel, L. & Notarangelo, L. D. Guidelines for genetic studies in single patients: lessons from primary immunodeficiencies. *J. Exp. Med.* **211**, 2137–2149 (2014).
16. World Health Organization. *Global Tuberculosis Report 2015*. (2015).
17. Bustamante, J., Boisson-Dupuis, S., Abel, L. & Casanova, J. L. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN-gamma immunity. *Semin. Immunol.* **26**, 454–470 (2014).
18. World Health Organisation. 2009 Update Tuberculosis Facts. [http://www.who.int/tb/publications/2009/factsheet\\_tb\\_2009update\\_dec09.pdf](http://www.who.int/tb/publications/2009/factsheet_tb_2009update_dec09.pdf) (2010). Available at: [http://www.who.int/tb/publications/2009/factsheet\\_tb\\_2009update\\_dec09.pdf](http://www.who.int/tb/publications/2009/factsheet_tb_2009update_dec09.pdf).
19. WHO. World Health Statistics 2015. (2015). Available at: [http://www.who.int/gho/publications/world\\_health\\_statistics/2015/en/](http://www.who.int/gho/publications/world_health_statistics/2015/en/).

20. World Health Organization. Global Tuberculosis: WHO Report 2012. (2012). Available at: [http://apps.who.int/iris/bitstream/10665/75938/1/9789241564502\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/75938/1/9789241564502_eng.pdf).
21. Fox, G. J. & Menzies, D. Epidemiology of tuberculosis immunology. *Adv. Exp. Med. Biol.* **783**, 1–32 (2013).
22. Salem, S. & Gros, P. Genetic determinants of susceptibility to Mycobacterial infections: IRF8, a new kid on the block. *Adv. Exp. Med. Biol.* **783**, 45–80 (2013).
23. Abel, L., El-Baghdadi, J., Bousfiha, A. A., Casanova, J. L. & Schurr, E. Human genetics of tuberculosis: a long and winding road. *Philos. Trans. R. Soc. London Series B Biol. Sci.* **369**, 20130428 (2014).
24. Fox, G. J., Orlova, M. & Schurr, E. Tuberculosis in Newborns: The Lessons of the ‘Lübeck Disaster’ (1929-1933). *PLoS Pathog.* **12**, e1005271 (2016).
25. Embley, T. M. & Stackebrandt, E. The molecular phylogeny and systematics of the actinomycetes. *Annu. Rev. Microbiol.* **48**, 257–289 (1994).
26. Fabre, M. *et al.* Molecular characteristics of ‘Mycobacterium canettii’ the smooth Mycobacterium tuberculosis bacilli. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **10**, 1165–1173 (2010).
27. Schlebusch, C. M. *et al.* Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* **338**, 374–379 (2012).
28. Chisholm, R. H., Trauer, J. M., Curnoe, D. & Tanaka, M. M. Controlled fire use in early humans might have triggered the evolutionary emergence of tuberculosis. *Proc. Natl. Acad. Sci.* **113**, 9051–9056 (2016).
29. Sreevatsan, S. *et al.* Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 9869–9874 (1997).
30. Wheeler, P. R. & Ratledge, C. *Tuberculosis: Pathogenesis, Protection, and Control.* (1994).

31. Bansal-Mutalik, R. & Nikaido, H. Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4958–4963 (2014).
32. Sani, M. *et al.* Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. *PLoS Pathog.* **6**, e1000794 (2010).
33. Kieser, K. J. & Rubin, E. J. How sisters grow apart: mycobacterial growth and division. *Nat. Rev. Microbiol.* **12**, 550–562 (2014).
34. Layre, E. *et al.* A comparative lipidomics platform for chemotaxonomic analysis of *Mycobacterium tuberculosis*. *Chem. Biol.* **18**, 1537–1549 (2011).
35. Léger, M. *et al.* The dual function of the *Mycobacterium tuberculosis* FadD32 required for mycolic acid biosynthesis. *Chem. Biol.* **16**, 510–519 (2009).
36. Singh, B. *et al.* Asymmetric growth and division in *Mycobacterium* spp.: compensatory mechanisms for non-medial septa. *Mol. Microbiol.* **88**, 64–76 (2013).
37. Santi, I., Dhar, N., Bousbaine, D., Wakamoto, Y. & McKinney, J. D. Single-cell dynamics of the chromosome replication and cell division cycles in mycobacteria. *Nat. Commun.* **4**, 2470 (2013).
38. Bhamidi, S. *et al.* A bioanalytical method to determine the cell wall composition of *Mycobacterium tuberculosis* grown in vivo. *Anal. Biochem.* **421**, 240–249 (2012).
39. Aldridge, B. B. *et al.* Asymmetry and aging of mycobacterial cells lead to variable growth and antibiotic susceptibility. *Science* **335**, 100–104 (2012).
40. Boon, C. & Dick, T. How *Mycobacterium tuberculosis* goes to sleep: the dormancy survival regulator DosR a decade later. *Future Microbiol.* **7**, 513–518 (2012).
41. Chao, M. C. & Rubin, E. J. Letting sleeping dos lie: does dormancy play a role in tuberculosis? *Annu. Rev. Microbiol.* **64**, 293–311 (2010).
42. Majumdar, S. D. *et al.* Appropriate DevR (DosR)-mediated signaling determines transcriptional response, hypoxic viability and virulence of *Mycobacterium tuberculosis*. *PloS One* **7**, e35847 (2012).

43. Leistikow, R. L. *et al.* The Mycobacterium tuberculosis DosR regulon assists in metabolic homeostasis and enables rapid recovery from nonrespiring dormancy. *J. Bacteriol.* **192**, 1662–1670 (2010).
44. Rustad, T. R., Harrell, M. I., Liao, R. & Sherman, D. R. The enduring hypoxic response of Mycobacterium tuberculosis. *PLoS One* **3**, e1502 (2008).
45. Bold, T. D., Banaei, N., Wolf, A. J. & Ernst, J. D. Suboptimal activation of antigen-specific CD4<sup>+</sup> effector cells enables persistence of M. tuberculosis in vivo. *PLoS Pathog.* **7**, e1002063 (2011).
46. Mehra, S. *et al.* The DosR Regulon Modulates Adaptive Immunity and Is Essential for Mycobacterium tuberculosis Persistence. *Am. J. Respir. Crit. Care Med.* **191**, 1185–1196 (2015).
47. Mariotti, S. *et al.* Dormant Mycobacterium tuberculosis fails to block phagosome maturation and shows unexpected capacity to stimulate specific human T lymphocytes. *J. Immunol. Baltim. Md 1950* **191**, 274–282 (2013).
48. Hershkovitz, I. *et al.* Detection and molecular characterization of 9,000-year-old Mycobacterium tuberculosis from a Neolithic settlement in the Eastern Mediterranean. *PLoS One* **3**, e3426 (2008).
49. Nuorala, E., Gotherstrom, A., Ahlstrom, T., Donoghue, H. D. & Spigelman, M. MTB complex DNA in a Scandinavian Neolithic passage grave. *Scientific Archaeology* **6**, (2004).
50. Ortner, D. J. & Putschar, W. G. *Identification of pathological conditions on human skeletal remains.* (Smithsonian Institution Press, 1981).
51. Zink, A., Haas, C. J., Reischl, U., Szeimies, U. & Nerlich, A. G. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *J. Med. Microbiol.* **50**, 355–366 (2001).
52. Canci, A., Minozzi, S. & Borgognini Tarli, S. M. New evidence of tuberculous spondylitis from Neolithic Liguria. *Int J Osteoarchaeol* **6**, 497–501 (1996).

53. Formicola, V., Milanesi, Q. & Scarsini, C. Evidence of spinal tuberculosis at the beginning of the fourth millennium BC from Arene Candide cave (Liguria, Italy). *Am. J. Phys. Anthropol.* **72**, 1–6 (1987).
54. Spigelman, M. & Lemma, E. The use of the polymerase chain reaction (PCR) to detect *Mycobacterium tuberculosis* in ancient skeletons. *Int. J. Osteoarchaeol.* **3**, 137–143 (1993).
55. Spigelman, M., Matheson, C., Lev, G., Greenblatt, C. & Donoghue, H. D. Confirmation of the presence of *Mycobacterium tuberculosis* complex-specific DNA in three archaeological specimens. (2002).
56. Salo, W. L., Aufderheide, A. C., Buikstra, J. & Holcomb, T. A. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 2091–2094 (1994).
57. Said, H. M. *et al.* A Novel Molecular Strategy for Surveillance of Multidrug Resistant Tuberculosis in High Burden Settings. *PloS One* **11**, e0146106 (2016).
58. Rao, K. R., Ahmed, N., Srinivas, S., Sechi, L. A. & Hasnain, S. E. Rapid identification of *Mycobacterium tuberculosis* Beijing genotypes on the basis of the mycobacterial interspersed repetitive unit locus 26 signature. *J. Clin. Microbiol.* **44**, 274–277 (2006).
59. Arenas, N. *et al.* Detecting active tuberculosis in Calarca-Quindio, Colombia, during 2005. *Rev. Salud Publica Bogota Colomb.* **10**, 279–289 (2008).
60. Driscoll, J. R. Spoligotyping for molecular epidemiology of the *Mycobacterium tuberculosis* complex. *Methods Mol. Biol. Clifton NJ* **551**, 117–128 (2009).
61. Liu, Q. *et al.* Molecular typing of *Mycobacterium tuberculosis* isolates circulating in Jiangsu province, China. *BMC Infect. Dis.* **11**, 288-2334-11–288 (2011).
62. Balcells, M. E. *et al.* A First Insight on the Population Structure of *Mycobacterium tuberculosis* Complex as Studied by Spoligotyping and MIRU-VNTRs in Santiago, Chile. *PloS One* **10**, e0118007 (2015).

63. Chen, Y. Y. *et al.* Molecular epidemiology of tuberculosis in Kaohsiung City located at southern Taiwan, 2000-2008. *PloS One* **10**, e0117061 (2015).
64. Parsons, L. M. *et al.* Rapid and simple approach for identification of Mycobacterium tuberculosis complex isolates by PCR-based genomic deletion analysis. *J. Clin. Microbiol.* **40**, 2339–2345 (2002).
65. Manchester, K. Tuberculosis and leprosy in antiquity: an interpretation. *Med. Hist.* **28**, 162–173 (1984).
66. Clark, G. A., Kelley, M. A., Grange, J. M. & Hill, M. C. The evolution of mycobacterial disease in human populations: a reevaluation. *Curr. Anthropol.* **28**, 45–62 (1987).
67. Cole, S. T. *et al.* Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**, 537–544 (1998).
68. Behr, M. A. *et al.* Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**, 1520–1523 (1999).
69. Gordon, S. V. *et al.* Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Mol. Microbiol.* **32**, 643–655 (1999).
70. Brosch, R. *et al.* Genomic analysis reveals variation between Mycobacterium tuberculosis H37Rv and the attenuated M. tuberculosis H37Ra strain. *Infect. Immun.* **67**, 5768–5774 (1999).
71. Brosch, R. *et al.* A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 3684–3689 (2002).
72. Mostowy, S. & Behr, M. A. The origin and evolution of Mycobacterium tuberculosis. *Clin. Chest Med.* **26**, 207–16, v–vi (2005).
73. Witas, H. W., Donoghue, H. D., Kubiak, D., Lewandowska, M. & Gładkowska-Rzeczycka, J. J. Molecular studies on ancient M. tuberculosis and M. leprae: methods of pathogen and host DNA analysis. *Eur. J. Clin. Microbiol. Infect. Dis. Off. Publ. Eur. Soc. Clin. Microbiol.* **34**, 1733–1749 (2015).

74. Biet, F. *et al.* Inter- and intra-subtype genotypic differences that differentiate *Mycobacterium avium* subspecies paratuberculosis strains. *BMC Microbiol.* **12**, 264-2180-12-264 (2012).
75. O'Garra, A. *et al.* The immune response in tuberculosis. *Annu. Rev. Immunol.* **31**, 475-527 (2013).
76. Guirado, E. & Schlesinger, L. S. Modeling the *Mycobacterium tuberculosis* Granuloma - the Critical Battlefield in Host Immunity and Disease. *Front. Immunol.* **4**, 98 (2013).
77. Galagan, J. E. Genomic insights into tuberculosis. *Nat. Rev.* **15**, 307-320 (2014).
78. Millman, A. C., Salman, M., Dayaram, Y. K., Connell, N. D. & Venketaraman, V. Natural killer cells, glutathione, cytokines, and innate immunity against *Mycobacterium tuberculosis*. *J. Interferon Cytokine Res. Off. J. Int. Soc. Interferon Cytokine Res.* **28**, 153-165 (2008).
79. Morris, D. *et al.* Unveiling the mechanisms for decreased glutathione in individuals with HIV infection. *Clin. Dev. Immunol.* **2012**, 734125 (2012).
80. Guerra, C. *et al.* Glutathione and adaptive immune responses against *Mycobacterium tuberculosis* infection in healthy and HIV infected individuals. *PloS One* **6**, e28378 (2011).
81. Sia, J. K., Georgieva, M. & Rengarajan, J. Innate Immune Defenses in Human Tuberculosis: An Overview of the Interactions between *Mycobacterium tuberculosis* and Innate Immune Cells. *J. Immunol. Res.* **2015**, 747543 (2015).
82. Philips, J. A. & Ernst, J. D. Tuberculosis pathogenesis and immunity. *Annu. Rev. Pathol. Mech. Dis.* **7**, 353-384 (2012).
83. Rajaram, M. V. S. *et al.* *Mycobacterium tuberculosis* activates human macrophage peroxisome proliferator-activated receptor gamma linking mannose receptor recognition to regulation of immune responses. *J. Immunol. Baltim. Md 1950* **185**, 929-942 (2010).
84. Liu, L., Liu, J., Niu, G., Xu, Q. & Chen, Q. *Mycobacterium tuberculosis* 19-kDa lipoprotein induces Toll-like receptor 2-dependent peroxisome proliferator-activated receptor  $\gamma$

- expression and promotes inflammatory responses in human macrophages. *Mol. Med. Rep.* **11**, 2921–2926 (2015).
85. Ogden, C. A. *et al.* C1q and mannose binding lectin engagement of cell surface calreticulin and CD91 initiates macropinocytosis and uptake of apoptotic cells. *J. Exp. Med.* **194**, 781–795 (2001).
  86. Mwandumba, H. C. *et al.* Mycobacterium tuberculosis resides in nonacidified vacuoles in endocytically competent alveolar macrophages from patients with tuberculosis and HIV infection. *J. Immunol. Baltim. Md 1950* **172**, 4592–4598 (2004).
  87. Mwandumba, H. C. *et al.* Alveolar macrophages from HIV-infected patients with pulmonary tuberculosis retain the capacity to respond to stimulation by lipopolysaccharide. *Microbes Infect. Inst. Pasteur* **9**, 1053–1060 (2007).
  88. Brightbill, H. D. *et al.* Host defense mechanisms triggered by microbial lipoproteins through toll-like receptors. *Science* **285**, 732–736 (1999).
  89. Flynn, J. L. *et al.* An essential role for interferon gamma in resistance to Mycobacterium tuberculosis infection. *J. Exp. Med.* **178**, 2249–2254 (1993).
  90. Ottenhoff, T. H., Kumararatne, D. & Casanova, J. L. Novel human immunodeficiencies reveal the essential role of type-I cytokines in immunity to intracellular bacteria. *Immunol. Today* **19**, 491–494 (1998).
  91. Chackerian, A. A., Perera, T. V. & Behar, S. M. Gamma interferon-producing CD4+ T lymphocytes in the lung correlate with resistance to infection with Mycobacterium tuberculosis. *Infect. Immun.* **69**, 2666–2674 (2001).
  92. Jouanguy, E. *et al.* Interferon-gamma-receptor deficiency in an infant with fatal bacille Calmette-Guerin infection. *N. Engl. J. Med.* **335**, 1956–1961 (1996).
  93. Rook, G. A., Steele, J., Ainsworth, M. & Champion, B. R. Activation of macrophages to inhibit proliferation of Mycobacterium tuberculosis: comparison of the effects of recombinant gamma-interferon on human monocytes and murine peritoneal macrophages. *Immunology* **59**, 333–338 (1986).

94. Kincaid, E. Z. & Ernst, J. D. Mycobacterium tuberculosis exerts gene-selective inhibition of transcriptional responses to IFN-gamma without inhibiting STAT1 function. *J. Immunol. Baltim. Md 1950* **171**, 2042–2049 (2003).
95. Ting, L. M., Kim, A. C., Cattamanchi, A. & Ernst, J. D. Mycobacterium tuberculosis inhibits IFN-gamma transcriptional responses without inhibiting activation of STAT1. *J. Immunol.* **163**, 3898–3906 (1999).
96. Cappelli, G. *et al.* Human macrophage gamma interferon decreases gene expression but not replication of Mycobacterium tuberculosis: analysis of the host-pathogen reciprocal influence on transcription in a comparison of strains H37Rv and CMT97. *Infect. Immun.* **69**, 7262–7270 (2001).
97. Nau, G. J. *et al.* Human macrophage activation programs induced by bacterial pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 1503–1508 (2002).
98. Wang, J. P., Rought, S. E., Corbeil, J. & Guiney, D. G. Gene expression profiling detects patterns of human macrophage responses following Mycobacterium tuberculosis infection. *Fems Immunol. Med. Microbiol.* **39**, 163–172 (2003).
99. Volpe, E. *et al.* Gene expression profiling of human macrophages at late time of infection with Mycobacterium tuberculosis. *Immunology* **118**, 449–460 (2006).
100. Ragno, S. *et al.* Changes in gene expression in macrophages infected with Mycobacterium tuberculosis: a combined transcriptomic and proteomic approach. *Immunology* **104**, 99–108 (2001).
101. McGarvey, J. A., Wagner, D. & Bermudez, L. E. Differential gene expression in mononuclear phagocytes infected with pathogenic and non-pathogenic mycobacteria. *Clin. Exp. Immunol.* **136**, 490–500 (2004).
102. Egen, J. G. *et al.* Intravital imaging reveals limited antigen presentation and T cell effector function in mycobacterial granulomas. *Immunity* **34**, 807–819 (2011).

103. Grakoui, A., John Wherry, E., Hanson, H. L., Walker, C. & Ahmed, R. Turning on the off switch: regulation of anti-viral T cell responses in the liver by the PD-1/PD-L1 pathway. *J. Hepatol.* **45**, 468–472 (2006).
104. Freeman, G. J., Wherry, E. J., Ahmed, R. & Sharpe, A. H. Reinvigorating exhausted HIV-specific T cells via PD-1-PD-1 ligand blockade. *J. Exp. Med.* **203**, 2223–2227 (2006).
105. Sharpe, A. H., Wherry, E. J., Ahmed, R. & Freeman, G. J. The function of programmed cell death 1 and its ligands in regulating autoimmunity and infection. *Nat. Immunol.* **8**, 239–245 (2007).
106. Berry, M. P. R. *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* **466**, 973–977 (2010).
107. Geijtenbeek, T. B. H. *et al.* Mycobacteria target DC-SIGN to suppress dendritic cell function. *J. Exp. Med.* **197**, 7–17 (2003).
108. Nigou, J., Zelle-Rieser, C., Gilleron, M., Thurnher, M. & Puzo, G. Mannosylated lipoarabinomannans inhibit IL-12 production by human dendritic cells: evidence for a negative signal delivered through the mannose receptor. *J. Immunol. Baltim. Md 1950* **166**, 7477–7485 (2001).
109. Trinchieri, G. Interleukin-12: a cytokine at the interface of inflammation and immunity. *Adv. Immunol.* **70**, 83–243 (1998).
110. Altare, F. *et al.* Inherited interleukin 12 deficiency in a child with bacille Calmette-Guerin and Salmonella enteritidis disseminated infection. *J. Clin. Invest.* **102**, 2035–2040 (1998).
111. Altare, F. *et al.* Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency. *Science* **280**, 1432–1435 (1998).
112. de Jong, R. *et al.* Severe mycobacterial and Salmonella infections in interleukin-12 receptor-deficient patients. *Science* **280**, 1435–1438 (1998).
113. Altare, F. *et al.* Mendelian susceptibility to mycobacterial infection in man. *Curr. Opin. Immunol.* **10**, 413–417 (1998).

114. Wakeham, J. *et al.* Lack of both types 1 and 2 cytokines, tissue inflammatory responses, and immune protection during pulmonary infection by *Mycobacterium bovis* bacille Calmette-Guérin in IL-12-deficient mice. *J. Immunol. Baltim. Md 1950* **160**, 6101–6111 (1998).
115. Hanekom, W. A. *et al.* *Mycobacterium tuberculosis* inhibits maturation of human monocyte-derived dendritic cells in vitro. *J. Infect. Dis.* **188**, 257–266 (2003).
116. Cella, M., Engering, A., Pinet, V., Pieters, J. & Lanzavecchia, A. Inflammatory stimuli induce accumulation of MHC class II complexes on dendritic cells. *Nature* **388**, 782–787 (1997).
117. De Gassart, A. *et al.* MHC class II stabilization at the surface of human dendritic cells is the result of maturation-dependent MARCH I down-regulation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3491–3496 (2008).
118. Hava, D. L. *et al.* Evasion of peptide, but not lipid antigen presentation, through pathogen-induced dendritic cell maturation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 11281–11286 (2008).
119. Kasai, M. *et al.* In vivo effect of anti-asialo GM1 antibody on natural killer activity. *Nature* **291**, 334–335 (1981).
120. Moretta, L., Biassoni, R., Bottino, C., Mingari, M. C. & Moretta, A. Human NK-cell receptors. *Immunol. Today* **21**, 420–422 (2000).
121. Pende, D. *et al.* Identification and molecular characterization of NKp30, a novel triggering receptor involved in natural cytotoxicity mediated by human natural killer cells. *J. Exp. Med.* **190**, 1505–1516 (1999).
122. Sivori, S. *et al.* p46, a novel natural killer cell-specific surface molecule that mediates cell activation. *J. Exp. Med.* **186**, 1129–1136 (1997).
123. Vitale, M. *et al.* NKp44, a Novel Triggering Surface Molecule Specifically Expressed by Activated Natural Killer Cells, Is Involved in Non-Major Histocompatibility Complex-restricted Tumor Cell Lysis. *J. Exp. Med.* **187**, 2065–2072 (1998).

124. Denis, M. Interleukin-12 (IL-12) augments cytolytic activity of natural killer cells toward Mycobacterium tuberculosis-infected human monocytes. *Cell. Immunol.* **156**, 529–536 (1994).
125. Voskoboinik, I., Dunstone, M. A., Baran, K., Whisstock, J. C. & Trapani, J. A. Perforin: structure, function, and role in human immunopathology. *Immunol. Rev.* **235**, 35–54 (2010).
126. Yin, X. M. *et al.* Bid-deficient mice are resistant to Fas-induced hepatocellular apoptosis. *Nature* **400**, 886–891 (1999).
127. Li, H., Zhu, H., Xu, C. J. & Yuan, J. Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell* **94**, 491–501 (1998).
128. Afonina, I. S., Cullen, S. P. & Martin, S. J. Cytotoxic and non-cytotoxic roles of the CTL/NK protease granzyme B. *Immunol. Rev.* **235**, 105–116 (2010).
129. Krensky, A. M. Granulysin: a novel antimicrobial peptide of cytolytic T lymphocytes and natural killer cells. *Biochem. Pharmacol.* **59**, 317–320 (2000).
130. Krensky, A. M. & Clayberger, C. Biology and clinical relevance of granulysin. *Tissue Antigens* **73**, 193–198 (2009).
131. Okada, S., Li, Q., Whitin, J. C., Clayberger, C. & Krensky, A. M. Intracellular mediators of granulysin-induced cell death. *J. Immunol. Baltim. Md 1950* **171**, 2556–2562 (2003).
132. Bao, Q. & Shi, Y. Apoptosome: a platform for the activation of initiator caspases. *Cell Death Differ.* **14**, 56–65 (2007).
133. Oddo, M. *et al.* Fas ligand-induced apoptosis of infected human macrophages reduces the viability on intracellular Mycobacterium tuberculosis. *J Immunol* **160**, 5448–5454 (1998).
134. Alderson, M. R. *et al.* CD40 expression by human monocytes: regulation by cytokines and activation of monocytes by the ligand for CD40. *J. Exp. Med.* **178**, 669–674 (1993).
135. Carbone, E. *et al.* A new mechanism of NK cell cytotoxicity activation: the CD40-CD40 ligand interaction. *J. Exp. Med.* **185**, 2053–2060 (1997).

136. Wang, R., Jaw, J. J., Stutzman, N. C., Zou, Z. & Sun, P. D. Natural killer cell-produced IFN- $\gamma$  and TNF- $\alpha$  induce target cell cytolysis through up-regulation of ICAM-1. *J. Leukoc. Biol.* **91**, 299–309 (2012).
137. Vankayalapati, R. *et al.* NK cells regulate CD8<sup>+</sup> T cell effector function in response to an intracellular pathogen. *J. Immunol. Baltim. Md 1950* **172**, 130–137 (2004).
138. Portevin, D., Via, L. E., Eum, S. & Young, D. Natural killer cells are recruited during pulmonary tuberculosis and their ex vivo responses to mycobacteria vary between healthy human donors in association with KIR haplotype. *Cell. Microbiol.* **14**, 1734–1744 (2012).
139. Esin, S. *et al.* Interaction of Mycobacterium tuberculosis cell wall components with the human natural killer cell receptors NKp44 and Toll-like receptor 2. *Scand. J. Immunol.* **77**, 460–469 (2013).
140. Lu, C.-C. *et al.* NK cells kill mycobacteria directly by releasing perforin and granulysin. *J. Leukoc. Biol.* **96**, 1119–1129 (2014).
141. Khader, S. A. *et al.* Interleukin 12p40 is required for dendritic cell migration and T cell priming after Mycobacterium tuberculosis infection. *J. Exp. Med.* **203**, 1805–1815 (2006).
142. Schaible, U. E. *et al.* Apoptosis facilitates antigen presentation to T lymphocytes through MHC-I and CD1 in tuberculosis. *Nat. Med.* **9**, 1039–1046 (2003).
143. Ngai, P. *et al.* Gamma interferon responses of CD4 and CD8 T-cell subsets are quantitatively different and independent of each other during pulmonary Mycobacterium bovis BCG infection. *Infect. Immun.* **75**, 2244–2252 (2007).
144. Kaufmann, S. H. & McMichael, A. J. Annulling a dangerous liaison: vaccination strategies against AIDS and tuberculosis. *Nat. Med.* **11**, S33–S44 (2005).
145. Wang, J., Santosuosso, M., Ngai, P., Zganiacz, A. & Xing, Z. Activation of CD8 T cells by mycobacterial vaccination protects against pulmonary tuberculosis in the absence of CD4 T cells. *J. Immunol. Baltim. Md 1950* **173**, 4590–4597 (2004).
146. Kaufmann, S. H. E. & Parida, S. K. Tuberculosis in Africa: learning from pathogenesis for biomarker identification. *Cell Host Microbe* **4**, 219–228 (2008).

147. Dorhoi, A. & Kaufmann, S. H. E. Fine-tuning of T cell responses during infection. *Curr. Opin. Immunol.* **21**, 367–377 (2009).
148. Trinchieri, G. Regulatory role of T cells producing both interferon gamma and interleukin 10 in persistent infection. *J. Exp. Med.* **194**, F53-57 (2001).
149. Joosten, S. A. & Ottenhoff, T. H. M. Human CD4 and CD8 regulatory T cells in infectious diseases and vaccination. *Hum. Immunol.* **69**, 760–770 (2008).
150. Dheda, K., Schwander, S. K., Zhu, B., van Zyl-Smit, R. N. & Zhang, Y. The immunology of tuberculosis: from bench to bedside. *Respirol. Carlton Vic* **15**, 433–450 (2010).
151. Abebe, F. & Bjune, G. The protective role of antibody responses during Mycobacterium tuberculosis infection. *Clin. Exp. Immunol.* **157**, 235–243 (2009).
152. Maglione, P. J. & Chan, J. How B cells shape the immune response against Mycobacterium tuberculosis. *Eur. J. Immunol.* **39**, 676–686 (2009).
153. Carroll, M. V., Lack, N., Sim, E., Krarup, A. & Sim, R. B. Multiple routes of complement activation by Mycobacterium bovis BCG. *Mol. Immunol.* **46**, 3367–3378 (2009).
154. Cottle, L. E. Mendelian susceptibility to mycobacterial disease. *Clin. Genet.* **79**, 17–22 (2011).
155. Wetering, D. van de, Paus, R. A. de, Dissel, J. T. van & Vosse, E. van de. IL-23 modulates CD56<sup>+</sup>/CD3<sup>-</sup> NK cell and CD56<sup>+</sup>/CD3<sup>+</sup> NK-like T cell function differentially from IL-12. *Int. Immunol.* **21**, 145–153 (2009).
156. Vosse, E. van de, Dissel, J. T. van & Ottenhoff, T. H. Genetic deficiencies of innate immune signalling in human infectious disease. *Lancet Infectious Dis.* **9**, 688–698 (2009).
157. Parham, C. *et al.* A receptor for the heterodimeric cytokine IL-23 is composed of IL-12Rbeta1 and a novel cytokine receptor subunit, IL-23R. *J. Immunol. Baltim. Md 1950* **168**, 5699–5708 (2002).
158. Paus, R. A. de, Wetering, D. van de, Dissel, J. T. van & Vosse, E. van de. IL-23 and IL-12 responses in activated human T cells retrovirally transduced with IL-23 receptor variants. *Mol. Immunol.* **45**, 3889–3895 (2008).

159. McGeachy, M. J. & Cua, D. J. Th17 cell differentiation: the long and winding road. *Immunity* **28**, 445–453 (2008).
160. Oppmann, B. *et al.* Novel p19 protein engages IL-12p40 to form a cytokine, IL-23, with biological activities similar as well as distinct from IL-12. *Immunity* **13**, 715–725 (2000).
161. Zheng, Y. *et al.* Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nat. Med.* **14**, 282–289 (2008).
162. Aujla, S. J. *et al.* IL-22 mediates mucosal host defense against Gram-negative bacterial pneumonia. *Nat. Med.* **14**, 275–281 (2008).
163. Pflanz, S. *et al.* IL-27, a heterodimeric cytokine composed of EB13 and p28 protein, induces proliferation of naive CD4<sup>+</sup> T cells. *Immunity* **16**, 779–790 (2002).
164. Smits, H. H. *et al.* Commensal Gram-negative bacteria prime human dendritic cells for enhanced IL-23 and IL-27 expression and enhanced Th1 development. *Eur. J. Immunol.* **34**, 1371–1380 (2004).
165. Yoshimura, T. *et al.* Two-sided roles of IL-27: induction of Th1 differentiation on naive CD4<sup>+</sup> T cells versus suppression of proinflammatory cytokine production including IL-23-induced IL-17 on activated CD4<sup>+</sup> T cells partially through STAT3-dependent mechanism. *J. Immunol. Baltim. Md 1950* **177**, 5377–5385 (2006).
166. Holscher, C. *et al.* The IL-27 receptor chain WSX-1 differentially regulates antibacterial immunity and survival during experimental tuberculosis. *J. Immunol. Baltim. Md 1950* **174**, 3534–3544 (2005).
167. Gengenbacher, M. & Kaufmann, S. H. Mycobacterium tuberculosis: success through dormancy. *FEMS Microbiol. Rev.* **36**, 514–532 (2012).
168. Qu, H. Q., Fisher-Hoch, S. P. & McCormick, J. B. Molecular immunity to mycobacteria: knowledge from the mutation and phenotype spectrum analysis of Mendelian susceptibility to mycobacterial diseases. *Int. J. Infect. Dis. IJID Off. Publ. Int. Soc. Infect. Dis.* **15**, e305–13 (2011).

169. Zumla, A., Raviglione, M., Hafner, R. & Reyn, C. F. von. Tuberculosis. *N. Engl. J. Med.* **368**, 745–755 (2013).
170. Cruz, A. T. & Starke, J. R. Clinical manifestations of tuberculosis in children. *Paediatr. Respir. Rev.* **8**, 107–117 (2007).
171. SM, N., AJ, B., S, A., E, W. & B, K. Paediatric Tuberculosis. *Lancet Infect. Dis.* **8**, 498–510 (2008).
172. Boisson-Dupuis, S. *et al.* Inherited and acquired immunodeficiencies underlying tuberculosis in childhood. *Immunol. Rev.* **264**, 103–120 (2015).
173. Kaufmann, S. H. How can immunology contribute to the control of tuberculosis? *Nat. Rev.* **1**, 20–30 (2001).
174. Schloß, M., Heckrodt, J., Schneider, C., Discher, T. & Krombach, G. A. Magnetic Resonance Imaging of the Lung as an Alternative for a Pregnant Woman with Pulmonary Tuberculosis. *J. Radiol. Case Rep.* **9**, 7–13 (2015).
175. Casanova, J. L. & Abel, L. Genetic dissection of immunity to mycobacteria: the human model. *Annu. Rev. Immunol.* **20**, 581–620 (2002).
176. Dubos, R. J. Discussion on treatment of tuberculous meningitis and survival of bacilli in tuberculous lesions. *Am. Rev. Tuberc.* **65**, 637–640 (1952).
177. Puffer, R. in 106 (MA: Harvard University Press., 1944).
178. Kirenga, B. J. *et al.* Tuberculosis risk factors among tuberculosis patients in Kampala, Uganda: implications for tuberculosis control. *BMC Public Health* **15**, 13 (2015).
179. Sreedharan, A. Immunology of tuberculosis. (2013).
180. Marquet, S. & Schurr, E. Genetics of susceptibility to infectious diseases: tuberculosis and leprosy as examples. *Drug Metab. Dispos. Biol. Fate Chem.* **29**, 479–483 (2001).
181. Roh, E. Y., Yoon, J. H., Shin, S., Song, E. Y. & Park, M. H. Association of TAP1 and TAP2 genes with susceptibility to pulmonary tuberculosis in Koreans. *APMIS Acta Pathol. Microbiol. Immunol. Scand.* (2015).

182. Juarez-Ortega, M. *et al.* Induction and treatment of anergy in murine leprosy. *Int. J. Exp. Pathol.* **96**, 31–41 (2015).
183. Comstock, G. W. Tuberculosis in twins: a re-analysis of the Proffit survey. *Am. Rev. Respir. Dis.* **117**, 621–624 (1978).
184. Kallmann, F. & Reisner, D. Twin studies on the significance of genetic factors in tuberculosis. *Am Rev Tuberc* **47**, 549–574 (1943).
185. Jepson, A. *et al.* Genetic regulation of acquired immune responses to antigens of *Mycobacterium tuberculosis*: a study of twins in West Africa. *Infect. Immun.* **69**, 3989–3994 (2001).
186. Stein, C. M. *et al.* Heritability analysis of cytokines as intermediate phenotypes of tuberculosis. *J. Infect. Dis.* **187**, 1679–1685 (2003).
187. Stead, W. W. Genetics and resistance to tuberculosis. Could resistance be enhanced by genetic engineering? *Ann. Intern. Med.* **116**, 937–941 (1992).
188. Stead, W. W. The origin and erratic global spread of tuberculosis. How the past explains the present and is the key to the future. *Clin. Chest Med.* **18**, 65–77 (1997).
189. Stead, W. W., Senner, J. W., Reddick, W. T. & Lofgren, J. P. Racial differences in susceptibility to infection by *Mycobacterium tuberculosis*. *N.Engl.J.Med.* **322**, 422–427 (1990).
190. Young, D. B., Perkins, M. D., Duncan, K. & 3rd, C. E. B. Confronting the scientific obstacles to global control of tuberculosis. *J. Clin. Invest.* **118**, 1255–1265 (2008).
191. Gagneux, S. Host-pathogen coevolution in human tuberculosis. *Philos. Trans. R. Soc. London Series B Biol. Sci.* **367**, 850–859 (2012).
192. Gagneux, S. *et al.* Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2869–2873 (2006).
193. Jong, B. C. de *et al.* Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in The Gambia. *J. Infect. Dis.* **198**, 1037–1043 (2008).

194. Manca, C. *et al.* Virulence of a Mycobacterium tuberculosis clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha /beta. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 5752–5757 (2001).
195. Nicol, M. P. & Wilkinson, R. J. The clinical consequences of strain diversity in Mycobacterium tuberculosis. *Trans. R. Soc. Trop. Med. Hyg.* **102**, 955–965 (2008).
196. Alcais, A., Fieschi, C., Abel, L. & Casanova, J. L. Tuberculosis in children and adults: two distinct genetic diseases. *J. Exp. Med.* **202**, 1617–1621 (2005).
197. Alcais, A. *et al.* Life-threatening infectious diseases of childhood: single-gene inborn errors of immunity? *Ann. N. Y. Acad. Sci.* **1214**, 18–33 (2010).
198. Apt, A. & Kramnik, I. Man and mouse TB: contradictions and solutions. *Tuberc. Edinb. Scotl.* **89**, 195–198 (2009).
199. Fortin, A., Abel, L., Casanova, J. L. & Gros, P. Host genetics of mycobacterial diseases in mice and men: forward genetic studies of BCG-osis and tuberculosis. *Annu. Rev. Genomics Hum. Genet.* **8**, 163–192 (2007).
200. Cooper, A. M. Cell-mediated immune responses in tuberculosis. *Annu. Rev. Immunol.* **27**, 393–422 (2009).
201. Suthar, A. B. *et al.* Antiretroviral therapy for prevention of tuberculosis in adults with HIV: a systematic review and meta-analysis. *PLoS Med.* **9**, e1001270 (2012).
202. Keane, J. *et al.* Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. *N. Engl. J. Med.* **345**, 1098–1104 (2001).
203. Flynn, J. L. & Chan, J. Immunology of tuberculosis. *Annu. Rev. Immunol.* **19**, 93–129 (2001).
204. North, R. J. & Jung, Y. J. Immunity to tuberculosis. *Annu. Rev. Immunol.* **22**, 599–623 (2004).
205. Flynn, J. L. *et al.* Tumor necrosis factor-alpha is required in the protective immune response against Mycobacterium tuberculosis in mice. *Immunity.* **2**, 561–572 (1995).

206. Lienhardt, C. *et al.* Investigation of the risk factors for tuberculosis: a case-control study in three countries in West Africa. *Int. J. Epidemiol.* **34**, 914–923 (2005).
207. Alcais, A., Remus, N., El Baghdadi, J., Abel, L. & Casanova, J. L. Genetic susceptibility to tuberculosis: from monogenic to polygenic inheritance. *Sepsis* **4**, 237–246 (2001).
208. Boisson, B., Quartier, P. & Casanova, J.-L. Immunological loss-of-function due to genetic gain-of-function in humans: autosomal dominance of the third kind. *Curr. Opin. Immunol.* **32**, 90–105 (2015).
209. Al-Herz, W. *et al.* Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency. *Front. Immunol.* **5**, 162 (2014).
210. Geha, R. S. *et al.* Primary immunodeficiency diseases: an update from the International Union of Immunological Societies Primary Immunodeficiency Diseases Classification Committee. *J. Allergy Clin. Immunol.* **120**, 776–794 (2007).
211. Valotti, M. *et al.* Long-lasting production of new T and B cells and T-cell repertoire diversity in patients with primary immunodeficiency who had undergone stem cell transplantation: a single-centre experience. *J. Immunol. Res.* **2014**, 240453 (2014).
212. Reichenbach, J. *et al.* Mycobacterial diseases in primary immunodeficiencies. *Curr. Opin. Allergy Clin. Immunol.* **1**, 503–511 (2001).
213. ESID database statistics. In: European Society for Immunodeficiencies. 2011. <http://www.esid.org/statistics.php>. Accessed.
214. Leiva, LE, Zelazco, M, Oleastro, M, Carneiro-Sampaio, M, Condino-Neto, A, Tavares Costa-Carvalho, B (2007) Primary immun.
215. African Society for Immunodeficiencies ASID. 2009. <http://www.asid.ma>. Accessed 11 Sept 2011.
216. Joshi, AY, Iyer, VN, Hagan, JB, Sauver, JLS, Boyce, TG (2009) Incidence and temporal trends of primary immunodeficiency:

217. Boyle, JM, Buckley, RH (2007) Population prevalence of diagnosed primary immunodeficiency diseases in the United States.
218. Bousfiha, A. A. *et al.* Primary immunodeficiency diseases worldwide: more common than generally thought. *J. Clin. Immunol.* **33**, 1–7 (2013).
219. Total number of patients in the USIDnet registry. In: US Immunodeficiency Network. 2004. <http://www.usidnet.org/index.cf>.
220. Barbouche, MR, Galal, N, Ben Mustapha, I, Jeddane, L, Mellouli, F, Ailal, F (2011) Primary immunodeficiencies in highly.
221. Ishimura, M, Takada, H, Doi, T, Imai, K, Sasahara, Y, Kanegane, H (2011) Nationwide survey of patients with primary immu.
222. Rezaei, N, Aghamohammadi, A, Moin, M, Pourpak, Z, Movahedi, M, Gharagozlou, M (2006) Frequency and clinical manifestatio.
223. Kirkpatrick, P, Riminton, S (2007) Primary immunodeficiency diseases in Australia and New Zealand. *J Clin Immunol* 27: pp.
224. Eley, B. & Esser, M. Investigation and management of primary immunodeficiency in South African children. *SAMJ South Afr. Med. J.* **104**, 793–793 (2014).
225. Esser, M., Banda, E., Moller, M. & Nortje, R. PRIMARY IMMUNODEFICIENCY DISEASE MANAGEMENT IN TUBERCULOSIS ENDEMIC REGIONS – ARE WE AWARE ENOUGH AND HOW DOES A REGISTRY ASSIST. (2015).
226. Filipe-Santos, O. *et al.* Inborn errors of IL-12/23- and IFN-gamma-mediated immunity: molecular, cellular, and clinical features. *Semin. Immunol.* **18**, 347–361 (2006).
227. Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33**, D514-7 (2005).
228. Boisson-Dupuis, S. *et al.* IL-12Rbeta1 deficiency in two of fifty children with severe tuberculosis from Iran, Morocco, and Turkey. *PloS One* **6**, e18524 (2011).

229. Tabarsi, P. *et al.* Lethal tuberculosis in a previously healthy adult with IL-12 receptor deficiency. *J. Clin. Immunol.* **31**, 537–539 (2011).
230. Dorman, E. *et al.* Clinical features of dominant and recessive interferon gamma receptor 1 deficiencies. *Lancet* **364**, 2113–2121 (2004).
231. Picard, C. *et al.* Inherited interleukin-12 deficiency: IL12B genotype and clinical phenotype of 13 patients from six kindreds. *Am. J. Hum. Genet.* **70**, 336–348 (2002).
232. Sasaki, Y. *et al.* Genetic basis of patients with bacille Calmette-Guérin osteomyelitis in Japan: identification of dominant partial interferon-gamma receptor 1 deficiency as a predominant type. *J. Infect. Dis.* **185**, 706–709 (2002).
233. Bustamante, J. *et al.* Novel primary immunodeficiencies revealed by the investigation of paediatric infectious diseases. *Curr. Opin. Immunol.* **20**, 39–48 (2008).
234. Patel, S. Y., Doffinger, R., Barcenas-Morales, G. & Kumararatne, D. S. Genetically determined susceptibility to mycobacterial infection. *J. Clin. Pathol.* **61**, 1006–1012 (2008).
235. Cooke, G. S. *et al.* Polymorphism within the interferon-gamma/receptor complex is associated with pulmonary tuberculosis. *Am. J. Respir. Crit. Care Med.* **174**, 339–343 (2006).
236. Jouanguy, E. *et al.* Partial interferon-gamma receptor 1 deficiency in a child with tuberculoïd bacillus Calmette-Guérin infection and a sibling with clinical tuberculosis. *J. Clin. Invest.* **100**, 2658–2664 (1997).
237. Stein, C. M. *et al.* Linkage and association analysis of candidate genes for TB and TNF $\alpha$  cytokine expression: evidence for association with IFNGR1, IL-10, and TNF receptor 1 genes. *Hum. Genet.* **121**, 663–673 (2007).
238. Storgaard, M., Varming, K., Herlin, T. & Obel, N. Novel mutation in the interferon-gamma-receptor gene and susceptibility to mycobacterial infections. *Scand. J. Immunol.* **64**, 137–139 (2006).

239. Dorman, S. E. & Holland, S. M. Mutation in the signal-transducing chain of the interferon-gamma receptor and susceptibility to mycobacterial infection. *J. Clin. Invest.* **101**, 2364–2369 (1998).
240. Cooper, A. M. *et al.* Mice lacking bioactive IL-12 can generate protective, antigen-specific cellular responses to mycobacterial infection only if the IL-12 p40 subunit is present. *J. Immunol. Baltim. Md 1950* **168**, 1322–1327 (2002).
241. Akahoshi, M. *et al.* Influence of interleukin-12 receptor beta1 polymorphisms on tuberculosis. *Hum. Genet.* **112**, 237–243 (2003).
242. Altare, F. *et al.* Interleukin-12 receptor beta1 deficiency in a patient with abdominal tuberculosis. *J. Infect. Dis.* **184**, 231–236 (2001).
243. Fieschi, C. *et al.* Low penetrance, broad resistance, and favorable outcome of interleukin 12 receptor beta1 deficiency: medical and immunological implications. *J. Exp. Med.* **197**, 527–535 (2003).
244. Ozbek, N. *et al.* Interleukin-12 receptor beta 1 chain deficiency in a child with disseminated tuberculosis. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **40**, e55-58 (2005).
245. Robinson, R. T. *et al.* Mycobacterium tuberculosis infection induces il12rb1 splicing to generate a novel IL-12Rbeta1 isoform that enhances DC migration. *J. Exp. Med.* **207**, 591–605 (2010).
246. Chapgier, A. *et al.* Human complete Stat-1 deficiency is associated with defective type I and II IFN responses in vitro but immunity to some low virulence viruses in vivo. *J. Immunol. Baltim. Md 1950* **176**, 5078–5083 (2006).
247. Tsumura, M. *et al.* Dominant-negative STAT1 SH2 domain mutations in unrelated patients with Mendelian susceptibility to mycobacterial disease. *Hum. Mutat.* **33**, 1377–1387 (2012).
248. Marquis, J. F., LaCourse, R., Ryan, L., North, R. J. & Gros, P. Disseminated and rapidly fatal tuberculosis in mice bearing a defective allele at IFN regulatory factor 8. *J. Immunol. Baltim. Md 1950* **182**, 3008–3015 (2009).

249. Bogunovic, D. *et al.* Mycobacterial disease and impaired IFN-gamma immunity in humans with inherited ISG15 deficiency. *Science* **337**, 1684–1688 (2012).
250. Zhang, X. *et al.* Human intracellular ISG15 prevents interferon-alpha/beta over-amplification and auto-inflammation. *Nature* **517**, 89–93 (2015).
251. Filipe-Santos, O. *et al.* X-linked susceptibility to mycobacteria is caused by mutations in NEMO impairing CD40-dependent IL-12 production. *J. Exp. Med.* **203**, 1745–1759 (2006).
252. Francesca Fusco *et al.* Incontinentia pigmenti: report on data from 2000 to 2013. *Orphanet J. Rare Dis.* **9**, (2014).
253. Schaller, J. Illness resembling lupus erythematosus in mothers of boys with chronic granulomatous disease. *Ann. Intern. Med.* **76**, 747–750 (1972).
254. Al-Muhsen, S. & Casanova, J. L. The genetic heterogeneity of mendelian susceptibility to mycobacterial diseases. *J. Allergy Clin. Immunol.* **122**, 1043-51–3 (2008).
255. Kilic, S. S. *et al.* A patient with tyrosine kinase 2 deficiency without hyper-IgE syndrome. *J. Pediatr.* **160**, 1055–1057 (2012).
256. Minegishi, Y. *et al.* Human tyrosine kinase 2 deficiency reveals its requisite roles in multiple cytokine signals involved in innate and acquired immunity. *Immunity* **25**, 745–755 (2006).
257. Woellner, C. *et al.* The hyper IgE syndrome and mutations in TYK2. *Immunity* **26**, 535; author reply 536 (2007).
258. Dupuis, S. *et al.* Impairment of mycobacterial but not viral immunity by a germline human STAT1 mutation. *Science* **293**, 300–303 (2001).
259. Altare, F. *et al.* Interleukin-12 receptor beta1 deficiency in a patient with abdominal tuberculosis. *J. Infect. Dis.* **184**, 231–236 (2001).
260. Caragol, I. *et al.* Clinical tuberculosis in 2 of 3 siblings with interleukin-12 receptor beta1 deficiency. *Clin. Infect. Dis.* **37**, 302–306 (2003).
261. Ozbek, N. *et al.* Interleukin-12 receptor beta 1 chain deficiency in a child with disseminated tuberculosis. *Clin. Infect. Dis.* **40**, e55–e58 (2005).

262. Jouanguy, E. *et al.* Partial interferon-gamma receptor 1 deficiency in a child with tuberculoid bacillus Calmette-Guerin infection and a sibling with clinical tuberculosis. *J.Clin.Invest* **100**, 2658–2664 (1997).
263. Picard, C. *et al.* Inherited interleukin-12 deficiency: IL12B genotype and clinical phenotype of 13 patients from six kindreds. *Am.J.Hum.Genet.* **70**, 336–348 (2002).
264. Chapgier, A. *et al.* Human complete Stat-1 deficiency is associated with defective type I and II IFN responses in vitro but immunity to some low virulence viruses in vivo. *J. Immunol. Baltim. Md 1950* **176**, 5078–5083 (2006).
265. Filipe-Santos, O. *et al.* X-linked susceptibility to mycobacteria is caused by mutations in NEMO impairing CD40-dependent IL-12 production. *J.Exp.Med.* **203**, 1745–1759 (2006).
266. Hoshina, T. *et al.* Clinical and host genetic characteristics of Mendelian susceptibility to mycobacterial diseases in Japan. *J. Clin. Immunol.* **31**, 309–314 (2011).
267. Casanova, J.-L. & Abel, L. Inborn errors of immunity to infection: the rule rather than the exception. *J. Exp. Med.* **202**, 197–201 (2005).
268. Dhawan, S. R. *et al.* Predictors of Neurological Outcome of Tuberculous Meningitis in Childhood A Prospective Cohort Study From a Developing Country. *J. Child Neurol.* 883073816668112 (2016). doi:10.1177/0883073816668112
269. Rali, P., Arshad, H. & Bihler, E. A Case of Tuberculous Meningitis with Tuberculoma in Nonimmunocompromised Immigrant. *Case Rep. Pulmonol.* **2016**, (2016).
270. Marx, G. E. & Chan, E. D. Tuberculous Meningitis: Diagnosis and Treatment Overview. *Tuberc. Res. Treat.* **2011**, e798764 (2011).
271. Koplay, M., Erdogan, H., Sivri, M., Nayman, A. & Paksoy, Y. Unusual reason of spinal cord infarction: tuberculous meningitis. *Acta Neurol. Belg.* (2015). doi:10.1007/s13760-015-0507-z
272. Garg, R. K. Tuberculosis of the central nervous system. *Postgrad. Med. J.* **75**, 133–140 (1999).

273. Thuong, N. T. *et al.* A polymorphism in human TLR2 is associated with increased susceptibility to tuberculous meningitis. *Genes Immun* **8**, 422–428 (2007).
274. Dissanayeke, S. R. *et al.* Polymorphic variation in TIRAP is not associated with susceptibility to childhood TB but may determine susceptibility to TBM in some ethnic groups. *PLoS ONE*. **4**, e6698 (2009).
275. Brancusi, F., Farrar, J. & Heemskerck, D. Tuberculous meningitis in adults: a review of a decade of developments focusing on prognostic factors for outcome. *Future Microbiol.* **7**, 1101–1116 (2012).
276. Al-Herz, W. *et al.* Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency. *Front. Immunol.* **2**, 54 (2011).
277. van Toorn, R., Schaaf, H. S., Solomons, R., Laubscher, J. A. & Schoeman, J. F. The value of transcranial Doppler imaging in children with tuberculous meningitis. *Childs Nerv. Syst. ChNS Off. J. Int. Soc. Pediatr. Neurosurg.* **30**, 1711–1716 (2014).
278. von Bezing, H., Andronikou, S., van Toorn, R. & Douglas, T. Are linear measurements and computerized volumetric ratios determined from axial MRI useful for diagnosing hydrocephalus in children with tuberculous meningitis? *Childs Nerv. Syst. ChNS Off. J. Int. Soc. Pediatr. Neurosurg.* **28**, 79–85 (2012).
279. Munch, Z. *et al.* Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. *Int. J. Tuberc. Lung Dis.* **7**, 271–277 (2003).
280. Beyers, N. *et al.* The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high-incidence community. *S.Afr.Med.J.* **86**, 40–1, 44 (1996).
281. Casanova, J.-L. Severe infectious diseases of childhood as monogenic inborn errors of immunity. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E7128–E7137 (2015).
282. Babraham Institute. Babraham Bioinformatics. *FastQC* Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

283. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
284. sourceforge. Picard. *Picard* Available at: <http://picard.sourceforge.net>.
285. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
286. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
287. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
288. Phillips, T. How Genetic Polymorphism Promotes Diversity and Lasts Over Generations. *The Balance* Available at: <https://www.thebalance.com/genetic-polymorphism-what-is-it-375594>. (Accessed: 6th December 2016)
289. Reference, G. H. What is a gene mutation and how do mutations occur? *Genetics Home Reference* Available at: <https://ghr.nlm.nih.gov/primer/mutationsanddisorders/genemutation>. (Accessed: 6th December 2016)
290. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
291. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
292. Gnad, F., Baucom, A., Mukhyala, K., Manning, G. & Zhang, Z. Assessment of computational methods for predicting the effects of missense mutations in human cancers. *BMC Genomics* **14 Suppl 3**, S7 (2013).
293. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).

294. Venselaar, H., Te Beek, T. A. H., Kuipers, R. K. P., Hekkelman, M. L. & Vriend, G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* **11**, 548 (2010).
295. Alipoor, B. *et al.* A Bioinformatics Approach to Prioritize Single Nucleotide Polymorphisms in TLRs Signaling Pathway Genes. *Int. J. Mol. Cell. Med.* **5**, 65–79 (2016).
296. Hassan, M. M. *et al.* Bioinformatics Approach for Prediction of Functional Coding/Noncoding Simple Polymorphisms (SNPs/Indels) in Human BRAF Gene. *Adv. Bioinforma.* **2016**, (2016).
297. de Voer, R. M. *et al.* Identification of Novel Candidate Genes for Early-Onset Colorectal Cancer Susceptibility. *PLoS Genet.* **12**, (2016).
298. Atik, T. *et al.* Novel MASP1 mutations are associated with an expanded phenotype in 3MC1 syndrome. *Orphanet J. Rare Dis.* **10**, (2015).
299. Costa, V. *et al.* Computational Analysis of Single Nucleotide Polymorphisms Associated with Altered Drug Responsiveness in Type 2 Diabetes. *Int. J. Mol. Sci.* **17**, (2016).
300. Akhoundi, F., Parvaneh, N. & Modjtaba, E.-B. In silico analysis of deleterious single nucleotide polymorphisms in human BUB1 mitotic checkpoint serine/threonine kinase B gene. *Meta Gene* **9**, 142–150 (2016).
301. Stoll, M. *et al.* Novel motor phenotypes in patients with VRK1 mutations without pontocerebellar hypoplasia. *Neurology* **87**, 65–70 (2016).
302. Kinnersley, B. *et al.* Search for new loci and low-frequency variants influencing glioma risk by exome-array analysis. *Eur. J. Hum. Genet.* **24**, 717–724 (2016).
303. Venselaar, H., te Beek, T. A., Kuipers, R. K., Hekkelman, M. L. & Vriend, G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* **11**, 548 (2010).
304. Venselaar, H., Te Beek, T. A. H., Kuipers, R. K. P., Hekkelman, M. L. & Vriend, G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* **11**, 548 (2010).

305. Schwarz, J. M., Rödelberger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
306. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361–362 (2014).
307. Schmidt, S. *et al.* Hypermutable non-synonymous sites are under stronger negative selection. *PLoS Genet.* **4**, e1000281 (2008).
308. Bromberg, Y., Yachdav, G. & Rost, B. SNAP predicts effect of mutations on protein function. *Bioinforma. Oxf. Engl.* **24**, 2397–2398 (2008).
309. Yue, P., Melamud, E. & Moulton, J. SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics* **7**, 166 (2006).
310. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
311. Ng, P. C. & Henikoff, S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res.* **12**, 436–446 (2002).
312. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
313. Kim, T. K. & Eberwine, J. H. Mammalian cell transfection: the present and the future. *Anal. Bioanal. Chem.* **397**, 3173–3178 (2010).
314. Sarkar, S., Davies, J. E., Huang, Z., Tunnacliffe, A. & Rubinsztein, D. C. Trehalose, a novel mTOR-independent autophagy enhancer, accelerates the clearance of mutant huntingtin and alpha-synuclein. *J. Biol. Chem.* **282**, 5641–5652 (2007).
315. Ficoll density gradient of leukopac for the separation of PBMCs. : In vivo antigen-driven plasmablast enrichment in combination with antigen-specific cell sorting to facilitate the isolation of rare monoclonal antibodies from human B cells : Nature Protocols : Nature Publishing Group. Available at:  
[http://www.nature.com/nprot/journal/v9/n7/fig\\_tab/nprot.2014.104\\_F5.html](http://www.nature.com/nprot/journal/v9/n7/fig_tab/nprot.2014.104_F5.html). (Accessed: 5th November 2015)

316. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
317. Daya, M. *et al.* A Panel of Ancestry Informative Markers for the Complex Five-Way Admixed South African Coloured Population. *PLoS ONE* **8**, e82224 (2013).
318. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
319. Yan, P. *et al.* Targeting autophagic regulation of NFkappaB in HTLV-I transformed cells by geldanamycin: implications for therapeutic interventions. *Autophagy* **3**, 600–603 (2007).
320. Qing, G., Yan, P., Qu, Z., Liu, H. & Xiao, G. Hsp90 regulates processing of NF-kappa B2 p100 involving protection of NF-kappa B-inducing kinase (NIK) from autophagy-mediated degradation. *Cell Res.* **17**, 520–530 (2007).
321. Nivon, M., Richet, E., Codogno, P., Arrigo, A.-P. & Kretz-Remy, C. Autophagy activation by NFkappaB is essential for cell survival after heat shock. *Autophagy* **5**, 766–783 (2009).
322. Loos, B., Hofmeyr, J.-H. S., Müller-Nedebock, K., Boonzaaier, L. & Kinnear, C. in *Autophagy: Cancer, Other Pathologies, Inflammation, Immunity, Infection, and Aging* 39–56 (Elsevier, 2014).
323. Loos, B., du Toit, A. & Hofmeyr, J.-H. S. Defining and measuring autophagosome flux—concept and reality. *Autophagy* **10**, 2087–2096 (2014).
324. Kirkegaard, K., Taylor, M. P. & Jackson, W. T. Cellular autophagy: surrender, avoidance and subversion by microorganisms. *Nat. Rev. Microbiol.* **2**, 301–314 (2004).
325. Klionsky, D. J., Cuervo, A. M. & Seglen, P. O. Methods for monitoring autophagy from yeast to human. *Autophagy* **3**, 181–206 (2007).
326. Mizushima, N. Methods for monitoring autophagy. *Int. J. Biochem. Cell Biol.* **36**, 2491–2502 (2004).
327. Kabeya, Y. *et al.* LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. *EMBO J.* **19**, 5720–5728 (2000).

328. Kabeya, Y. *et al.* LC3, GABARAP and GATE16 localize to autophagosomal membrane depending on form-II formation. *J. Cell Sci.* **117**, 2805–2812 (2004).
329. Mizushima, N. & Yoshimori, T. How to interpret LC3 immunoblotting. *Autophagy* **3**, 542–545 (2007).
330. Conley, M. E. & Casanova, J.-L. Discovery of single-gene inborn errors of immunity by next generation sequencing. *Curr. Opin. Immunol.* **30**, 17–23 (2014).
331. Pabinger, S. *et al.* A survey of tools for variant analysis of next-generation genome sequencing data. *Brief. Bioinform.* **15**, 256–278 (2014).
332. Johar, A. S. *et al.* Candidate gene discovery in autoimmunity by using extreme phenotypes, next generation sequencing and whole exome capture. *Autoimmun. Rev.* **14**, 204–209 (2015).
333. MacArthur, D. G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469–476 (2014).
334. Yang, Y. *et al.* Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. *N. Engl. J. Med.* **369**, 1502–1511 (2013).
335. Picard, C. *et al.* Primary Immunodeficiency Diseases: an Update on the Classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015. *J. Clin. Immunol.* **35**, 696–726 (2015).
336. van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet. TIG* **30**, 418–426 (2014).
337. MAP3K14 mitogen-activated protein kinase kinase kinase 14 [Homo sapiens (human)] - Gene - NCBI. Available at: <http://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=ShowDetailView&TermToSearch=9020>. (Accessed: 13th March 2016)
338. Rowe, A. M. *et al.* A cell-intrinsic requirement for NF- $\kappa$ B-inducing kinase in CD4 and CD8 T cell memory. *J. Immunol. Baltim. Md 1950* **191**, 3663–3672 (2013).

339. Xiao, G., Harhaj, E. W. & Sun, S. C. NF-kappaB-inducing kinase regulates the processing of NF-kappaB2 p100. *Mol. Cell* **7**, 401–409 (2001).
340. Willmann, K. L. *et al.* Biallelic loss-of-function mutation in NIK causes a primary immunodeficiency with multifaceted aberrant lymphoid immunity. *Nat. Commun.* **5**, 5360 (2014).
341. Jost, P. J. & Ruland, J. Aberrant NF- $\kappa$ B signaling in lymphoma: mechanisms, consequences, and therapeutic implications. *Blood* **109**, 2700–2707 (2007).
342. Oancea, G. *et al.* Structural arrangement of the transmission interface in the antigen ABC transport complex TAP. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 5551–5556 (2009).
343. Abele, R. & Tampé, R. The ABCs of immunology: structure and function of TAP, the transporter associated with antigen processing. *Physiol. Bethesda Md* **19**, 216–224 (2004).
344. Herget, M. *et al.* Conformation of peptides bound to the transporter associated with antigen processing (TAP). *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1349–1354 (2011).
345. Abele, R. & Tampé, R. Modulation of the antigen transport machinery TAP by friends and enemies. *FEBS Lett.* **580**, 1156–1163 (2006).
346. Israni, A. V. *et al.* Tubercular meningitis in children: Clinical, pathological, and radiological profile and factors associated with mortality. *J. Neurosci. Rural Pract.* **7**, 400–404 (2016).
347. Al-Muhsen, S. & Casanova, J. L. The genetic heterogeneity of mendelian susceptibility to mycobacterial diseases. *J Allergy Clin Immunol* **122**, 1043–1051 (2008).
348. Beisiegel, M. *et al.* Combination of host susceptibility and Mycobacterium tuberculosis virulence define gene expression profile in the host. *Eur. J. Immunol.* **39**, 3369–3384 (2009).
349. S. Sameer Kumar, G. Gene Expression Profiling of Tuberculous Meningitis. *J. Proteomics Bioinform.* **4**, (2011).

350. Hoeve, M. A. *et al.* IL-12 receptor deficiency revisited: IL-23-mediated signaling is also impaired in human genetic IL-12 receptor beta1 deficiency. *Eur. J. Immunol.* **33**, 3393–3397 (2003).
351. van de Vosse, E., Lichtenauer-Kaligis, E. G., van Dissel, J. T. & Ottenhoff, T. H. Genetic variations in the interleukin-12/interleukin-23 receptor (beta1) chain, and implications for IL-12 and IL-23 receptor structure and function. *Immunogenetics* **54**, 817–29 (2003).
352. Presky, D. H. *et al.* Analysis of the multiple interactions between IL-12 and the high affinity IL-12 receptor complex. *J. Immunol. Baltim. Md 1950* **160**, 2174–2179 (1998).
353. de Jong, R. *et al.* Severe mycobacterial and Salmonella infections in interleukin-12 receptor-deficient patients. *Science* **280**, 1435–1438 (1998).
354. Altare, F. *et al.* Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency. *Science* **280**, 1432–1435 (1998).
355. Beaucoudrey, L. de *et al.* Mutations in STAT3 and IL12RB1 impair the development of human IL-17-producing T cells. *J. Exp. Med.* **205**, 1543–1550 (2008).
356. Wellcome Trust Case Control Consortium *et al.* Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.* **39**, 1329–1337 (2007).
357. Duerr, R. H. *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006).
358. Ma, A. & Malynn, B. A. A20: linking a complex regulator of ubiquitylation to immunity and human disease. *Nat. Rev. Immunol.* **12**, 774–785 (2012).
359. Reveille, J. D. *et al.* Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nat. Genet.* **42**, 123–127 (2010).
360. van de Vosse, E., van Dissel, J. T. & Ottenhoff, T. H. M. Genetic deficiencies of innate immune signalling in human infectious disease. *Lancet Infect. Dis.* **9**, 688–698 (2009).

361. Sakai, T., Matsuoka, M., Aoki, M., Nosaka, K. & Mitsuya, H. Missense mutation of the interleukin-12 receptor beta1 chain-encoding gene is associated with impaired immunity against *Mycobacterium avium* complex infection. *Blood* **97**, 2688–2694 (2001).
362. de Beaucoudrey, L. *et al.* Revisiting human IL-12R $\beta$ 1 deficiency: a survey of 141 patients from 30 countries. *Medicine (Baltimore)* **89**, 381–402 (2010).
363. Aytekin, C. *et al.* BCG Lymphadenitis and Recurrent Oral Candidiasis in an Infant with a New Mutation Leading to Interleukin-12 Receptor Beta-1 Deficiency. *J. Investig. Allergol. Clin. Immunol. Off. Organ Int. Assoc. Asthmology INTERASMA Soc. Latinoam. Alerg. E Immunol.* **21**, 401–404 (2011).
364. Lichtenauer-Kaligis, E. G. R. *et al.* Severe *Mycobacterium bovis* BCG infections in a large series of novel IL-12 receptor beta1 deficient patients and evidence for the existence of partial IL-12 receptor beta1 deficiency. *Eur. J. Immunol.* **33**, 59–69 (2003).
365. Robinson, R. T. *et al.* *Mycobacterium tuberculosis* infection induces *il12rb1* splicing to generate a novel IL-12Rbeta1 isoform that enhances DC migration. *J. Exp. Med.* **207**, 591–605 (2010).
366. Ray, A. A., Fountain, J. J., Miller, H. E., Cooper, A. M. & Robinson, R. T. IL12R $\beta$ 1 $\Delta$ TM Is a Secreted Product of *il12rb1* That Promotes Control of Extrapulmonary Tuberculosis. *Infect. Immun.* **83**, 560–571 (2015).
367. Cooper, A. M., Magram, J., Ferrante, J. & Orme, I. M. Interleukin 12 (IL-12) is crucial to the development of protective immunity in mice intravenously infected with *mycobacterium tuberculosis*. *J. Exp. Med.* **186**, 39–45 (1997).
368. Bogunovic, D. *et al.* Mycobacterial disease and impaired IFN- $\gamma$  immunity in humans with inherited ISG15 deficiency. *Science* **337**, 1684–1688 (2012).
369. Fan, J.-B. & Zhang, D.-E. ISG15 regulates IFN- $\gamma$  immunity in human mycobacterial disease. *Cell Res.* **23**, 173–175 (2013).

370. D’Cunha, J., Knight, E., Haas, A. L., Truitt, R. L. & Borden, E. C. Immunoregulatory properties of ISG15, an interferon-induced cytokine. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 211–215 (1996).
371. Bogunovic, D., Boisson-Dupuis, S. & Casanova, J.-L. ISG15: leading a double life as a secreted molecule. *Exp. Mol. Med.* **45**, e18 (2013).
372. Hare, N. J. *et al.* Microparticles released from *Mycobacterium tuberculosis* -infected human macrophages contain increased levels of the type I interferon inducible proteins including ISG15. *PROTEOMICS* **15**, 3020–3029 (2015).
373. Zhang, D. & Zhang, D.-E. Interferon-stimulated gene 15 and the protein ISGylation system. *J. Interferon Cytokine Res. Off. J. Int. Soc. Interferon Cytokine Res.* **31**, 119–130 (2011).
374. Jeon, Y. J., Yoo, H. M. & Chung, C. H. ISG15 and immune diseases. *Biochim. Biophys. Acta* **1802**, 485–496 (2010).
375. Zhang, X. *et al.* Human intracellular ISG15 prevents interferon- $\alpha/\beta$  over-amplification and auto-inflammation. *Nature* **517**, 89–93 (2015).
376. Pilzer, D., Gasser, O., Moskovich, O., Schifferli, J. A. & Fishelson, Z. Emission of membrane vesicles: roles in complement resistance, immunity and cancer. *Springer Semin. Immunopathol.* **27**, 375–387 (2005).
377. Zhao, C., Denison, C., Huibregtse, J. M., Gygi, S. & Krug, R. M. Human ISG15 conjugation targets both IFN-induced and constitutively expressed proteins functioning in diverse cellular pathways. *Proc. Natl. Acad. Sci.* **102**, 10200–10205 (2005).
378. Giannakopoulos, N. V. *et al.* Proteomic identification of proteins conjugated to ISG15 in mouse and human cells. *Biochem. Biophys. Res. Commun.* **336**, 496–506 (2005).
379. Malakhov, M. P. *et al.* High-throughput Immunoblotting. UBIQUITIN-LIKE PROTEIN ISG15 MODIFIES KEY REGULATORS OF SIGNAL TRANSDUCTION. *J. Biol. Chem.* **278**, 16608–16613 (2003).

380. Lenschow, D. J. *et al.* IFN-stimulated gene 15 functions as a critical antiviral molecule against influenza, herpes, and Sindbis viruses. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1371–1376 (2007).
381. Yuan, W. Influenza B virus NS1 protein inhibits conjugation of the interferon (IFN)-induced ubiquitin-like ISG15 protein. *EMBO J.* **20**, 362–371 (2001).
382. Dauphinee, S. M. *et al.* Contribution of increased ISG15, ISGylation and deregulated type I IFN signaling in Usp18 mutant mice during the course of bacterial infections. *Genes Immun.* **15**, 282–292 (2014).
383. Shaw, M. H. *et al.* Tyk2 negatively regulates adaptive Th1 immunity by mediating IL-10 signaling and promoting IFN-gamma-dependent IL-10 reactivation. *J. Immunol. Baltim. Md 1950* **176**, 7263–7271 (2006).
384. Redford, P. S., Murray, P. J. & O’Garra, A. The role of IL-10 in immune regulation during M. tuberculosis infection. *Mucosal Immunol.* **4**, 261–270 (2011).
385. Kreins, A. Y. *et al.* Human TYK2 deficiency: Mycobacterial and viral infections without hyper-IgE syndrome. *J. Exp. Med.* **212**, 1641–1662 (2015).
386. Heimall, J., Freeman, A. & Holland, S. M. Pathogenesis of hyper IgE syndrome. *Clin. Rev. Allergy Immunol.* **38**, 32–38 (2010).
387. Minegishi, Y. *et al.* Human Tyrosine Kinase 2 Deficiency Reveals Its Requisite Roles in Multiple Cytokine Signals Involved in Innate and Acquired Immunity. *Immunity* **25**, 745–755 (2006).
388. Qin, X.-B., Zhang, W.-J., Zou, L., Huang, P.-J. & Sun, B.-J. Identification potential biomarkers in pulmonary tuberculosis and latent infection based on bioinformatics analysis. *BMC Infect. Dis.* **16**, 500 (2016).
389. Vanoni, M. A., Vitali, T. & Zucchini, D. MICAL, the flavoenzyme participating in cytoskeleton dynamics. *Int. J. Mol. Sci.* **14**, 6920–6959 (2013).
390. Fenteany, G. & Glogauer, M. Cytoskeletal remodeling in leukocyte function. *Curr. Opin. Hematol.* **11**, 15–24 (2004).

391. Leemans, J. C. *et al.* CD44 is a macrophage binding site for Mycobacterium tuberculosis that mediates macrophage recruitment and protective immunity against tuberculosis. *J. Clin. Invest.* **111**, 681–689 (2003).
392. Casanova, J.-L., Holland, S. M. & Notarangelo, L. D. Inborn errors of human JAKs and STATs. *Immunity* **36**, 515–528 (2012).
393. Kilic, S. S. *et al.* A patient with tyrosine kinase 2 deficiency without hyper-IgE syndrome. *J. Pediatr.* **160**, 1055–1057 (2012).
394. Glocker, E.-O. *et al.* Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N. Engl. J. Med.* **361**, 2033–2045 (2009).
395. Glocker, E.-O. *et al.* Infant colitis--it's in the genes. *Lancet Lond. Engl.* **376**, 1272 (2010).
396. Glocker, E.-O., Kotlarz, D., Klein, C., Shah, N. & Grimbacher, B. IL-10 and IL-10 receptor defects in humans. *Ann. N. Y. Acad. Sci.* **1246**, 102–107 (2011).
397. Chandesris, M.-O. *et al.* Autosomal Dominant STAT3 Deficiency and Hyper-IgE Syndrome: Molecular, Cellular, and Clinical Features From a French National Survey. *Medicine (Baltimore)* **91**, e1–e19 (2012).
398. Sowerwine, K. J., Holland, S. M. & Freeman, A. F. Hyper-IgE syndrome update. *Ann. N. Y. Acad. Sci.* **1250**, 25–32 (2012).
399. Gonzaga-Jauregui, C., Lupski, J. R. & Gibbs, R. A. Human Genome Sequencing in Health and Disease. *Annu. Rev. Med.* **63**, 35–61 (2012).
400. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
401. Mathe, E. *et al.* Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res.* **34**, 1317–1325 (2006).
402. Leong, I. U., Stuckey, A., Lai, D., Skinner, J. R. & Love, D. R. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two

- metaservers to classify long QT syndrome gene mutations. *BMC Med. Genet.* **16**, 34 (2015).
403. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* gkr407 (2011). doi:10.1093/nar/gkr407
404. Mi, H., Poudel, S., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* **44**, D336-342 (2016).
405. Tian, J. *et al.* Predicting the phenotypic effects of non-synonymous single nucleotide polymorphisms based on support vector machines. *BMC Bioinformatics* **8**, 450 (2007).
406. Ferrer-Costa, C., Orozco, M. & de la Cruz, X. Sequence-based prediction of pathological mutations. *Proteins* **57**, 811–819 (2004).
407. Li, B. *et al.* Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinforma. Oxf. Engl.* **25**, 2744–2750 (2009).
408. Acharya, V. & Nagarajaram, H. A. Hansa: an automated method for discriminating disease and neutral human nsSNPs. *Hum. Mutat.* **33**, 332–337 (2012).
409. Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L. & Casadio, R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat.* **30**, 1237–1244 (2009).
410. González-Pérez, A. & López-Bigas, N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.* **88**, 440–449 (2011).
411. Stone, E. A. & Sidow, A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* **15**, 978–986 (2005).
412. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).

413. Shintre, C. A. *et al.* Structures of ABCB10, a human ATP-binding cassette transporter in apo- and nucleotide-bound states. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9710–9715 (2013).
414. Oswald, C., Holland, I. B. & Schmitt, L. The motor domains of ABC-transporters. What can structures tell us? *Naunyn. Schmiedebergs Arch. Pharmacol.* **372**, 385–399 (2006).
415. Rees, D. C., Johnson, E. & Lewinson, O. ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* **10**, 218–227 (2009).
416. Ye, J., Osborne, A. R., Groll, M. & Rapoport, T. A. RecA-like motor ATPases--lessons from structures. *Biochim. Biophys. Acta* **1659**, 1–18 (2004).
417. Davidson, A. L., Dassa, E., Orelle, C. & Chen, J. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev. MMBR* **72**, 317–364, table of contents (2008).
418. Dassa, E. Natural history of ABC systems: not only transporters. *Essays Biochem.* **50**, 19–42 (2011).
419. Procko, E., O'Mara, M. L., Bennett, W. F. D., Tieleman, D. P. & Gaudet, R. The mechanism of ABC transporters: general lessons from structural and functional studies of an antigenic peptide transporter. *FASEB J.* **23**, 1287–1302 (2009).
420. Procko, E., Raghuraman, G., Wiley, D. C., Raghavan, M. & Gaudet, R. Identification of domain boundaries within the N-termini of TAP1 and TAP2 and their importance in tapasin binding and tapasin-mediated increase in peptide loading of MHC class I. *Immunol. Cell Biol.* **83**, 475–482 (2005).
421. Kelly, A., Powis, S. H. & al, et. Assembly and Function of the Two ABC Transporter Proteins Encoded in the Human Major Histocompatibility Complex. *Nature* **355**, 641–4 (1992).
422. Kueppers, P., Gupta, R. P., Stindt, J., Smits, S. H. J. & Schmitt, L. Functional Impact of a Single Mutation within the Transmembrane Domain of the Multidrug ABC Transporter Pdr5. *Biochemistry (Mosc.)* **52**, 2184–2195 (2013).

423. Kallabi, F. *et al.* Molecular characterization of X-linked adrenoleukodystrophy in a Tunisian family: identification of a novel missense mutation in the ABCD1 gene. *Neurodegener. Dis.* **12**, 207–211 (2013).
424. Liu, S. *et al.* A Novel Missense Mutation (I26M) in DUOXA2 Causing Congenital Goiter Hypothyroidism Impairs NADPH Oxidase Activity but Not Protein Expression. *J. Clin. Endocrinol. Metab.* **100**, 1225–1229 (2015).
425. Bo, S. *et al.* Isoleucine-to-methionine substitution at residue 148 variant of PNPLA3 gene and metabolic outcomes in gestational diabetes. *Am. J. Clin. Nutr.* **101**, 310–318 (2015).
426. Romeo, S. *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* **40**, 1461–1465 (2008).
427. Musso, G., Gambino, R., Cassader, M. & Pagano, G. Meta-analysis: natural history of non-alcoholic fatty liver disease (NAFLD) and diagnostic accuracy of non-invasive tests for liver disease severity. *Ann. Med.* **43**, 617–649 (2011).
428. Sookoian, S. & Pirola, C. J. Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease. *Hepatol. Baltim. Md* **53**, 1883–1894 (2011).
429. Takahata, S., Senju, N., Osaki, Y., Yoshida, T. & Ida, T. Amino Acid Substitutions in Mosaic Penicillin-Binding Protein 2 Associated with Reduced Susceptibility to Cefixime in Clinical Isolates of *Neisseria gonorrhoeae*. *Antimicrob. Agents Chemother.* **50**, 3638–3645 (2006).
430. Tomazin, R. *et al.* Stable binding of the herpes simplex virus ICP47 protein to the peptide binding site of TAP. *EMBO J.* **15**, 3256–3266 (1996).
431. Ahn, K. *et al.* Molecular mechanism and species specificity of TAP inhibition by herpes simplex virus ICP47. *EMBO J.* **15**, 3247–3255 (1996).
432. Kyritsis, C. *et al.* Molecular mechanism and structural aspects of transporter associated with antigen processing inhibition by the cytomegalovirus protein US6. *J. Biol. Chem.* **276**, 48031–48039 (2001).

433. Hewitt, E. W., Gupta, S. S. & Lehner, P. J. The human cytomegalovirus gene product US6 inhibits ATP binding by TAP. *EMBO J.* **20**, 387–396 (2001).
434. Lapinski, P. E., Neubig, R. R. & Raghavan, M. Walker A lysine mutations of TAP1 and TAP2 interfere with peptide translocation but not peptide binding. *J. Biol. Chem.* **276**, 7526–7533 (2001).
435. Alberts, P., Daumke, O., Deverson, E. V., Howard, J. C. & Knittler, M. R. Distinct functional properties of the TAP subunits coordinate the nucleotide-dependent transport cycle. *Curr. Biol. CB* **11**, 242–251 (2001).
436. Rajalingam, R., Singal, D. P. & Mehra, N. K. Transporter associated with antigen-processing (TAP) genes and susceptibility to tuberculoid leprosy and pulmonary tuberculosis. *Tissue Antigens* **49**, 168–172 (1997).
437. Gomez, L. M. *et al.* Analysis of IL1B, TAP1, TAP2 and IKBL polymorphisms on susceptibility to tuberculosis. *Tissue Antigens* **67**, 290–296 (2006).
438. Xu, C. *et al.* Genetic polymorphisms of LMP/TAP gene and hepatitis B virus infection risk in the Chinese population. *J. Clin. Immunol.* **27**, 534–541 (2007).
439. Naderi, M., Hashemi, M. & Amininia, S. Association of TAP1 and TAP2 Gene Polymorphisms with Susceptibility to Pulmonary Tuberculosis. *Iran. J. Allergy Asthma Immunol.* **15**, 62–68 (2016).
440. Wang, D. *et al.* Association of LMP/TAP gene polymorphisms with tuberculosis susceptibility in Li population in China. *PLoS One* **7**, e33051 (2012).
441. Xiao, G. & Sun, S.-C. Negative Regulation of the Nuclear Factor  $\kappa$ B-inducing Kinase by a cis-Acting Domain. *J. Biol. Chem.* **275**, 21081–21085 (2000).
442. Liu, J. *et al.* Structure of the Nuclear Factor  $\kappa$ B-inducing Kinase (NIK) Kinase Domain Reveals a Constitutively Active Conformation. *J. Biol. Chem.* **287**, 27326–27334 (2012).
443. Malinin, N. L., Boldin, M. P., Kovalenko, A. V. & Wallach, D. MAP3K-related kinase involved in NF- $\kappa$ B induction by TNF, CD95 and IL-1. *Nature* **385**, 540–544 (1997).

444. Lin, X. *et al.* Molecular determinants of NF-kappaB-inducing kinase action. *Mol. Cell. Biol.* **18**, 5899–5907 (1998).
445. Sambataro, F. *et al.* Catechol-O-Methyltransferase Valine158Methionine Polymorphism Modulates Brain Networks Underlying Working Memory Across Adulthood. *Biol. Psychiatry* **66**, 540–548 (2009).
446. Strack, V., Bossenmaier, B., Stoyanov, B., Mushack, J. & Häring, H. U. A 973 valine to methionine mutation of the human insulin receptor: interaction with insulin-receptor substrate-1 and Shc in HEK 293 cells. *Diabetologia* **40**, 1135–1140 (1997).
447. Kazemi-Esfarjani, P. *et al.* Substitution of valine-865 by methionine or leucine in the human androgen receptor causes complete or partial androgen insensitivity, respectively with distinct androgen receptor phenotypes. *Mol. Endocrinol. Baltim. Md* **7**, 37–46 (1993).
448. Chen, Z.-Y., Bath, K., McEwen, B., Hempstead, B. & Lee, F. Impact of genetic variant BDNF (Val66Met) on brain structure and function. *Novartis Found. Symp.* **289**, 180–195 (2008).
449. Soong, B.-W. *et al.* A homozygous NOTCH3 mutation p.R544C and a heterozygous TREX1 variant p.C99MfsX3 in a family with hereditary small vessel disease of the brain. *J. Chin. Med. Assoc. JCMA* **76**, 319–324 (2013).
450. Gohar, F. *et al.* Secretory Activity of Neutrophils Correlates With Genotype in Familial Mediterranean Fever. *Arthritis Rheumatol.* n/a-n/a (2016). doi:10.1002/art.39784
451. Hartel, B. P. *et al.* A combination of two truncating mutations in USH2A causes more severe and progressive hearing impairment in Usher syndrome type IIa. *Hear. Res.* **339**, 60–68 (2016).
452. Meka, P. bhushann *et al.* HIF-1 $\alpha$  (1772C>T) polymorphism as marker for breast cancer development. *Tumor Biol.* **36**, 3215–3220 (2014).
453. Ahmed, I. *et al.* TGIF1 is a Potential Candidate Gene for High Myopia in Ethnic Kashmiri Population. *Curr. Eye Res.* **39**, 282–290 (2014).

454. Han, F.-F., Guo, C.-L., Gong, L.-L., Jin, Z. & Liu, L.-H. Effects of the NQO1 609C>T polymorphism on leukemia susceptibility: evidence from a meta-analysis. *Asian Pac. J. Cancer Prev. APJCP* **14**, 5311–5316 (2013).
455. Chang, Y.-G. *et al.* Different roles for two ubiquitin-like domains of ISG15 in protein modification. *J. Biol. Chem.* **283**, 13370–13377 (2008).
456. Sridharan, H., Zhao, C. & Krug, R. M. Species specificity of the NS1 protein of influenza B virus: NS1 binds only human and non-human primate ubiquitin-like ISG15 proteins. *J. Biol. Chem.* **285**, 7852–7856 (2010).
457. Li, L. *et al.* Crystal Structure of Human ISG15 Protein in Complex with Influenza B Virus NS1B. *J. Biol. Chem.* **286**, 30258–30262 (2011).
458. Ham, S. J. *et al.* Interaction between RING1 (R1) and the Ubiquitin-like (UBL) Domains Is Critical for the Regulation of Parkin Activity. *J. Biol. Chem.* **291**, 1803–1816 (2016).
459. Im, E., Yoo, L., Hyun, M., Shin, W. H. & Chung, K. C. Covalent ISG15 conjugation positively regulates the ubiquitin E3 ligase activity of parkin. *Open Biol.* **6**, 160193 (2016).
460. Molnár, J., Szakács, G. & Tusnády, G. E. Characterization of Disease-Associated Mutations in Human Transmembrane Proteins. *PLoS ONE* **11**, (2016).
461. Raja, H., Salomão, D. R., Viswanatha, D. S. & Pulido, J. S. PREVALENCE OF MYD88 L265P MUTATION IN HISTOLOGICALLY PROVEN, DIFFUSE LARGE B-CELL VITREORETINAL LYMPHOMA: *Retina* **36**, 624–628 (2016).
462. Scurry, A. N. *et al.* Structural and Functional Abnormalities of the Neuromuscular Junction in the Trembler-J Homozygote Mouse Model of Congenital Hypomyelinating Neuropathy. *J. Neuropathol. Exp. Neurol.* **75**, 334–346 (2016).
463. McNally, R., Toms, A. V. & Eck, M. J. Crystal Structure of the FERM-SH2 Module of Human Jak2. *PLoS ONE* **11**, (2016).
464. Stark, G. R. & Darnell, J. E. The JAK-STAT pathway at twenty. *Immunity* **36**, 503–514 (2012).

465. Ghoreschi, K., Laurence, A. & O'Shea, J. J. Janus kinases in immune cell signaling. *Immunol. Rev.* **228**, 273–287 (2009).
466. Haan, C., Kreis, S., Margue, C. & Behrmann, I. Jaks and cytokine receptors--an intimate relationship. *Biochem. Pharmacol.* **72**, 1538–1546 (2006).
467. Toms, A. V. *et al.* Structure of a pseudokinase-domain switch that controls oncogenic activation of Jak kinases. *Nat. Struct. Mol. Biol.* **20**, 1221–1223 (2013).
468. Shan, Y. *et al.* Molecular basis for pseudokinase-dependent autoinhibition of JAK2 tyrosine kinase. *Nat. Struct. Mol. Biol.* **21**, 579–584 (2014).
469. Wallweber, H. J. A., Tam, C., Franke, Y., Starovasnik, M. A. & Lupardus, P. J. Structural basis of recognition of interferon- $\alpha$  receptor by tyrosine kinase 2. *Nat. Struct. Mol. Biol.* **21**, 443–448 (2014).
470. Royer, Y., Staerk, J., Costuleanu, M., Courtoy, P. J. & Constantinescu, S. N. Janus kinases affect thrombopoietin receptor cell surface localization and stability. *J. Biol. Chem.* **280**, 27251–27261 (2005).
471. Liu, B. A. *et al.* SRC Homology 2 Domain Binding Sites in Insulin, IGF-1 and FGF receptor mediated signaling networks reveal an extensive potential interactome. *Cell Commun. Signal. CCS* **10**, 27 (2012).
472. McNally, R. & Eck, M. J. JAK-cytokine receptor recognition, unboxed. *Nat. Struct. Mol. Biol.* **21**, 431–433 (2014).
473. Valiev, M., Yang, J., Adams, J. A., Taylor, S. S. & Weare, J. H. Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanics simulations. *J. Phys. Chem. B* **111**, 13455–13464 (2007).
474. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
475. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).
476. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).

477. Canessa, C. M., Horisberger, J. D., Louvard, D. & Rossier, B. C. Mutation of a cysteine in the first transmembrane segment of Na,K-ATPase alpha subunit confers ouabain resistance. *EMBO J.* **11**, 1681–1687 (1992).
478. Lee, W. H. *et al.* The retinoblastoma susceptibility gene encodes a nuclear phosphoprotein associated with DNA binding activity. *Nature* **329**, 642–645 (1987).
479. Chen, P. L., Scully, P., Shew, J. Y., Wang, J. Y. & Lee, W. H. Phosphorylation of the retinoblastoma gene product is modulated during the cell cycle and cellular differentiation. *Cell* **58**, 1193–1198 (1989).
480. Bignon, Y. J. *et al.* A single Cys706 to Phe substitution in the retinoblastoma protein causes the loss of binding to SV40 T antigen. *Cell Growth Differ. Mol. Biol. J. Am. Assoc. Cancer Res.* **1**, 647–651 (1990).
481. Brown, T. R. *et al.* Deletion of the steroid-binding domain of the human androgen receptor gene in one family with complete androgen insensitivity syndrome: evidence for further genetic heterogeneity in this syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8151–8155 (1988).
482. Trifiro, M. *et al.* The 56/58 kDa androgen-binding protein in male genital skin fibroblasts with a deleted androgen receptor gene. *Mol. Cell. Endocrinol.* **75**, 37–47 (1991).
483. Warriar, N., Yu, C., Pagé, N. & Govindan, M. V. Substitution of Cys-560 by Phe, Trp, Tyr, and Ser in the first zinc finger of human androgen receptor affects hormonal sensitivity and transcriptional activation. *J. Biol. Chem.* **269**, 29016–29023 (1994).
484. Zhao, L., Ma, Y., Seemann, J. & Huang, L. J. A regulating role of the JAK2 FERM domain in hyperactivation of JAK2(V617F). *Biochem. J.* **426**, 91–98 (2010).
485. Lupardus, P. J. *et al.* Structure of the pseudokinase-kinase domains from protein kinase TYK2 reveals a mechanism for Janus kinase (JAK) autoinhibition. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8025–8030 (2014).

486. Alvarado-Facundo, E. *et al.* Glycosylation of Residue 141 of Subtype H7 Influenza A Hemagglutinin (HA) Affects HA-Pseudovirus Infectivity and Sensitivity to Site A Neutralizing Antibodies. *PLoS ONE* **11**, (2016).
487. Yamasaki, T. *et al.* Dynamic Changes in Striatal mGluR1 But Not mGluR5 during Pathological Progression of Parkinson's Disease in Human Alpha-Synuclein A53T Transgenic Rats: A Multi-PET Imaging Study. *J. Neurosci.* **36**, 375–384 (2016).
488. Salto, L. M. *et al.* The Ala54Thr Polymorphism of the Fatty Acid Binding Protein 2 Gene Modulates HDL Cholesterol in Mexican-Americans with Type 2 Diabetes. *Int. J. Environ. Res. Public Health* **13**, (2016).
489. Robinson, R. T. IL12R $\beta$ 1: the cytokine receptor that we used to know. *Cytokine* **71**, 348–359 (2015).
490. Ouederni, M. *et al.* Clinical features of Candidiasis in patients with inherited interleukin 12 receptor  $\beta$ 1 deficiency. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **58**, 204–213 (2014).
491. Chua, A. O. *et al.* Expression cloning of a human IL-12 receptor component. A new member of the cytokine receptor superfamily with strong homology to gp130. *J. Immunol.* **153**, 128–136 (1994).
492. Zou, J., Presky, D. H., Wu, C. Y. & Gubler, U. Differential associations between the cytoplasmic regions of the interleukin-12 receptor subunits beta1 and beta2 and JAK kinases. *J. Biol. Chem.* **272**, 6073–6077 (1997).
493. Campbell, I. D. & Spitzfaden, C. Building proteins with fibronectin type III modules. *Struct. Lond. Engl.* **1993** **2**, 333–337 (1994).
494. Erickson, H. P. Stretching fibronectin. *J. Muscle Res. Cell Motil.* **23**, 575–580 (2002).
495. Hedberg, C. *et al.* Hereditary myopathy with early respiratory failure is associated with misfolding of the titin fibronectin III 119 subdomain. *Neuromuscul. Disord.* **24**, 373–379 (2014).

496. Lu, W., Wu, G. & Zhang, B. Association Between X-Ray Cross-Complementing Group 3 (XRCC3) Thr241Met Polymorphism and Risk of Thyroid Cancer: A Meta-Analysis. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **21**, 3978–3985 (2015).
497. Bastos, H. N. *et al.* Association of Polymorphisms in Genes of the Homologous Recombination DNA Repair Pathway and Thyroid Cancer Risk. *Thyroid* **19**, 1067–1075 (2009).
498. Sturgis, E. M., Zhao, C., Zheng, R. & Wei, Q. Radiation Response Genotype and Risk of Differentiated Thyroid Cancer: A Case-Control Analysis. *The Laryngoscope* **115**, 938–945 (2005).
499. Zhang, Y.-Z. *et al.* Compound Library Screening Identified Cardiac Glycoside Digitoxin as an Effective Growth Inhibitor of Gefitinib-Resistant Non-Small Cell Lung Cancer via Downregulation of  $\alpha$ -Tubulin and Inhibition of Microtubule Formation. *Molecules* **21**, 374 (2016).
500. Hypogammaglobulinemia: Practice Essentials, Background, Pathophysiology. (2016).
501. Yin, L. *et al.* Defective lymphotoxin-beta receptor-induced NF-kappaB transcriptional activity in NIK-deficient mice. *Science* **291**, 2162–2165 (2001).
502. Miyawaki, S. *et al.* A new mutation, aly, that induces a generalized lack of lymph nodes accompanied by immunodeficiency in mice. *Eur. J. Immunol.* **24**, 429–434 (1994).
503. Shinkura, R. *et al.* A lymphoplasia is caused by a point mutation in the mouse gene encoding Nf-kappa b-inducing kinase. *Nat. Genet.* **22**, 74–77 (1999).
504. Jin, J. *et al.* The kinase TBK1 controls IgA class switching by negatively regulating noncanonical NF- $\kappa$ B signaling. *Nat. Immunol.* **13**, 1101–1109 (2012).
505. de Leon-Boenig, G. *et al.* The Crystal Structure of the Catalytic Domain of the NF- $\kappa$ B Inducing Kinase Reveals a Narrow but Flexible Active Site. *Structure* **20**, 1704–1714 (2012).

506. Verbeek, D. S., Goedhart, J., Bruinsma, L., Sinke, R. J. & Reits, E. A. PKC $\gamma$  mutations in spinocerebellar ataxia type 14 affect C1 domain accessibility and kinase activity leading to aberrant MAPK signaling. *J. Cell Sci.* **121**, 2339–2349 (2008).
507. Sun, G. & Budde, R. J. Mutations in the N-terminal regulatory region reduce the catalytic activity of Csk, but do not affect its recognition of Src. *Arch. Biochem. Biophys.* **367**, 167–172 (1999).
508. Cheng, H. Y., Schiavone, A. P. & Smithgall, T. E. A Point Mutation in the N-Terminal Coiled-Coil Domain Releases c-Fes Tyrosine Kinase Activity and Survival Signaling in Myeloid Leukemia Cells. *Mol. Cell. Biol.* **21**, 6170–6180 (2001).
509. Smithgall, T. E. *et al.* The c-Fes family of protein-tyrosine kinases. *Crit. Rev. Oncog.* **9**, 43–62 (1998).
510. Nolen, B., Taylor, S. & Ghosh, G. Regulation of protein kinases; controlling activity through activation segment conformation. *Mol. Cell* **15**, 661–675 (2004).
511. Yamamoto, A. *et al.* Bafilomycin A1 prevents maturation of autophagic vacuoles by inhibiting fusion between autophagosomes and lysosomes in rat hepatoma cell line, H-4-II-E cells. *Cell Struct. Funct.* **23**, 33–42 (1998).
512. Sarkar, S. *et al.* Small molecules enhance autophagy and reduce toxicity in Huntington's disease models. *Nat. Chem. Biol.* **3**, 331–338 (2007).
513. Tanida, I., Minematsu-Ikeguchi, N., Ueno, T. & Kominami, E. Lysosomal turnover, but not a cellular level, of endogenous LC3 is a marker for autophagy. *Autophagy* **1**, 84–91 (2005).
514. Levine, B., Mizushima, N. & Virgin, H. W. Autophagy in immunity and inflammation. *Nature* **469**, 323–335 (2011).
515. Deretic, V. & Levine, B. Autophagy, immunity, and microbial adaptations. *Cell Host Microbe* **5**, 527–549 (2009).
516. Mizushima, N. & Komatsu, M. Autophagy: renovation of cells and tissues. *Cell* **147**, 728–741 (2011).

517. Levine, B., Mizushima, N. & Virgin, H. W. Autophagy in immunity and inflammation. *Nature* **469**, 323–335 (2011).
518. Choi, A. M. K., Ryter, S. W. & Levine, B. Autophagy in human health and disease. *N. Engl. J. Med.* **368**, 651–662 (2013).
519. Rubinsztein, D. C., Mariño, G. & Kroemer, G. Autophagy and aging. *Cell* **146**, 682–695 (2011).
520. White, E. Deconvoluting the context-dependent role for autophagy in cancer. *Nat. Rev. Cancer* **12**, 401–410 (2012).
521. Murrow, L. & Debnath, J. Autophagy as a Stress-Response and Quality-Control Mechanism: Implications for Cell Injury and Human Disease. *Annu. Rev. Pathol. Mech. Dis.* **8**, 105–137 (2013).
522. Nixon, R. A. The role of autophagy in neurodegenerative disease. *Nat. Med.* **19**, 983–997 (2013).
523. Gutierrez, M. G. *et al.* Autophagy is a defense mechanism inhibiting BCG and Mycobacterium tuberculosis survival in infected macrophages. *Cell* **119**, 753–766 (2004).
524. Yuk, J. M., Yoshimori, T. & Jo, E. K. Autophagy and bacterial infectious diseases. *Exp. Mol. Med.* **44**, 99–108 (2012).
525. Dale, E., Davis, M. & Faustman, D. L. A role for transcription factor NF-kappaB in autoimmunity: possible interactions of genes, sex, and the immune response. *Adv. Physiol. Educ.* **30**, 152–158 (2006).
526. Rahman, S. *et al.* Compartmentalization of immune responses in human tuberculosis: few CD8<sup>+</sup> effector T cells but elevated levels of FoxP3<sup>+</sup> regulatory t cells in the granulomatous lesions. *Am. J. Pathol.* **174**, 2211–2224 (2009).
527. Aziz, S., Al-Anazi, A. R., Al-Hedaithy, M. A., Al-Shobaili, H. A. & Al-Aska, A. I. Mycobacterium tuberculosis and CD4<sup>+</sup> T-lymphopenia. A grave combination. *Saudi Med. J.* **26**, 1655–1657 (2005).

528. Al-Aska, A. *et al.* CD4+ T-lymphopenia in HIV negative tuberculous patients at King Khalid University Hospital in Riyadh, Saudi Arabia. *Eur. J. Med. Res.* **16**, 285–288 (2011).
529. Uppal, S. S., Tewari, S. C., Verma, S. & Dhot, P. S. Comparison of CD4 and CD8 lymphocyte counts in HIV-negative pulmonary TB patients with those in normal blood donors and the effect of antitubercular treatment: hospital-based flow cytometric study. *Cytometry B Clin. Cytom.* **61**, 20–26 (2004).
530. Davoudi, S. *et al.* CD4+ cell counts in patients with different clinical manifestations of tuberculosis. *Braz. J. Infect. Dis. Off. Publ. Braz. Soc. Infect. Dis.* **12**, 483–486 (2008).
531. Pilheu, J. A. *et al.* CD4+ T-lymphocytopenia in severe pulmonary tuberculosis without evidence of human immunodeficiency virus infection. *Int. J. Tuberc. Lung Dis. Off. J. Int. Union Tuberc. Lung Dis.* **1**, 422–426 (1997).
532. Singhal, M., Banavalikar, J. N., Sharma, S. & Saha, K. Peripheral blood T lymphocyte subpopulations in patients with tuberculosis and the effect of chemotherapy. *Tubercle* **70**, 171–178 (1989).
533. Kony, S. J. *et al.* Tuberculosis-associated severe CD4+ T-lymphocytopenia in HIV-seronegative patients from Dakar. SIDAK Research Group. *J. Infect.* **41**, 167–171 (2000).
534. Lewinsohn, D. A. *et al.* Mycobacterium tuberculosis-specific CD8+ T cells preferentially recognize heavily infected cells. *Am. J. Respir. Crit. Care Med.* **168**, 1346–1352 (2003).
535. Lin, P. L. & Flynn, J. L. CD8 T cells and Mycobacterium tuberculosis infection. *Semin. Immunopathol.* **37**, 239–249 (2015).
536. Prezzemolo, T. *et al.* Functional Signatures of Human CD4 and CD8 T Cell Responses to Mycobacterium tuberculosis. *Front. Immunol.* **5**, (2014).
537. Giorgi, J. V. & Hultin, L. Method for determining favorable prognosis in an HIV positive subject using HLA-DR+ /CD38- CD8bright cells. (1995).
538. Bürgisser, P., Hammann, C., Kaufmann, D., Battegay, M. & Rutschmann, O. T. Expression of CD28 and CD38 by CD8+ T lymphocytes in HIV-1 infection correlates with markers of

- disease severity and changes towards normalization under treatment. *Clin. Exp. Immunol.* **115**, 458–463 (1999).
539. Sleiman, M. *et al.* NK Cell Killer Ig-like Receptor Repertoire Acquisition and Maturation Are Strongly Modulated by HLA Class I Molecules. *J. Immunol.* **192**, 2602–2610 (2014).
540. Addison, E. G. *et al.* Ligation of CD8alpha on human natural killer cells prevents activation-induced apoptosis and enhances cytolytic activity. *Immunology* **116**, 354–361 (2005).
541. Ploegh, H. L., Orr, H. T. & Strominger, J. L. Major histocompatibility antigens: The human (HLA-A,-B,-C) and murine (H-2K, H-2D) class I molecules. *Cell* **24**, 287–299 (1981).
542. Orr, M. T. & Lanier, L. L. Natural Killer Cell Education and Tolerance. *Cell* **142**, 847–856 (2010).
543. Lenzini, L., Rottoli, P. & Rottoli, L. The spectrum of human tuberculosis. *Clin. Exp. Immunol.* **27**, 230–237 (1977).
544. Ashtekar, M. D. *et al.* T lymphocytes in pulmonary tuberculosis. *Indian J. Med. Res.* **97**, 14–17 (1993).
545. Lucas, C. L. *et al.* Heterozygous splice mutation in PIK3R1 causes human immunodeficiency with lymphoproliferation due to dominant activation of PI3K. *J. Exp. Med.* **211**, 2537–2547 (2014).
546. Deau, M.-C. *et al.* A human immunodeficiency caused by mutations in the PIK3R1 gene. *J. Clin. Invest.* **124**, 3923–3928 (2014).
547. Ferreira, M. A. R. *et al.* Quantitative Trait Loci for CD4:CD8 Lymphocyte Ratio Are Associated with Risk of Type 1 Diabetes and HIV-1 Immune Control. *Am. J. Hum. Genet.* **86**, 88–92 (2010).
548. Sunder, S. R., Hanumanth, S. R., Gaddam, S., Jonnalagada, S. & Valluri, V. L. Association of TAP 1 and 2 gene polymorphisms with human immunodeficiency virus-tuberculosis co-infection. *Hum. Immunol.* **72**, 908–911 (2011).

549. Nunes-Alves, C. *et al.* Human and Murine Clonal CD8+ T Cell Expansions Arise during Tuberculosis Because of TCR Selection. *PLoS Pathog.* **11**, e1004849 (2015).
550. Lajoie, J. *et al.* Novel TAP1 polymorphisms in indigenous Zimbabweans: their potential implications on TAP function and in human diseases. *Hum. Immunol.* **64**, 823–829 (2003).
551. Höhler, T. *et al.* TAP-polymorphisms in juvenile onset psoriasis and psoriatic arthritis. *Hum. Immunol.* **51**, 49–54 (1996).
552. Tang, J. *et al.* TAPI polymorphisms in several human ethnic groups: characteristics, evolution, and genotyping strategies. *Hum. Immunol.* **62**, 256–268 (2001).
553. Moins-Teisserenc, H. *et al.* TAP2 gene polymorphism contributes to genetic susceptibility to multiple sclerosis. *Hum. Immunol.* **42**, 195–202 (1995).
554. Takeuchi, F. *et al.* Polymorphism of TAP1 and TAP2 in Japanese patients with rheumatoid arthritis. *Tissue Antigens* **49**, 280–282 (1997).
555. Stein, C. M. Genetic epidemiology of tuberculosis susceptibility: impact of study design. *PLoS Pathog.* **7**, e1001189 (2011).
556. Chang, S. T., Linderman, J. J. & Kirschner, D. E. Effect of multiple genetic polymorphisms on antigen presentation and susceptibility to Mycobacterium tuberculosis infection. *Infect. Immun.* **76**, 3221–3232 (2008).
557. Deng, H. W. Population admixture may appear to mask, change or reverse genetic effects of genes underlying complex traits. *Genetics* **159**, 1319–1323 (2001).
558. Möller, M., de Wit, E. & Hoal, E. G. Past, present and future directions in human genetic susceptibility to tuberculosis. *FEMS ImmunolMedMicrobiol* **58**, 3–26 (2010).
559. Bodmer, W. & Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* **40**, 695–701 (2008).
560. Ma, X. *et al.* Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. *PLoS ONE.* **2**, e1318 (2007).
561. Baker, A. R. *et al.* Genetic Variation in TLR Genes in Ugandan and South African Populations and Comparison with HapMap Data. *PLoS ONE* **7**, (2012).

562. McCarroll, S. A. & Altshuler, D. M. Copy-number variation and association studies of human disease. *Nat. Genet.* **39**, S37-42 (2007).
563. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
564. Jakobsson, M. *et al.* Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **451**, 998–1003 (2008).
565. Velez, D. R. *et al.* NOS2A, TLR4, and IFNGR1 interactions influence pulmonary tuberculosis susceptibility in African-Americans. *Hum.Genet.* **126**, 643–653 (2009).
566. Velez, D. R. *et al.* Association of SLC11A1 with tuberculosis and interactions with NOS2A and TLR2 in African-Americans and Caucasians. *IntJTubercLung Dis* **13**, 1068–1076 (2009).
567. Stein, C. M. & Baker, A. R. Tuberculosis as a complex trait: impact of genetic epidemiological study design. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* **22**, 91–99 (2011).
568. Stein, C. M. *et al.* Linkage and association analysis of candidate genes for TB and TNFalpha cytokine expression: evidence for association with IFNGR1, IL-10, and TNF receptor 1 genes. *Hum.Genet.* **121**, 663–673 (2007).
569. Vesenbeckh, S. M. *et al.* The Use of Interferon Gamma Release Assays in the Diagnosis of Active Tuberculosis. *Tuberc. Res. Treat.* **2012**, e768723 (2012).
570. Diagnostic Standards and Classification of Tuberculosis in Adults and Children. This official statement of the American Thoracic Society and the Centers for Disease Control and Prevention was adopted by the ATS Board of Directors, July 1999. This statement was endorsed by the Council of the Infectious Disease Society of America, September 1999. *Am. J. Respir. Crit. Care Med.* **161**, 1376–1395 (2000).
571. Cobat, A. *et al.* Two loci control tuberculin skin test reactivity in an area hyperendemic for tuberculosis. *J.Exp.Med.* **206**, 2583–2591 (2009).

572. Stein, C. M. *et al.* Genome scan of M. tuberculosis infection and disease in Ugandans. *PLoS.ONE.* **3**, e4094 (2008).
573. Thye, T. *et al.* IL10 haplotype associated with tuberculin skin test response but not with pulmonary TB. *PLoS.ONE.* **4**, e5420 (2009).
574. Motsinger-Reif, A. A. *et al.* Polymorphisms in IL-1beta, vitamin D receptor Fok1, and Toll-like receptor 2 are associated with extrapulmonary tuberculosis. *BMC Med. Genet.* **11**, 37 (2010).
575. Flores-Villanueva, P. O. *et al.* A functional promoter polymorphism in monocyte chemoattractant protein-1 is associated with increased susceptibility to pulmonary tuberculosis. *J.Exp.Med.* **202**, 1649–1658 (2005).
576. Hoal, E. G. *et al.* SLC11A1 (NRAMP1) but not SLC11A2 (NRAMP2) polymorphisms are associated with susceptibility to tuberculosis in a high-incidence community in South Africa. *Int. J. Tuberc. Lung Dis. Off. J. Int. Union Tuberc. Lung Dis.* **8**, 1464–1471 (2004).
577. Taype, C. A. *et al.* Association between SLC11A1 polymorphisms and susceptibility to different clinical forms of tuberculosis in the Peruvian population. *Infect.Genet.Evol.* **6**, 361–367 (2006).
578. Edwards, B. J., Haynes, C., Levenstien, M. A., Finch, S. J. & Gordon, D. Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. *BMC Genet.* **6**, 18 (2005).
579. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9**, 356–369 (2008).
580. Gagneux, S. & Small, P. M. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. *Lancet Infect. Dis.* **7**, 328–337 (2007).
581. Thwaites, G. *et al.* Relationship between Mycobacterium tuberculosis genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. *J. Clin. Microbiol.* **46**, 1363–1368 (2008).

582. Caws, M. *et al.* The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. *PLoS.Pathog.* **4**, e1000034 (2008).
583. Moller, M., Wit, E. de & Hoal, E. G. Past, present and future directions in human genetic susceptibility to tuberculosis. *FEMS Immunol. Med. Microbiol.* **58**, 3–26 (2010).
584. Dewey, F. E. *et al.* Phased whole-genome genetic risk in a family quartet using a major allele reference sequence. *PLoS Genet.* **7**, e1002280 (2011).
585. McMillan, C. T. *et al.* Genetic and neuroanatomic associations in sporadic frontotemporal lobar degeneration. *Neurobiol. Aging* **35**, 1473–1482 (2014).
586. Kitts, A. L., Phan, L., Ward, M. & Holmes, J. B. *The Database of Short Genetic Variation (dbSNP)*. (2014).
587. Day, I. N. M. dbSNP in the detail and copy number complexities. *Hum. Mutat.* **31**, 2–4 (2010).
588. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
589. Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11983–11988 (2011).
590. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
591. The Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46**, 818–825 (2014).
592. Ramsay, M., Sankoh, O. & Consortium, as members of the A.-G. study and the H. African partnerships through the H3Africa Consortium bring a genomic dimension to longitudinal population studies on the continent. *Int. J. Epidemiol.* **45**, 305–308 (2016).
593. Gurdasani, D. *et al.* The African Genome Variation Project shapes medical genetics in Africa. *Nature* **517**, 327–332 (2015).

594. Daya, M., van der Merwe, L., van Helden, P. D., Möller, M. & Hoal, E. G. Investigating the Role of Gene-Gene Interactions in TB Susceptibility. *PLoS ONE* **10**, (2015).
595. Sologuren, I. *et al.* Partial recessive IFN- $\gamma$ R1 deficiency: genetic, immunological and clinical features of 14 patients from 11 kindreds. *Hum. Mol. Genet.* **20**, 1509–1523 (2011).
596. MacLennan, C. *et al.* Interleukin (IL)-12 and IL-23 are key cytokines for immunity against Salmonella in humans. *J. Infect. Dis.* **190**, 1755–1757 (2004).
597. Ziv, E. & Burchard, E. G. Human population structure and genetic association studies. *Pharmacogenomics.* **4**, 431–441 (2003).
598. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459–463 (2010).
599. Sullivan, G. M. & Feinn, R. Using Effect Size—or Why the P Value Is Not Enough. *J. Grad. Med. Educ.* **4**, 279–282 (2012).
600. Button, K. S. *et al.* Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376 (2013).
601. Berry, E. M., Coustère-Yakir, C. & Grover, N. B. The significance of non-significance. *QJM Mon. J. Assoc. Physicians* **91**, 647–653 (1998).
602. Gelman, A., Hill, J. & Yajima, M. Why We (Usually) Don't Have to Worry About Multiple Comparisons. *J. Res. Educ. Eff.* **5**, 189–211 (2012).
603. Mimouni, J. [Our experiences in three years of BCG vaccination at the center of the O.P.H.S. at Constantine; study of observed cases (25 cases of complications from BCG vaccination)]. *Algér. Médicale* **55**, 1138–1147 (1951).
604. Bustamante, J., Boisson-Dupuis, S., Abel, L. & Casanova, J.-L. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN- $\gamma$  immunity. *Semin. Immunol.* **26**, 454–470 (2014).
605. Doffinger, R. *et al.* Autoantibodies to interferon-gamma in a patient with selective susceptibility to mycobacterial infection and organ-specific autoimmunity. *Clin. Infect. Dis.* **38**, E10–E14 (2004).

606. Puel, A. *et al.* Autoantibodies against IL-17A, IL-17F, and IL-22 in patients with chronic mucocutaneous candidiasis and autoimmune polyendocrine syndrome type I. *J. Exp. Med.* **207**, 291–297 (2010).
607. Browne, S. K. *et al.* Adult-onset immunodeficiency in Thailand and Taiwan. *N. Engl. J. Med.* **367**, 725–734 (2012).
608. Chi, C.-Y. *et al.* Anti-IFN- $\gamma$  autoantibodies in adults with disseminated nontuberculous mycobacterial infections are associated with HLA-DRB1\*16:02 and HLA-DQB1\*05:02 and the reactivation of latent varicella-zoster virus infection. *Blood* **121**, 1357–1366 (2013).
609. Casanova, J. L. & Abel, L. Human genetics of infectious diseases: a unified theory. *EMBO J* **26**, 915–922 (2007).
610. Casanova, J.-L. & Abel, L. The genetic theory of infectious diseases: a brief history and selected illustrations. *Annu. Rev. Genomics Hum. Genet.* **14**, 215–243 (2013).
611. Casanova, J.-L. & Abel, L. Genetic dissection of immunity to mycobacteria: the human model. *Annu. Rev. Immunol.* **20**, 581–620 (2002).
612. Holland, S. M. Immunotherapy of mycobacterial infections. *Semin. Respir. Infect.* **16**, 47–59 (2001).
613. Alangari, A. A. *et al.* Treatment of disseminated mycobacterial infection with high-dose IFN- $\gamma$  in a patient with IL-12R $\beta$ 1 deficiency. *Clin. Dev. Immunol.* **2011**, 691956 (2011).
614. Mirsaeidi, M. Personalized medicine approach in mycobacterial disease. *Int. J. Mycobacteriology* **1**, 59–64 (2012).

APPENDIX I – **PID Warning Signs (Adapted from Jeffrey Modell Foundation)**

- Eight or more new ear infections within 1 year
- Two or more serious sinus infections within 1 year
- Two or more months on antibiotics with little effect
- Two or more pneumonias within 1 year
- Failure of an infant to gain weight or grow normally
- Recurrent, deep skin or organ abscesses
- Persistent thrush in mouth or elsewhere on skin, after age 1
- Need for intravenous antibiotics to clear infections
- Two or more deep-seated infections
- Parasitoses (e.g. Pneumocystis jiroveci Pneumonia [PJP] and Giardia)
- Auto-immune manifestations, especially in the very young
- A family history of primary immunodeficiency (or unexplained early death)
- BCG dissemination
- AND Recurrent meningococcal infections

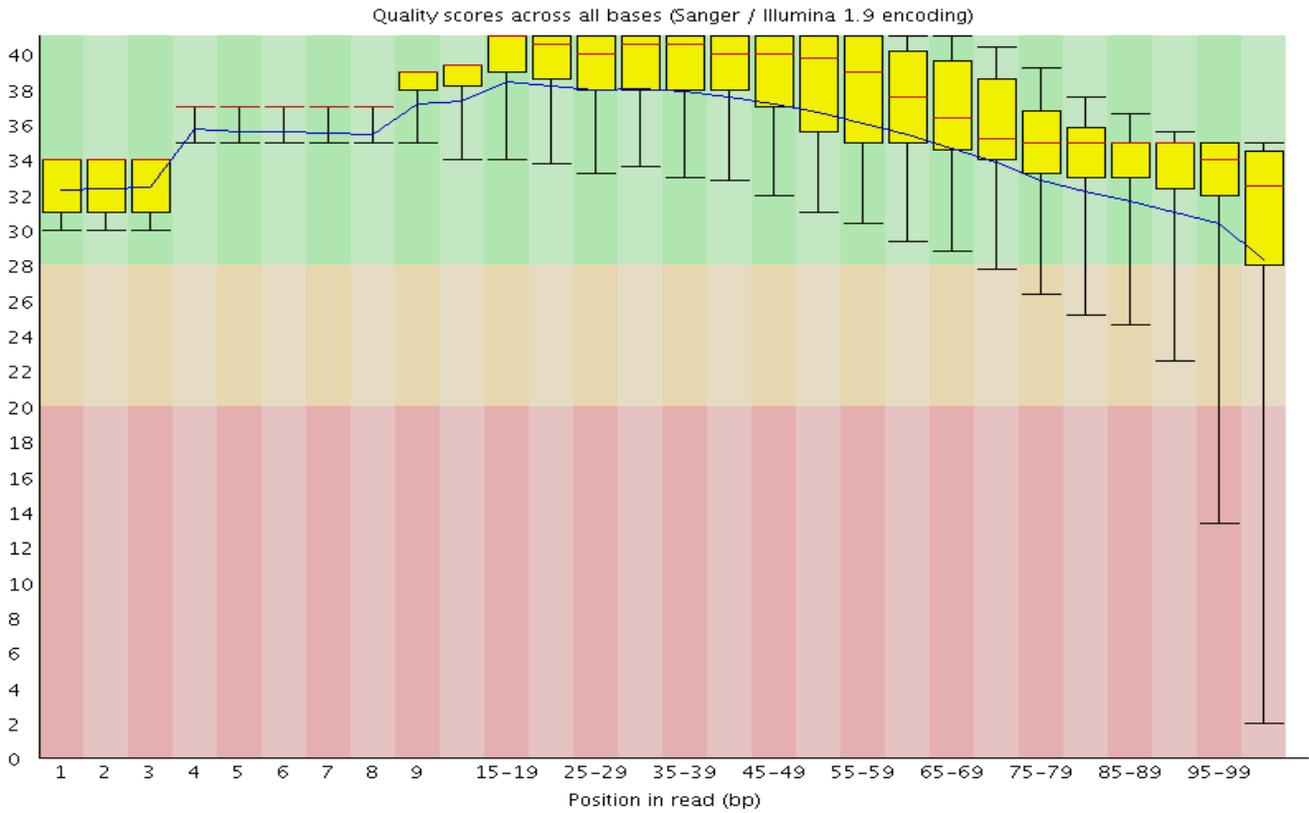
APPENDIX II – **FastQC results for exome sequencing data**

The (1) basic statistics and per base sequence quality of each is shown:

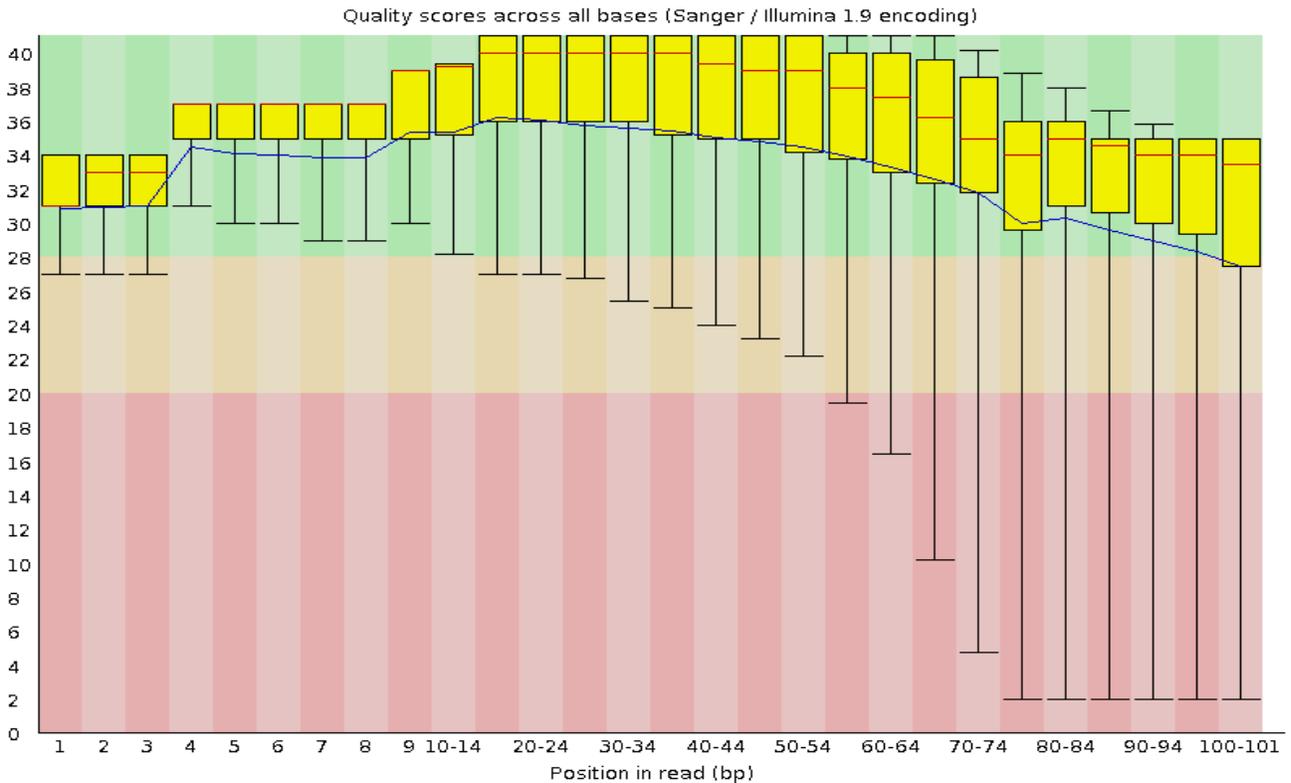
	Total sequences	Sequences flaggad as poor quality	Sequence length	%GC
<b>MSMD PATIENTS AND CONTROLS</b>				
<b>PID_012</b>				
Lane 1	12414535	0	101	49
Lane 2	12338928	0	101	48
Lane 3	12412209	0	101	48
<b>Mother of PID_012</b>				
Lane 2	16609275	0	101	46
Lane 3	16737017	0	101	46
Lane 4	16511972	0	101	46
<b>PID_040</b>				
Lane 2	23860903	0	101	45
Lane 3	24075143	0	101	45
Lane 4	23719586	0	101	45
<b>Father of PID_040</b>				
Lane 2	18842277	0	101	45
Lane 3	18964404	0	101	45
Lane 4	18726093	0	101	45
<b>Mother of PID_040</b>				
Lane 2	15186807	0	101	45
Lane 3	15294029	0	101	45
Lane 4	15081027	0	101	45
<b>PID_060</b>				
Lane 4	15112669	0	101	49
Lane 5	14941272	0	101	49
Lane 6	7505323	0	101	49
Lane 7	7608658	0	101	49
<b>Father of PID_060</b>				
Lane 2	16469583	0	101	45
Lane 3	16573569	0	101	45
Lane 4	16304077	0	101	45
<b>Mother of PID_060</b>				
Lane 2	11230150	0	101	45
Lane 3	11305282	0	101	45
Lane 4	11182370	0	101	45
<b>TBM PATIENTS</b>				
<b>D5370</b>				
Lane 3	23259890	0	126	48
Lane 4	23372529	0	126	48
<b>D5371</b>				
Lane 2	18974781	0	126	50
Lane 3	18858978	0	126	50
<b>D5372</b>				
Lane 3	22253666	0	126	48
Lane 4	22253666	0	126	48
<b>D5373</b>				
Lane 1	24320328	0	126	48
Lane 2	24290063	0	126	48
<b>D5374</b>				
Lane 1	23722813	0	126	48

Lane 2	23747873	0	126	48
<b>D5375</b>				
Lane 3	18758939	0	126	48
Lane 4	18882784	0	126	48
<b>D5376</b>				
Lane 3	19118095	0	126	48
Lane 4	19198537	0	126	48
<b>D5377</b>				
Lane 3	18967122	0	126	48
Lane 4	19083977	0	126	48
<b>D5378</b>				
Lane 1	21579521	0	126	50
Lane 2	21469082	0	126	50
<b>D5379</b>				
Lane 2	17673631	0	126	48
Lane 3	17817547	0	126	48
<b>TBM CONTROLS</b>				
<b>D5380</b>				
Lane 1	26239015	0	126	48
Lane 2	26367095	0	126	48
<b>D5381</b>				
Lane 1	21736959	0	126	48
Lane 2	21745532	0	126	48
<b>D5382</b>				
Lane 3	20927179	0	126	48
Lane 4	21027724	0	126	49
<b>D5383</b>				
Lane 3	22572748	0	126	47
Lane 4	22552798	0	126	47
<b>D5384</b>				
Lane 1	19346091	0	126	48
Lane 2	19428625	0	126	48
<b>D5385</b>				
Lane 3	21793591	0	126	48
Lane 4	21886180	0	126	48
<b>D5386</b>				
Lane 1	21754783	0	126	50
Lane 2	21643615	0	126	50
<b>D5387</b>				
Lane 1	20900575	0	126	50
Lane 2	20768619	0	126	50
<b>D5388</b>				
Lane 1	22970836	0	126	48
Lane 2	23065217	0	126	48
<b>D5389</b>				
Lane 3	20923574	0	126	48
Lane 4	21003713	0	126	48

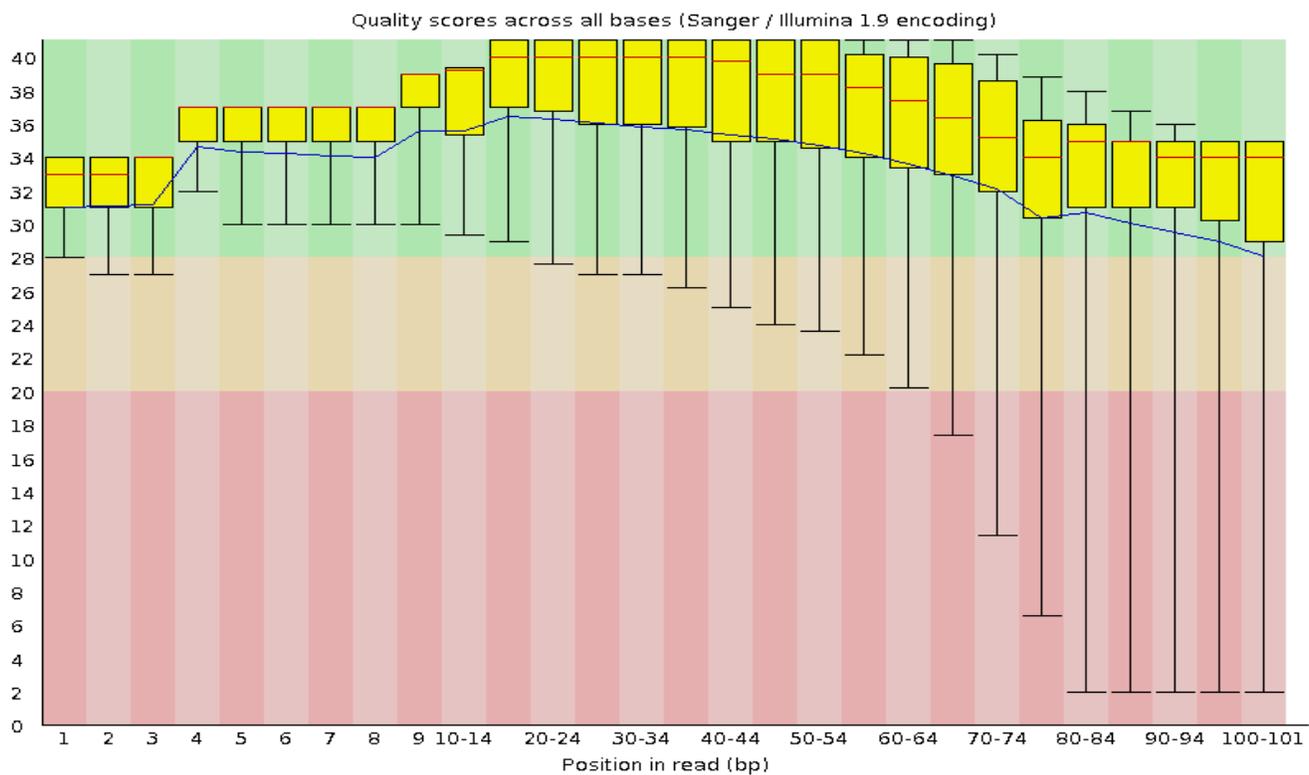
### PID\_012: Per base sequence quality



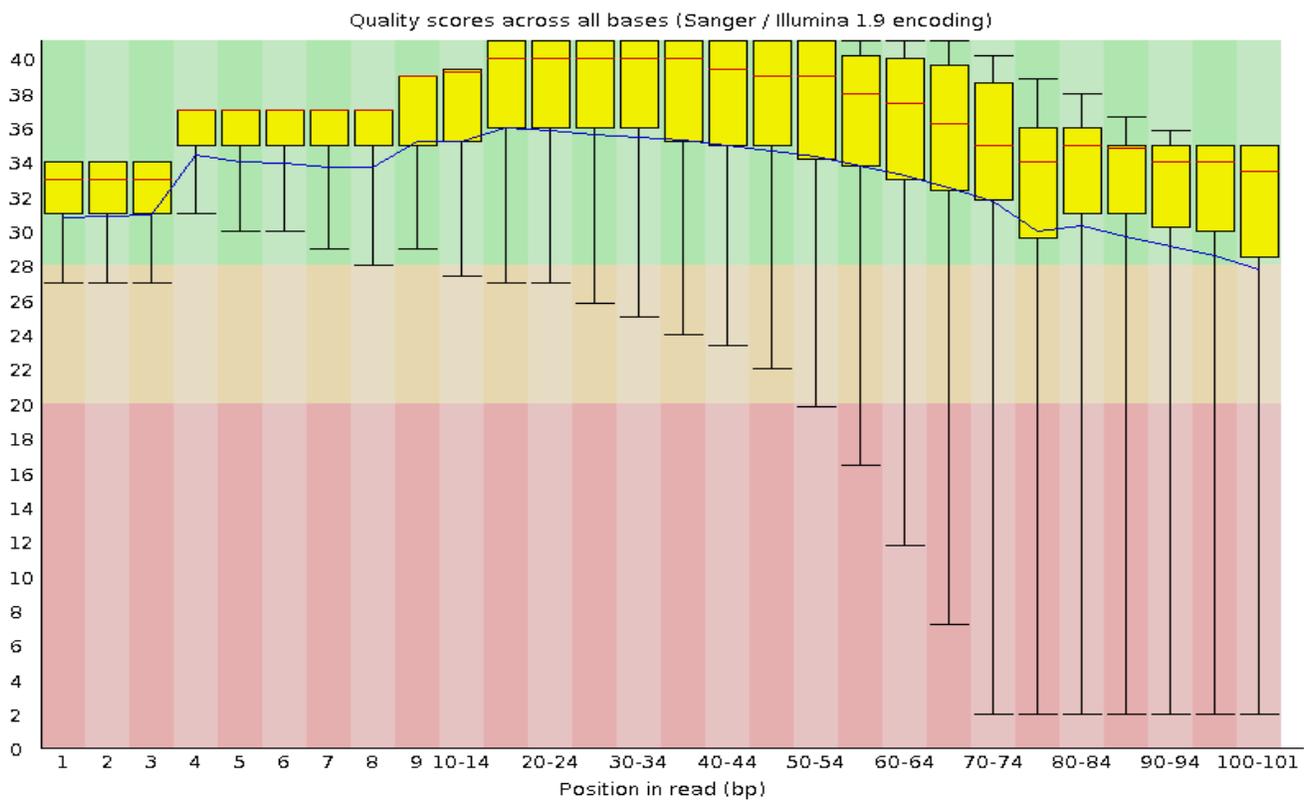
### Mother of PID\_012: Per base sequence quality



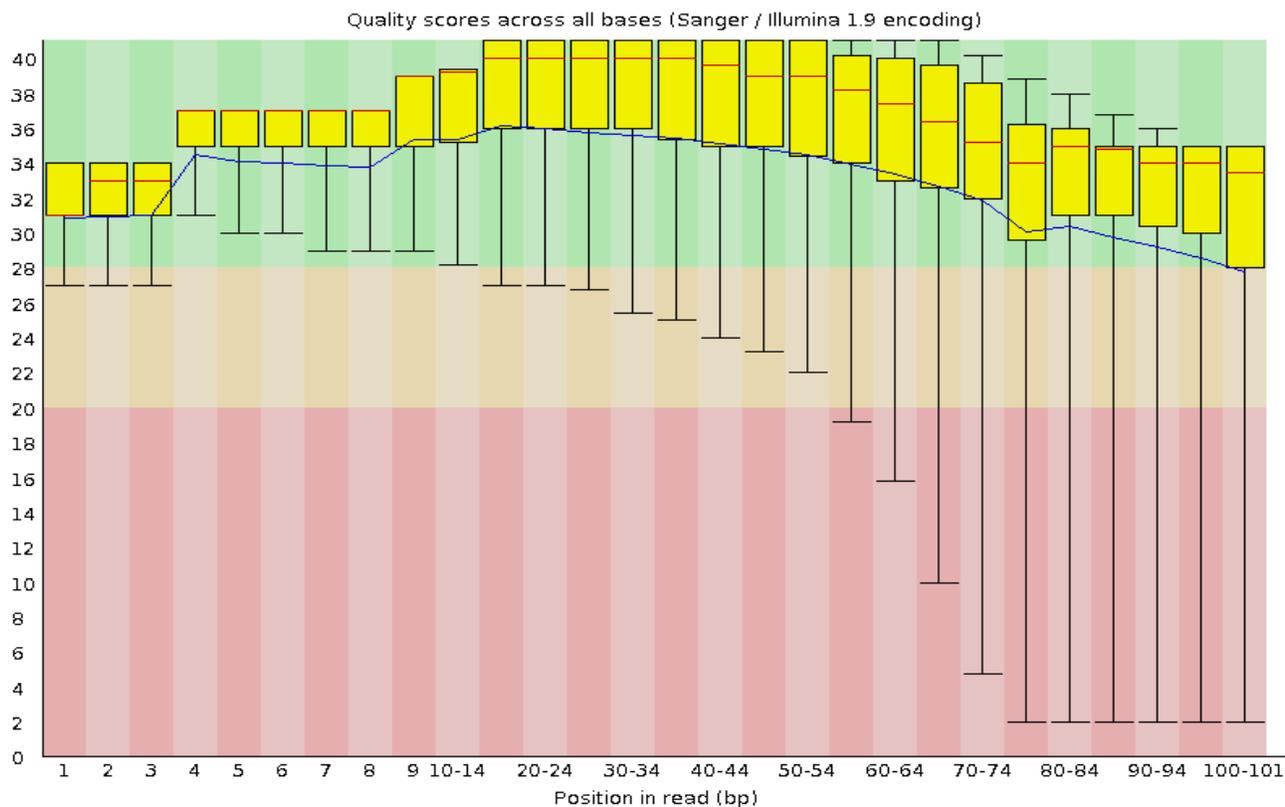
### PID\_040: Per base sequence quality



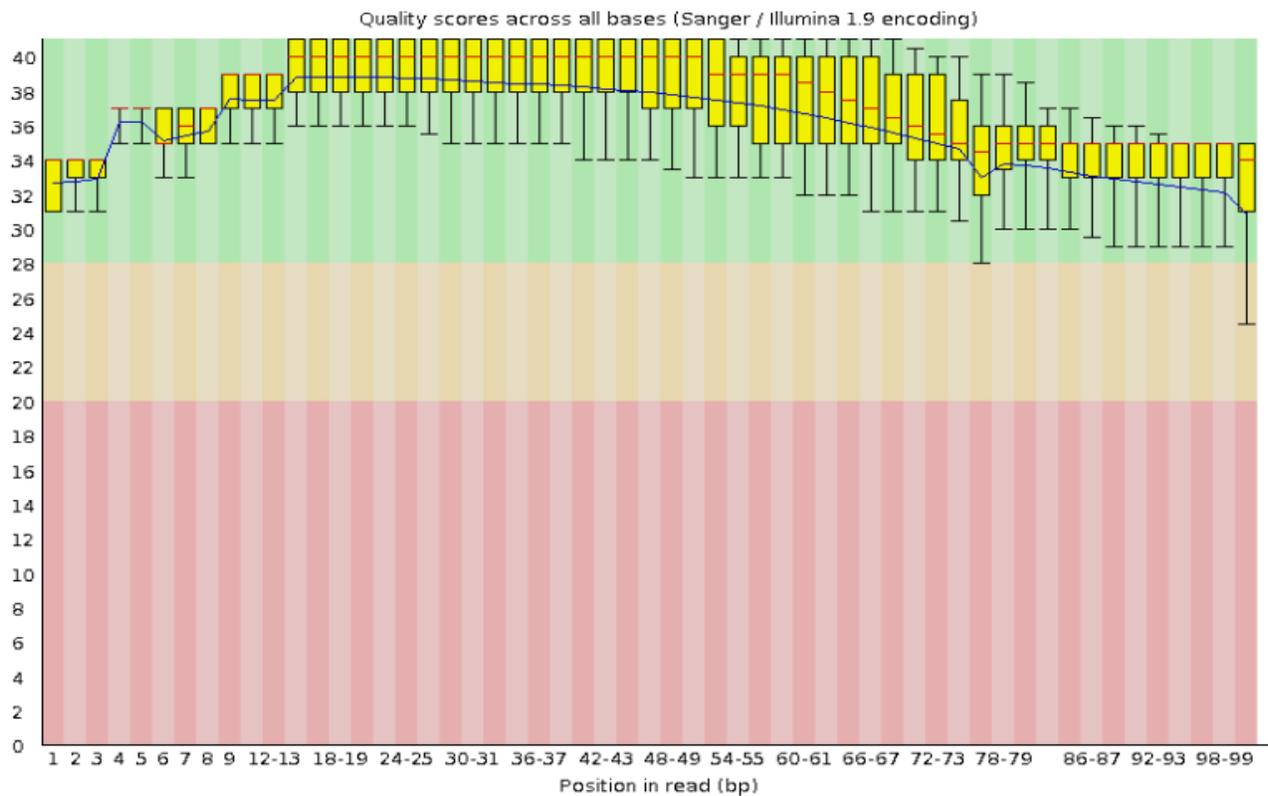
### Father of PID\_040: Per base sequence quality



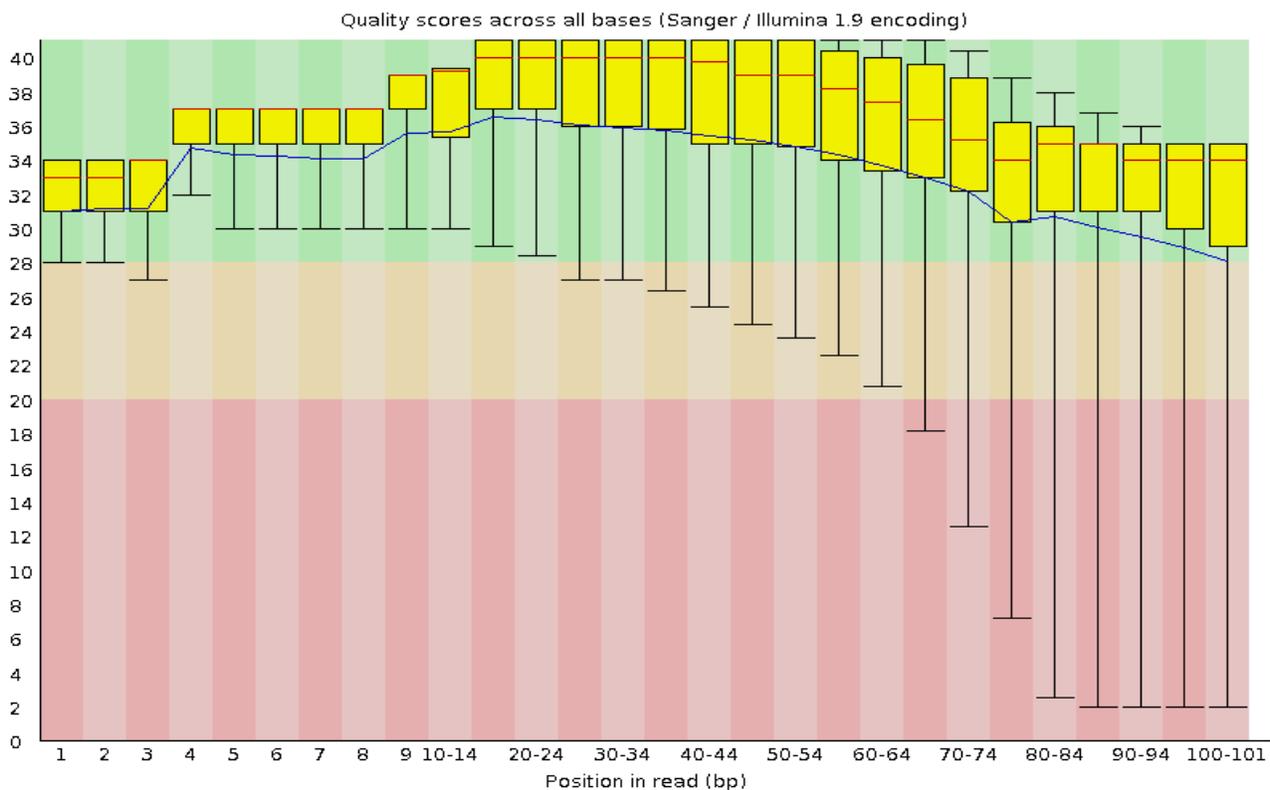
### Mother of PID\_040: Per base sequence quality



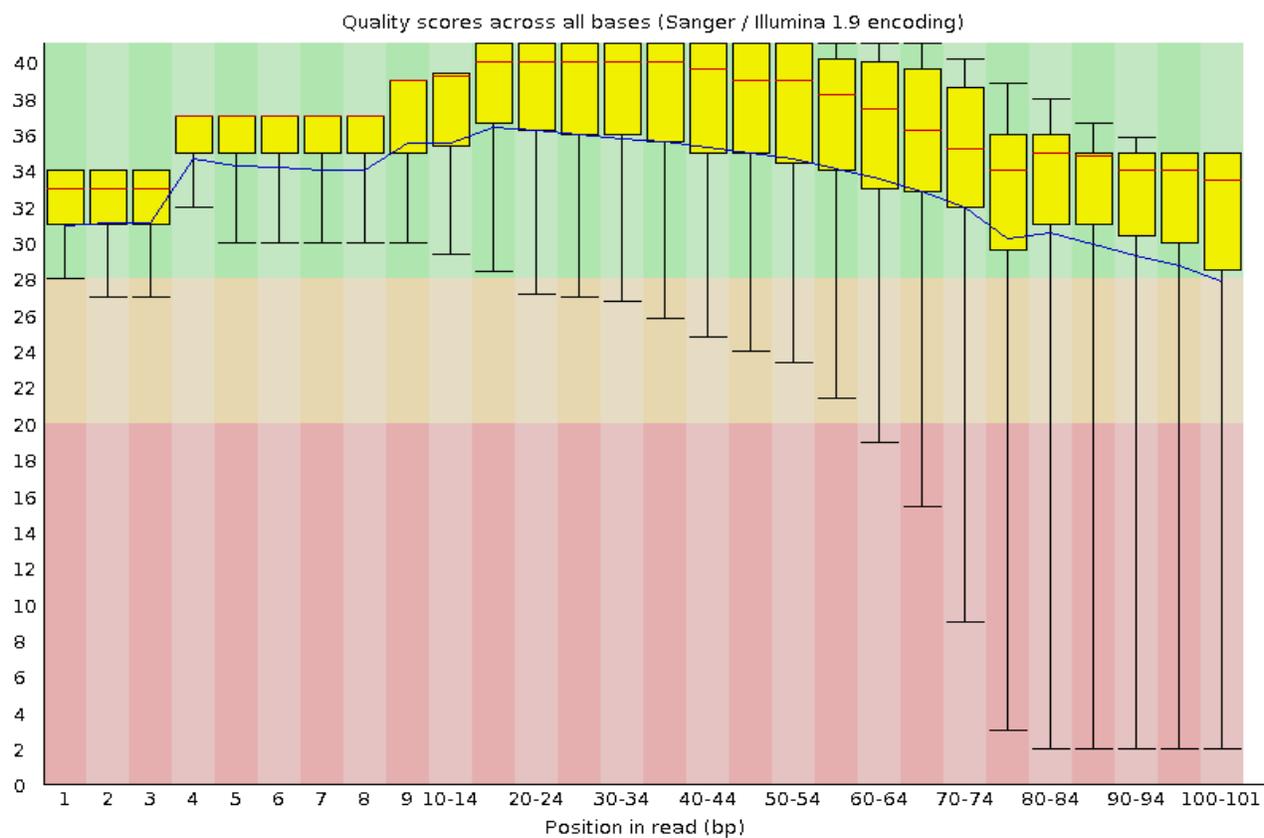
### PID\_060: Per base sequence quality



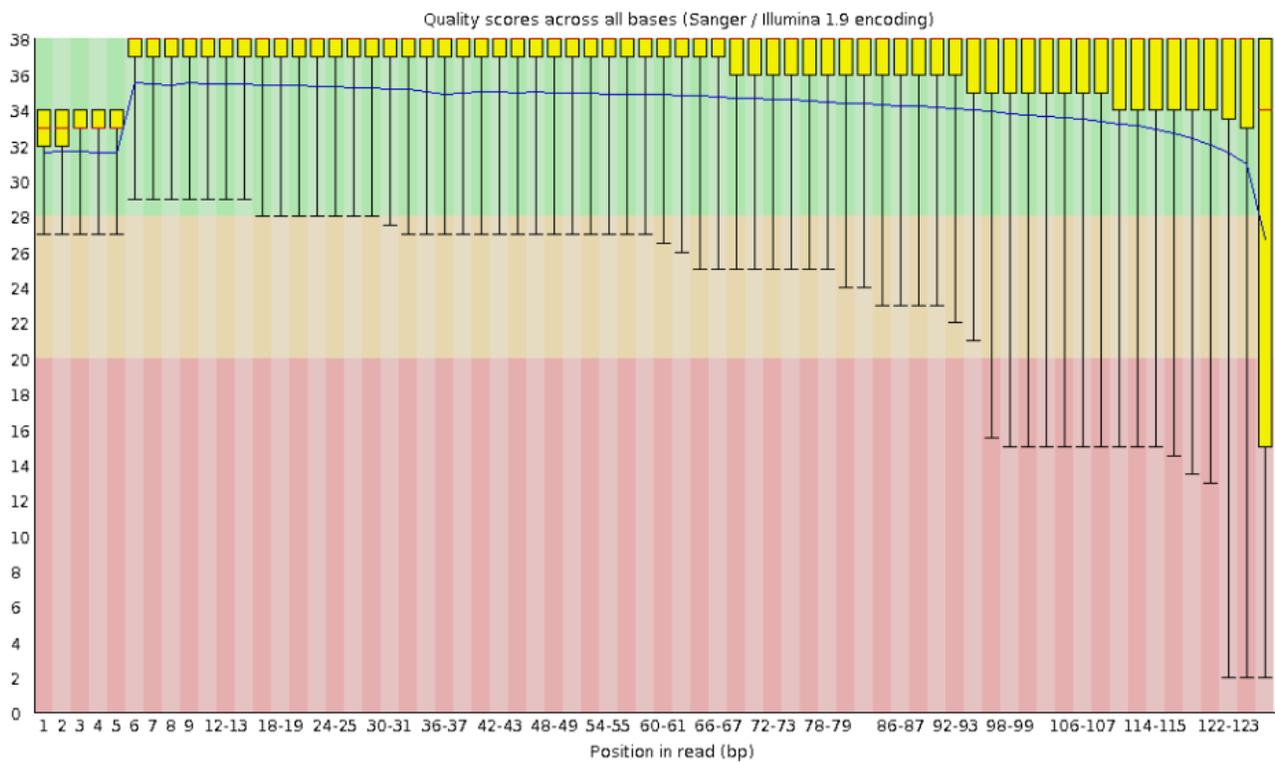
### Father of PID\_060: Per base sequence quality



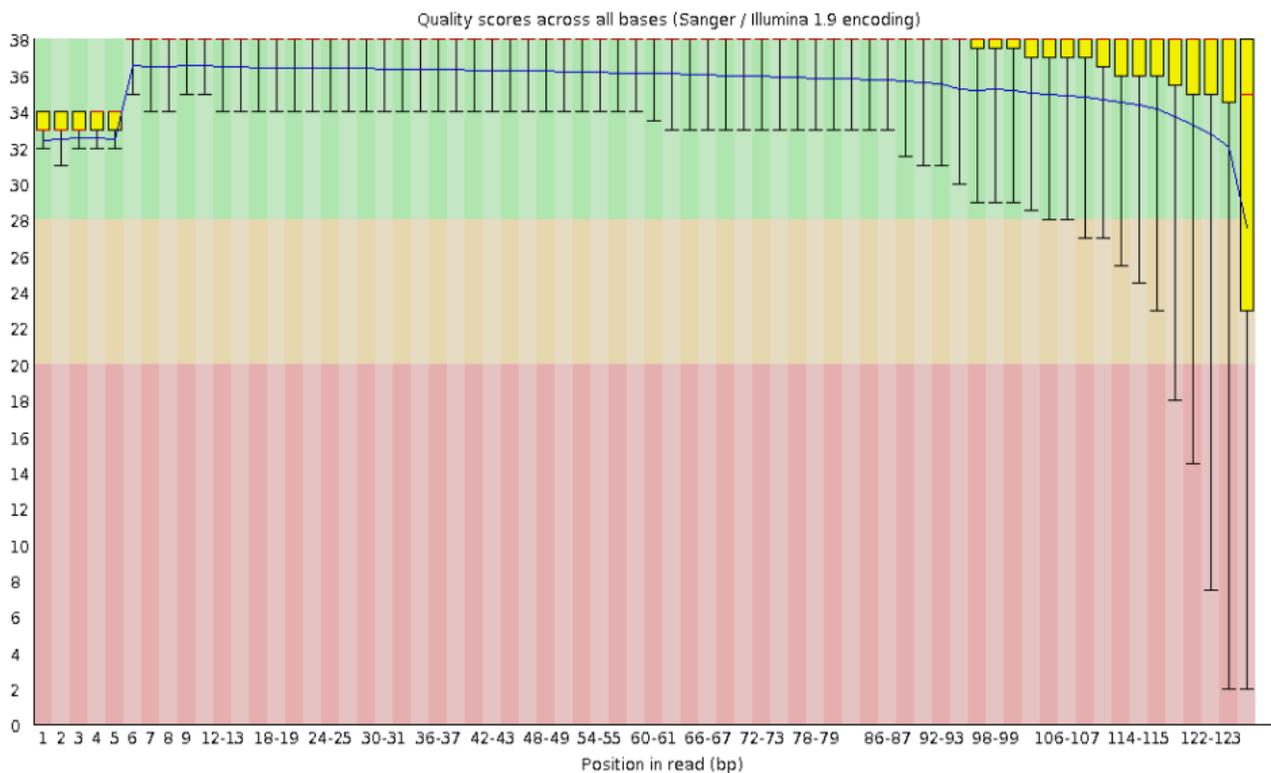
### Mother of PID\_060: Per base sequence quality



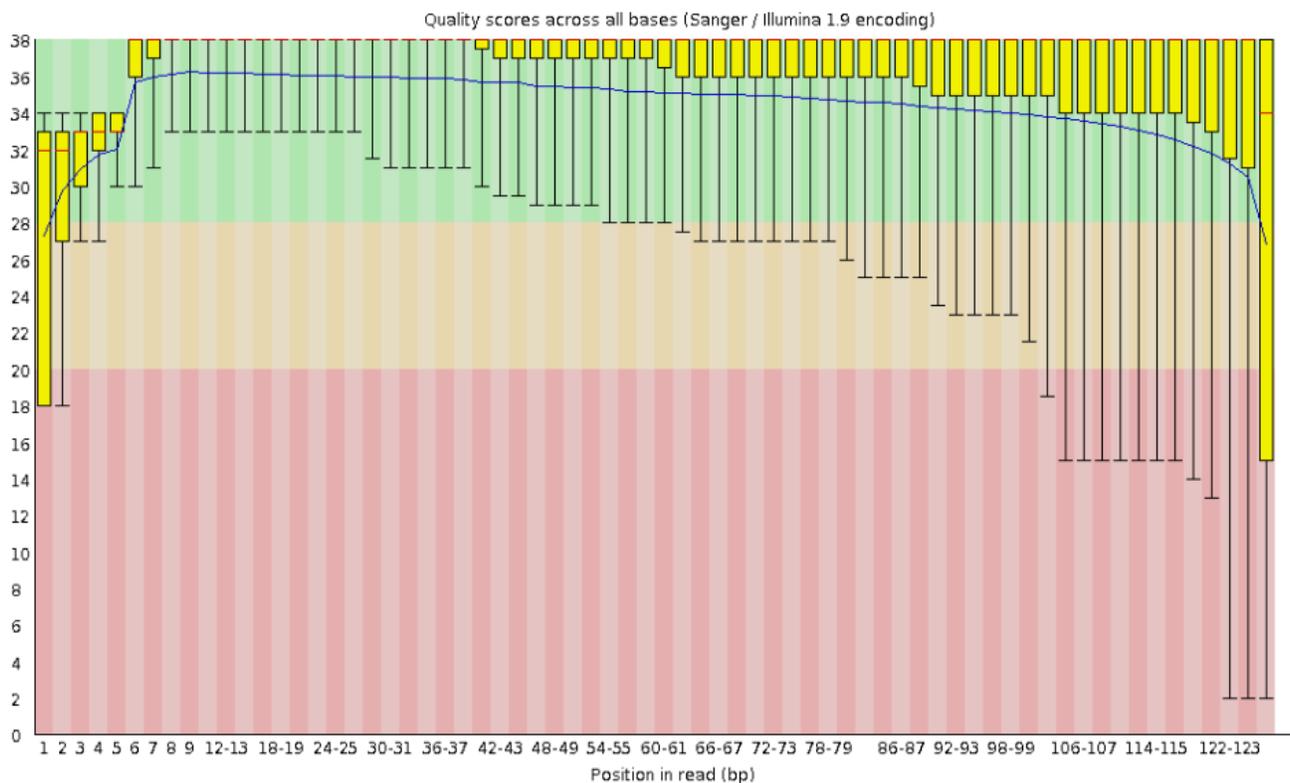
### TBM patient D5370: Per base sequence quality



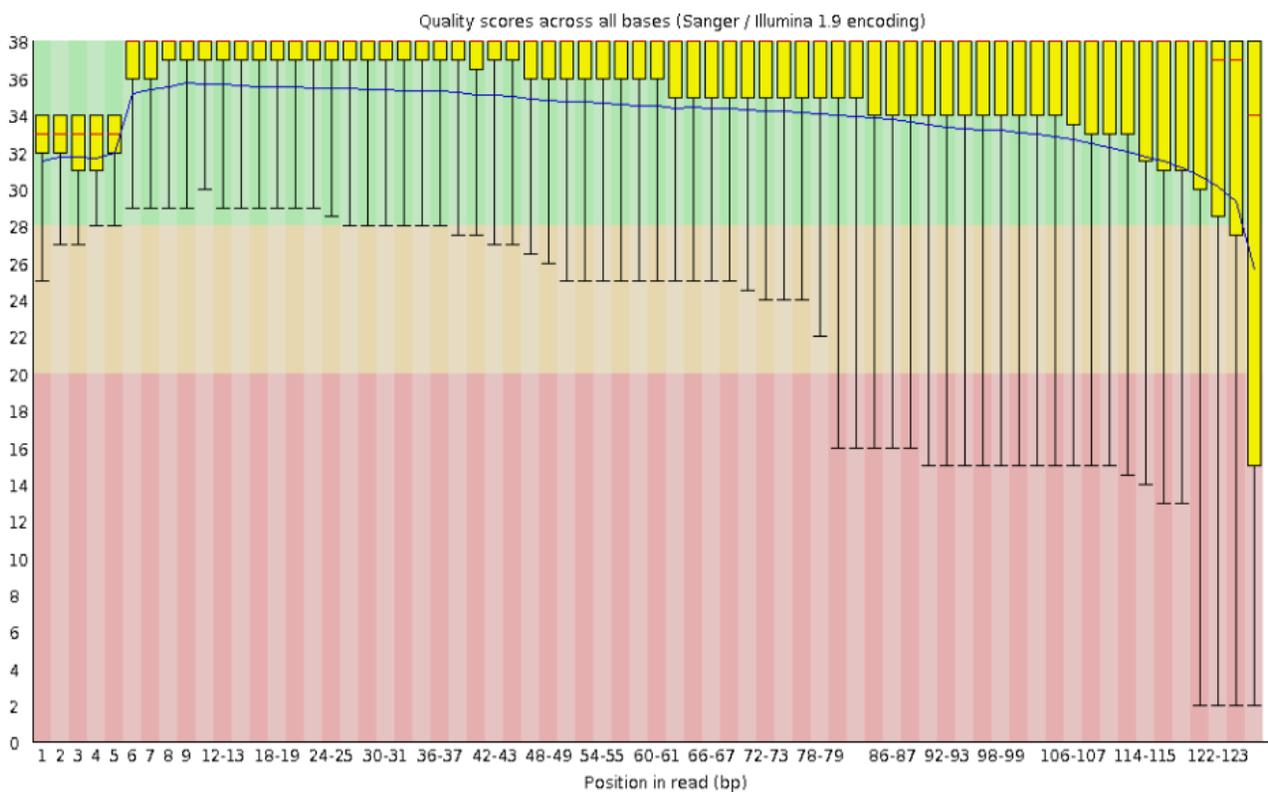
### TBM patient D5371: Per base sequence quality



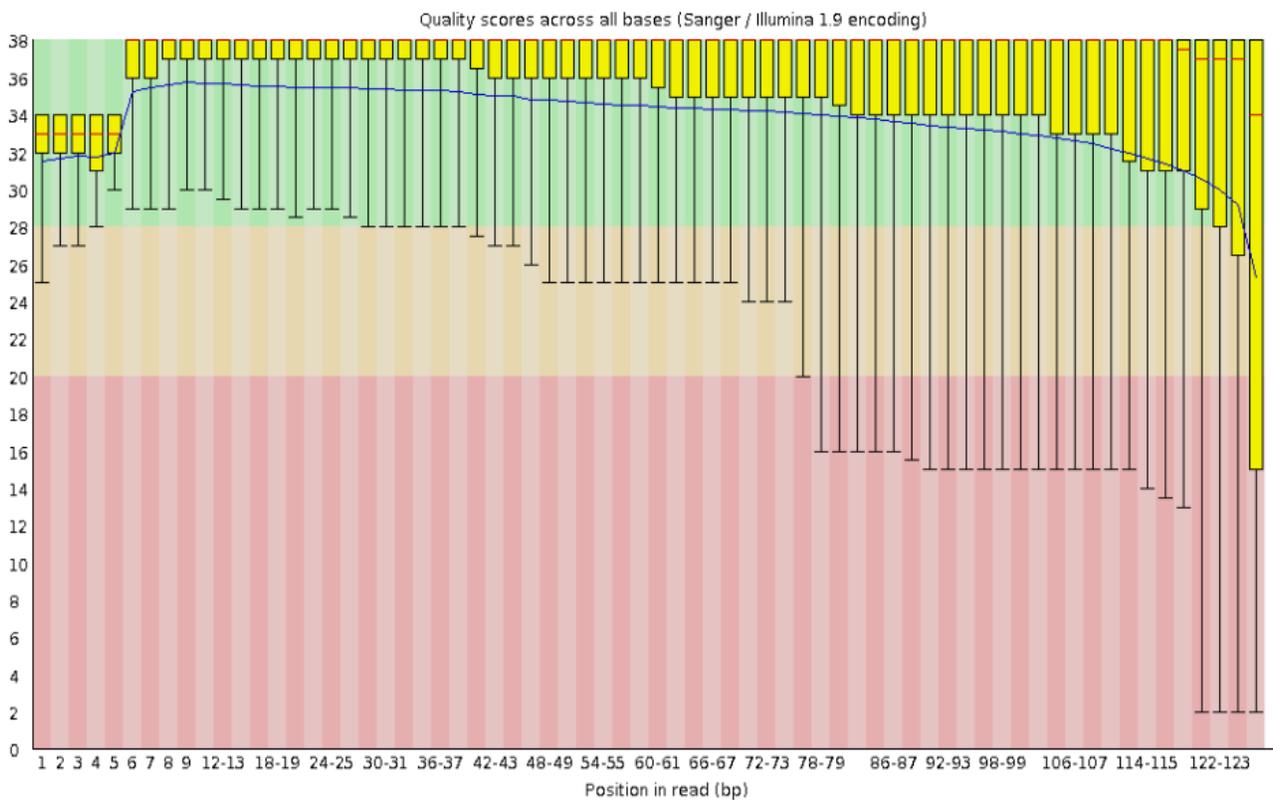
### TBM patient D5372: Per base sequence quality



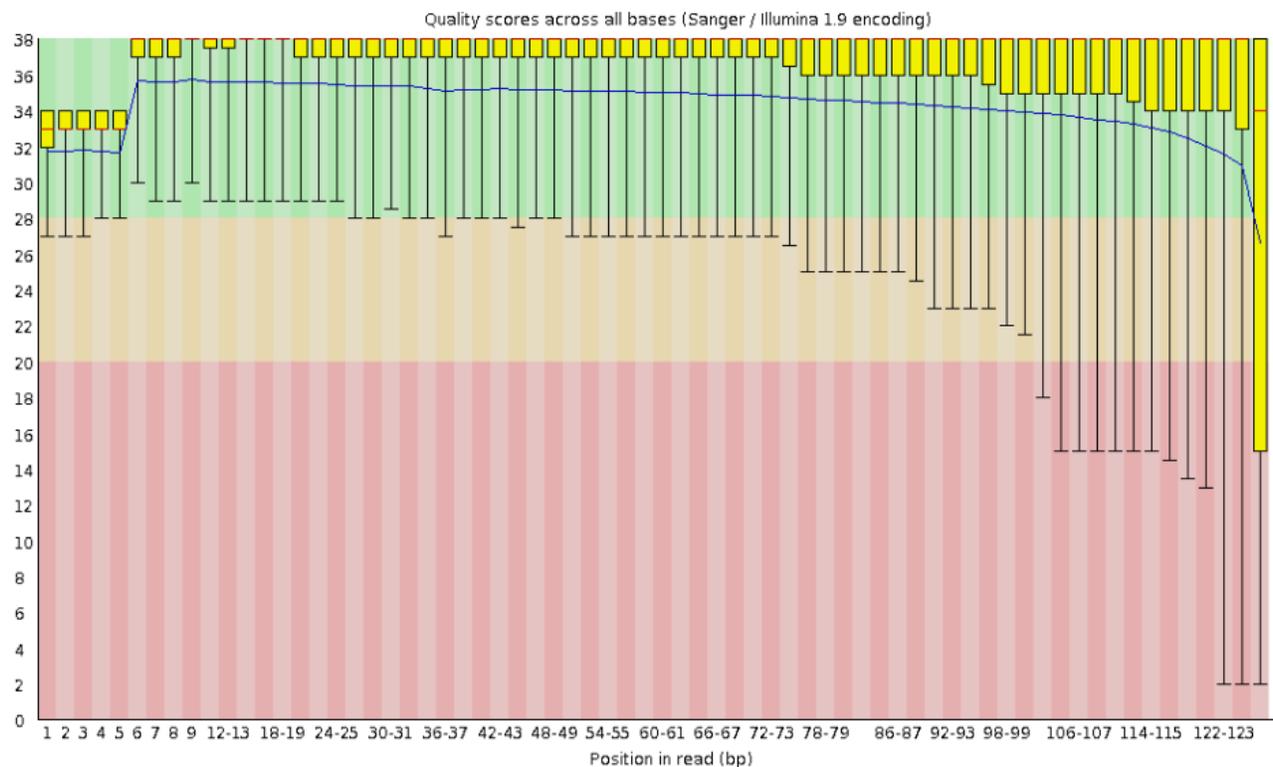
### TBM patient D5373: Per base sequence quality



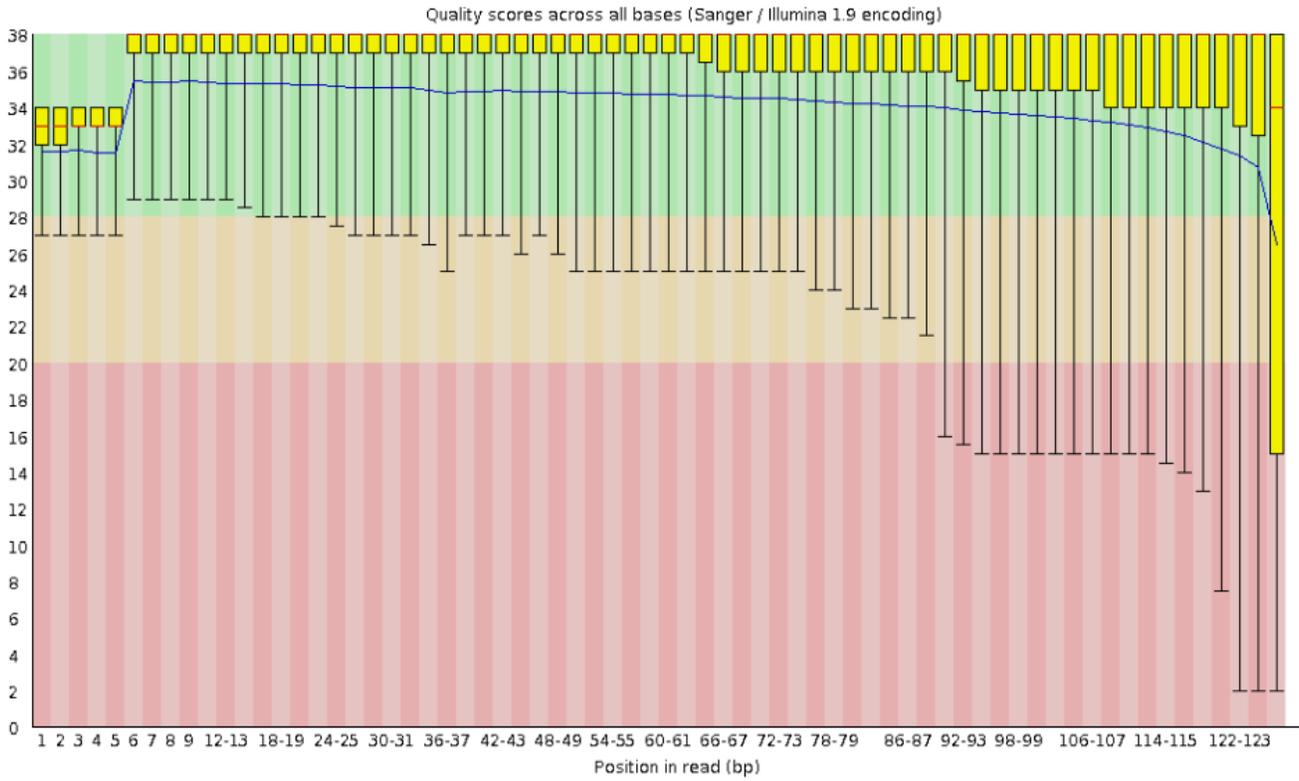
### TBM patient D5374: Per base sequence quality



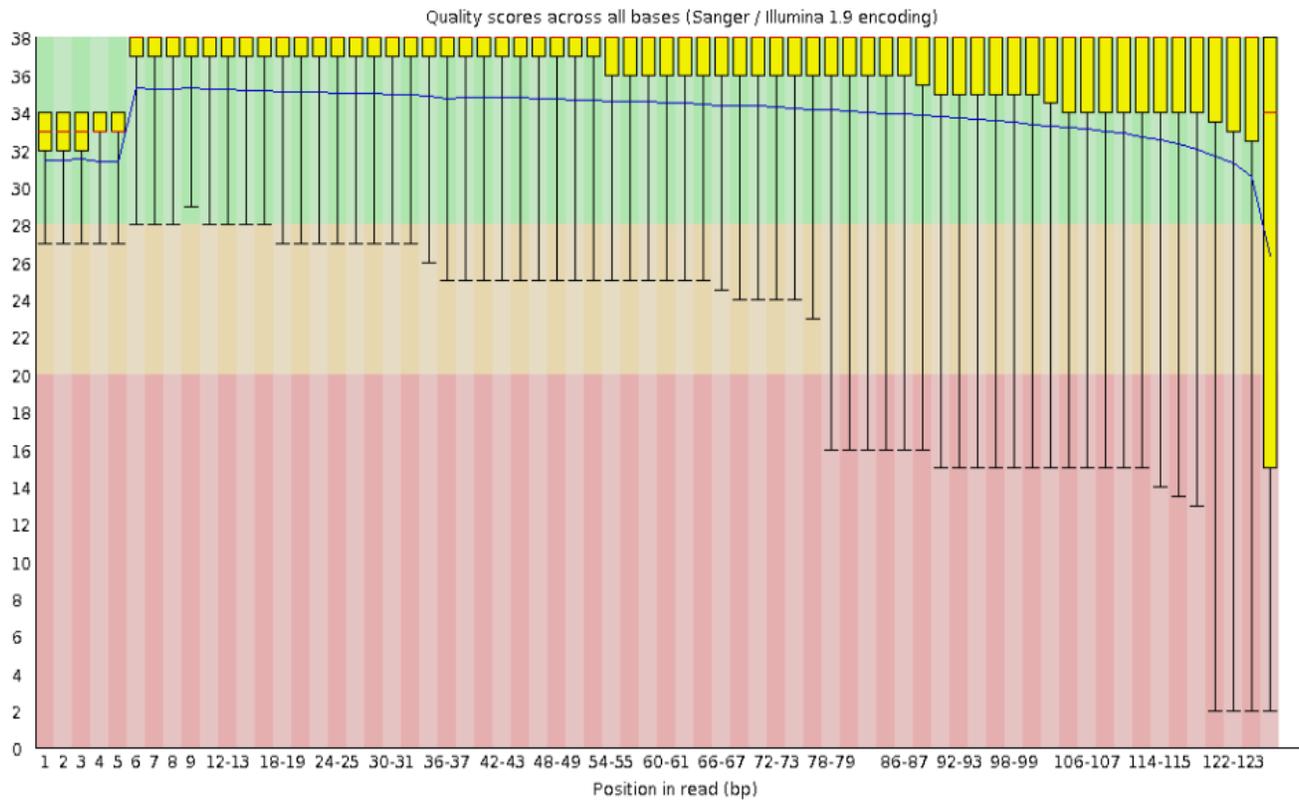
### TBM patient D5375: Per base sequence quality



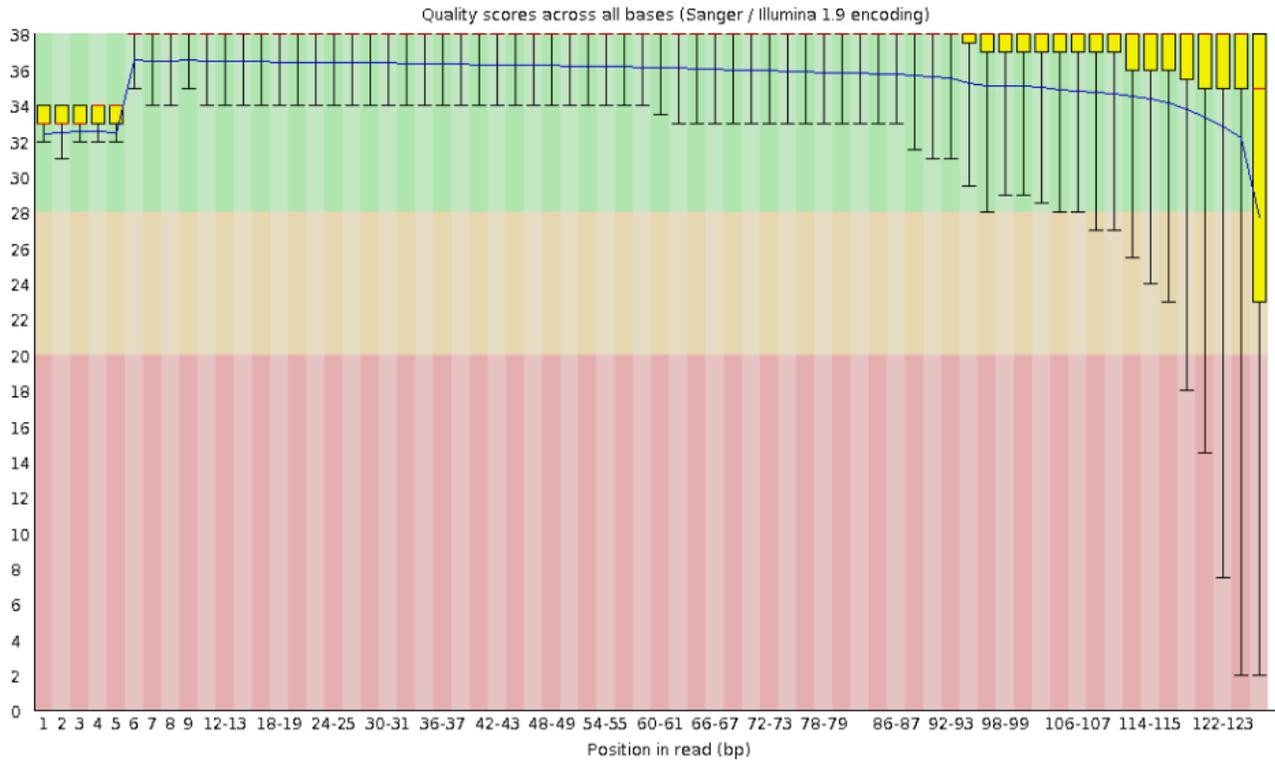
### TBM patient D5376: Per base sequence quality



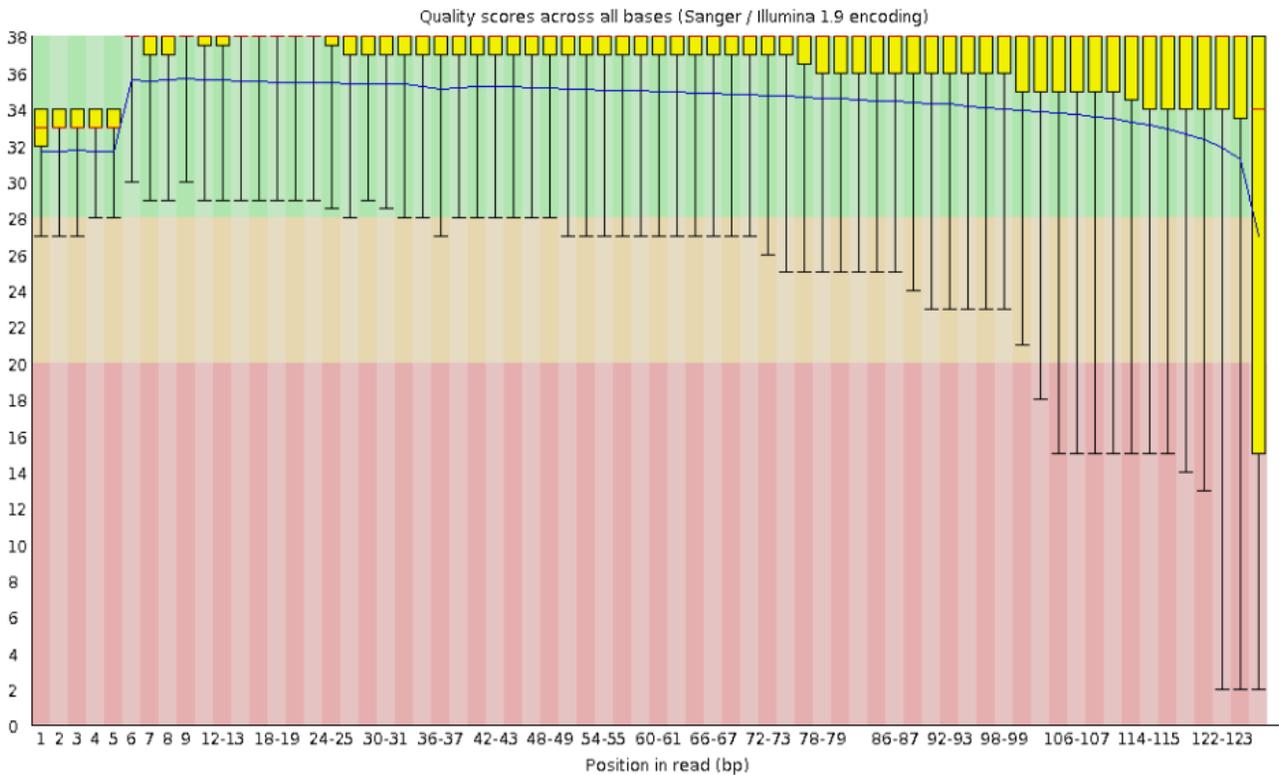
### TBM patient D5377: Per base sequence quality



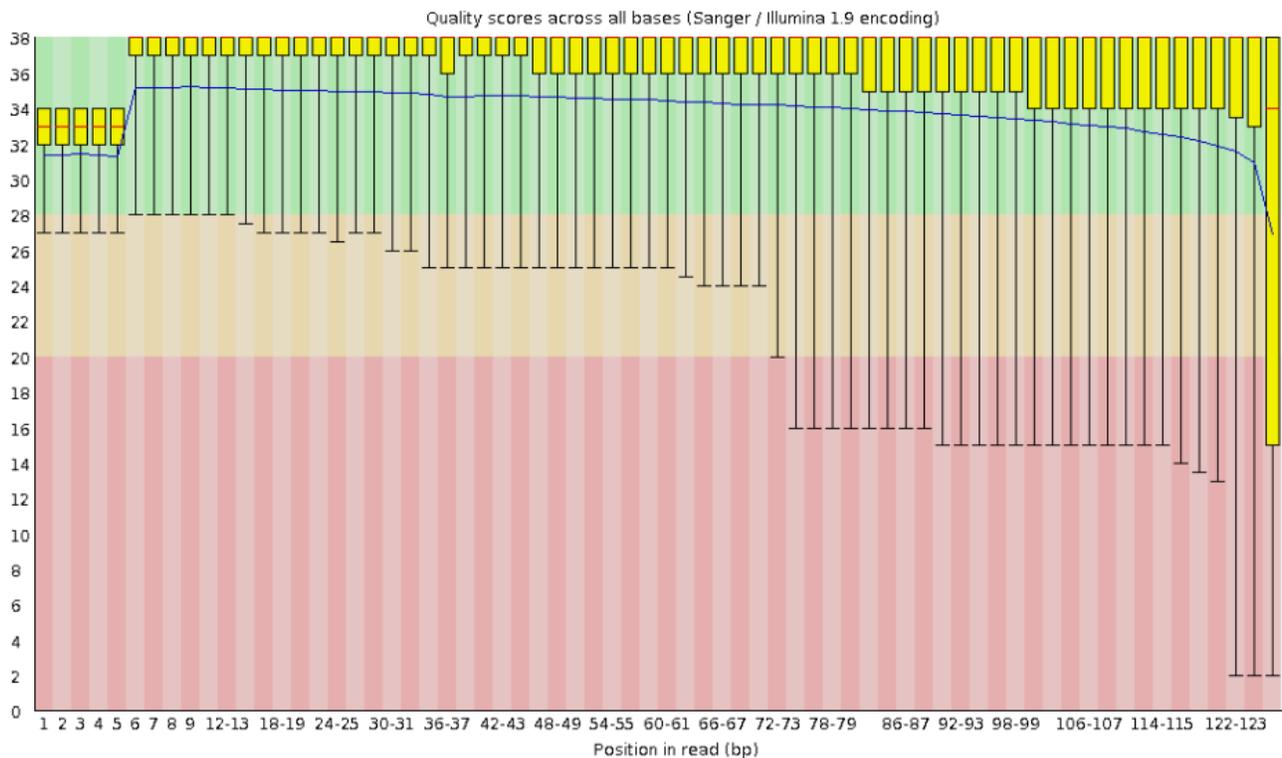
### TBM patient D5378: Per base sequence quality



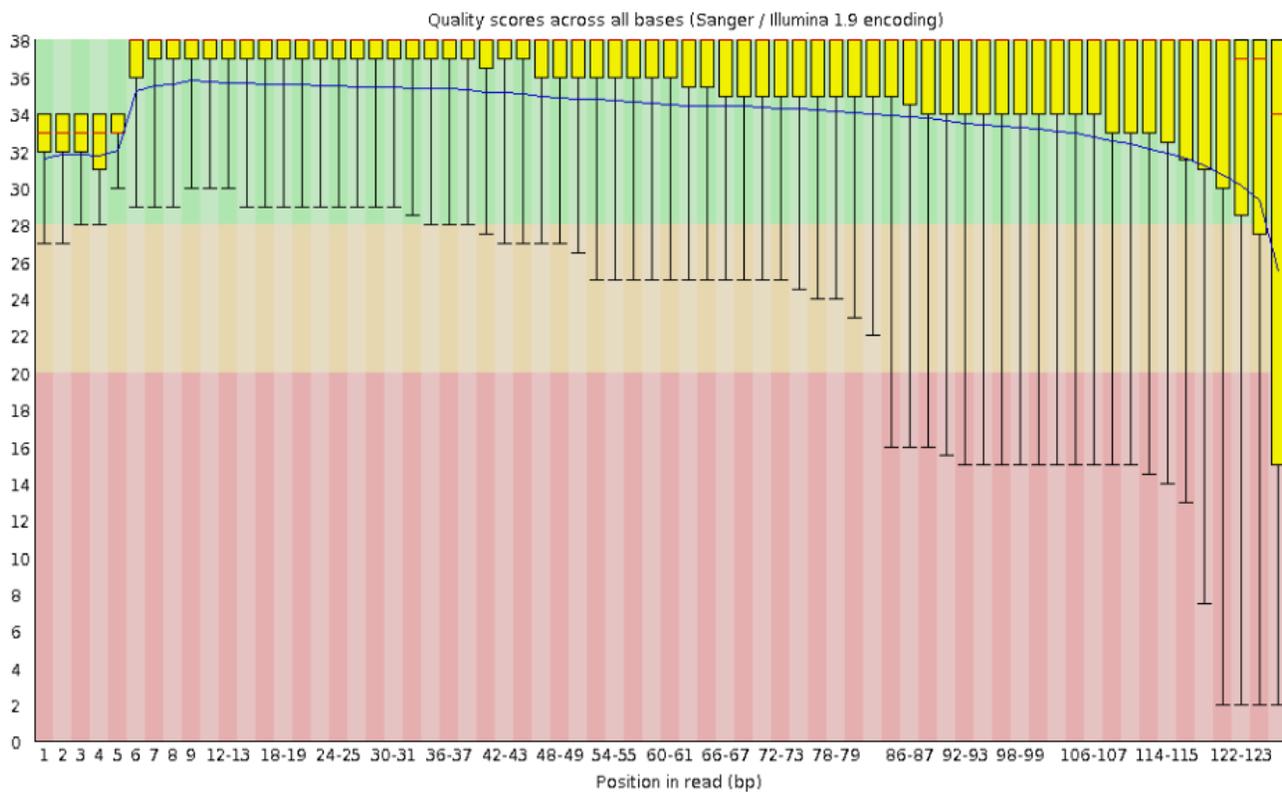
### TBM patient D5379: Per base sequence quality



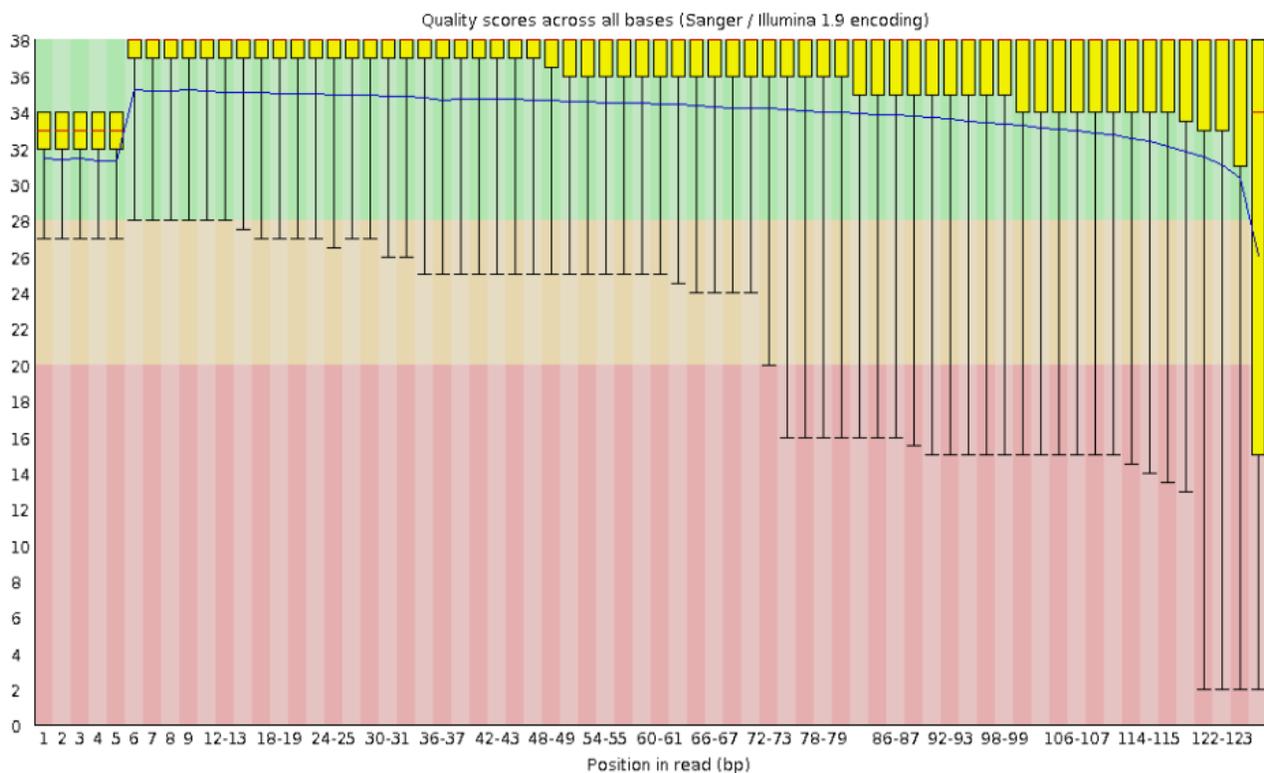
### TBM control D5380: Per base sequence quality



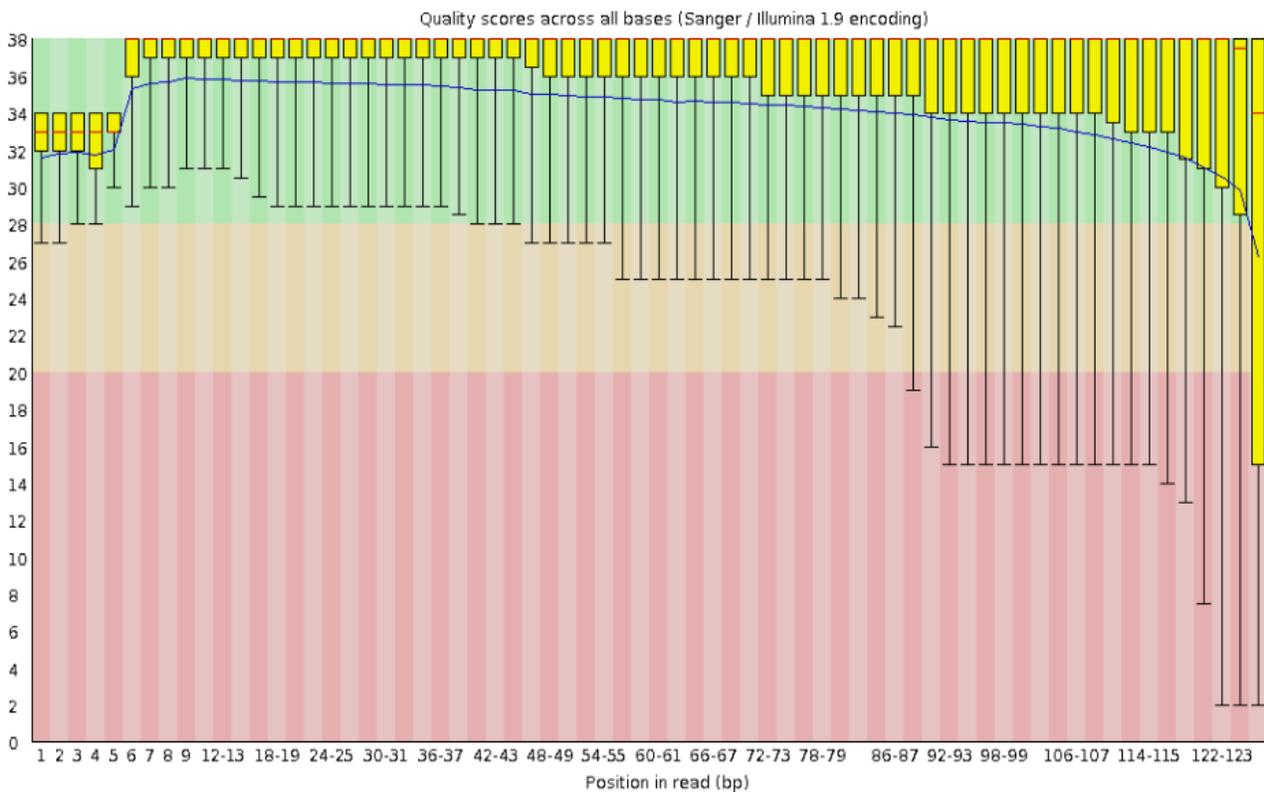
### TBM control D5381: Per base sequence quality



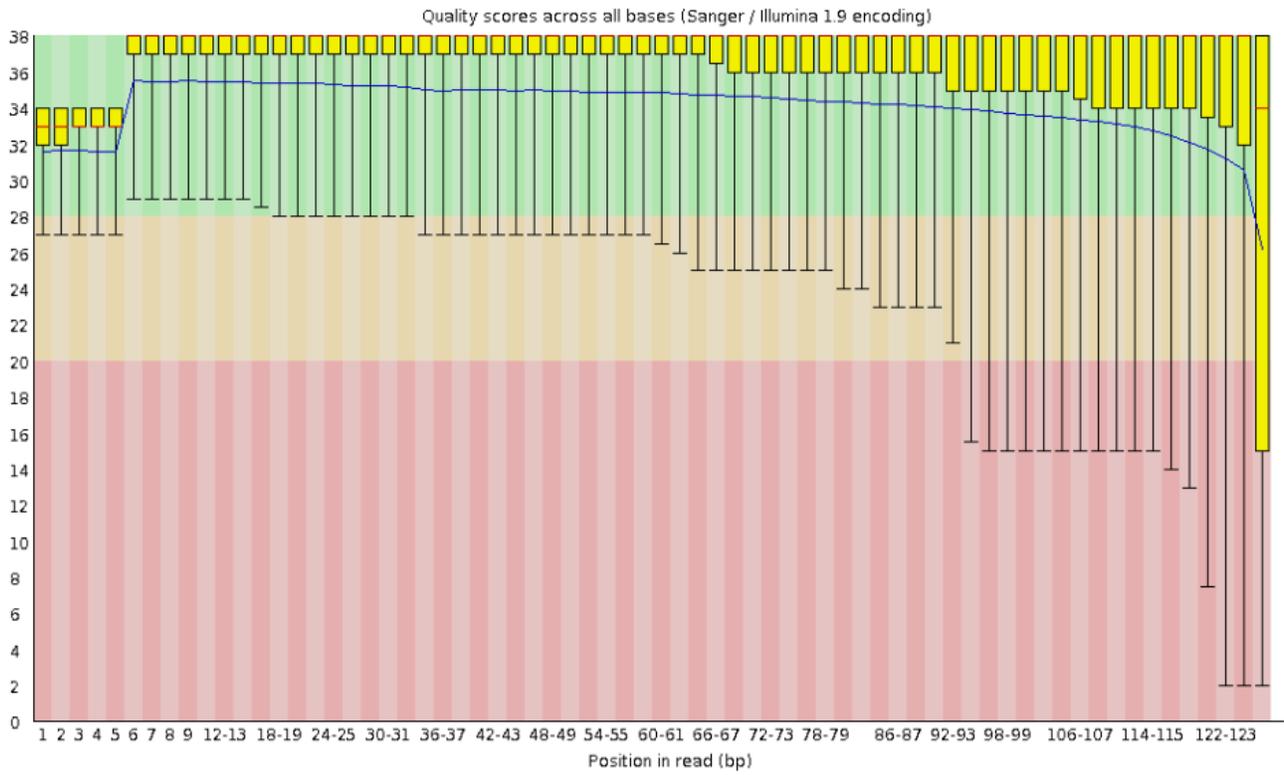
### TBM control D5382: Per base sequence quality



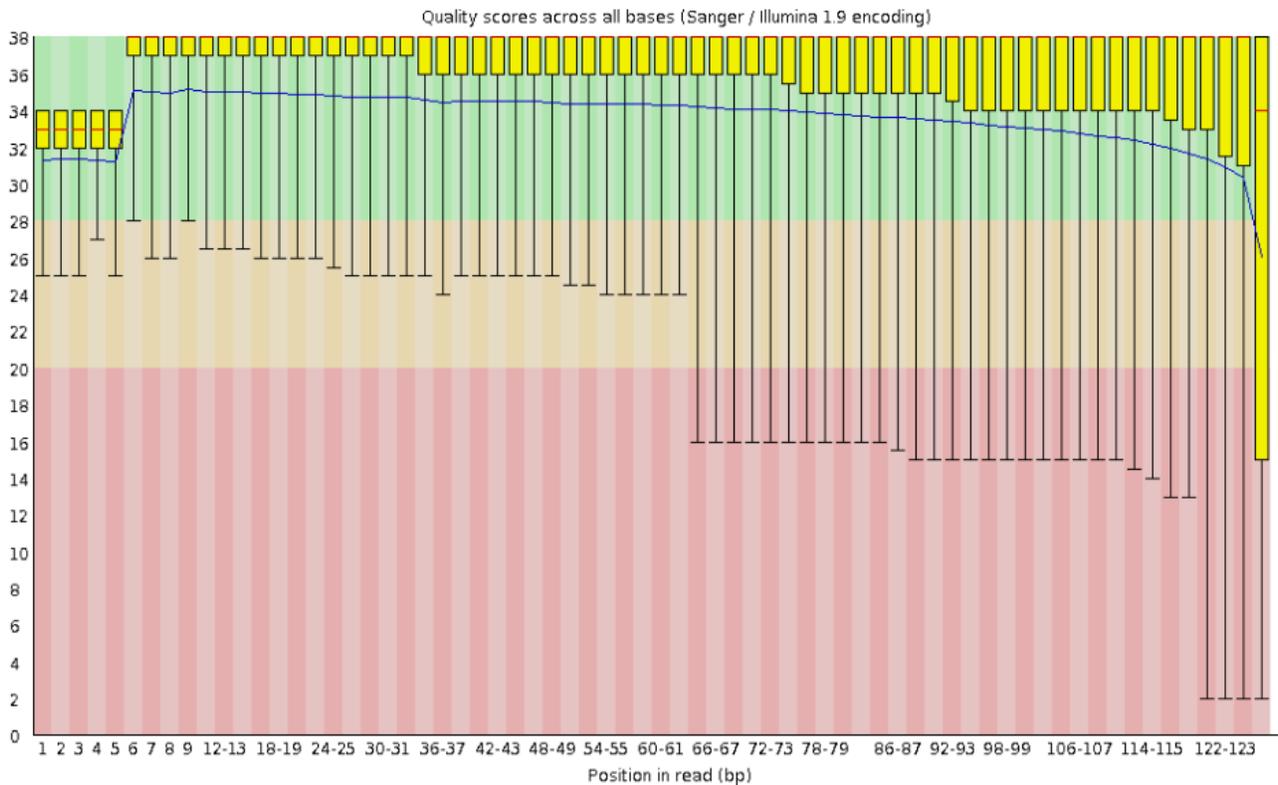
### TBM control D5383: Per base sequence quality



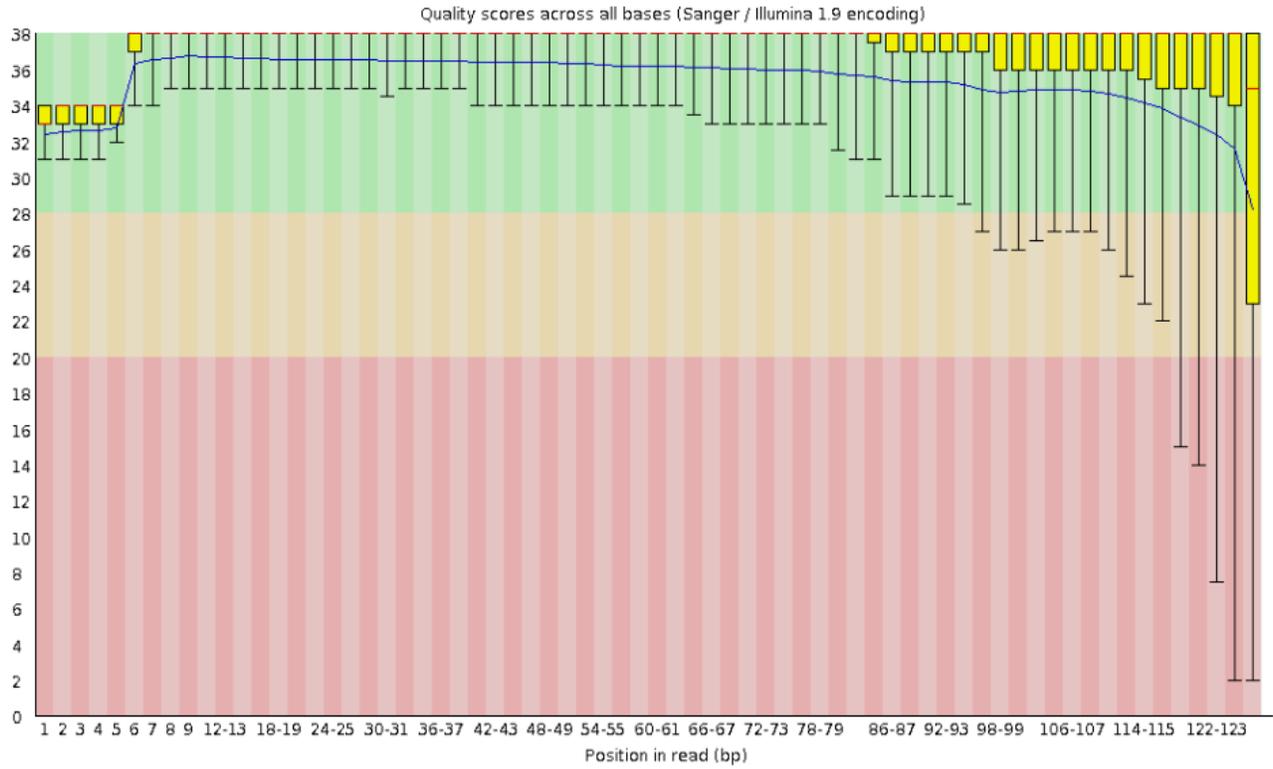
### TBM control D5384: Per base sequence quality



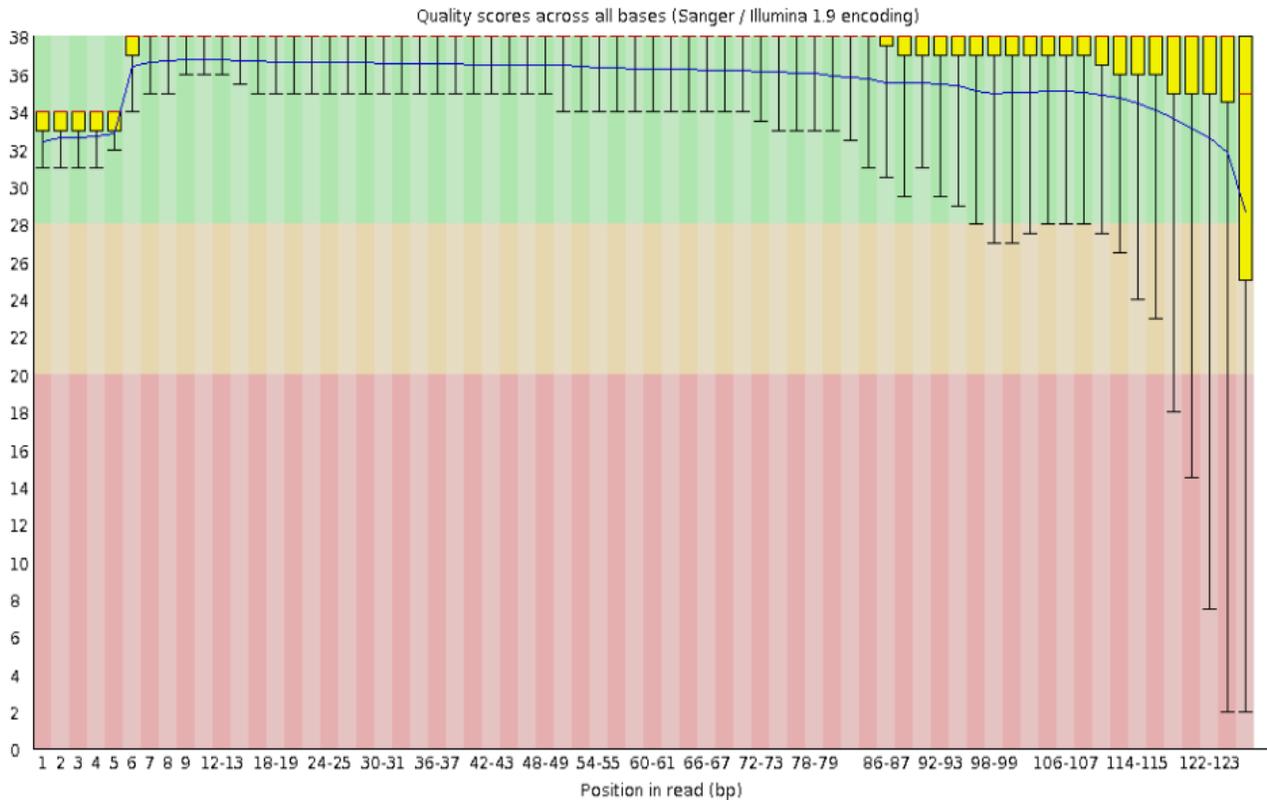
### TBM control D5385: Per base sequence quality



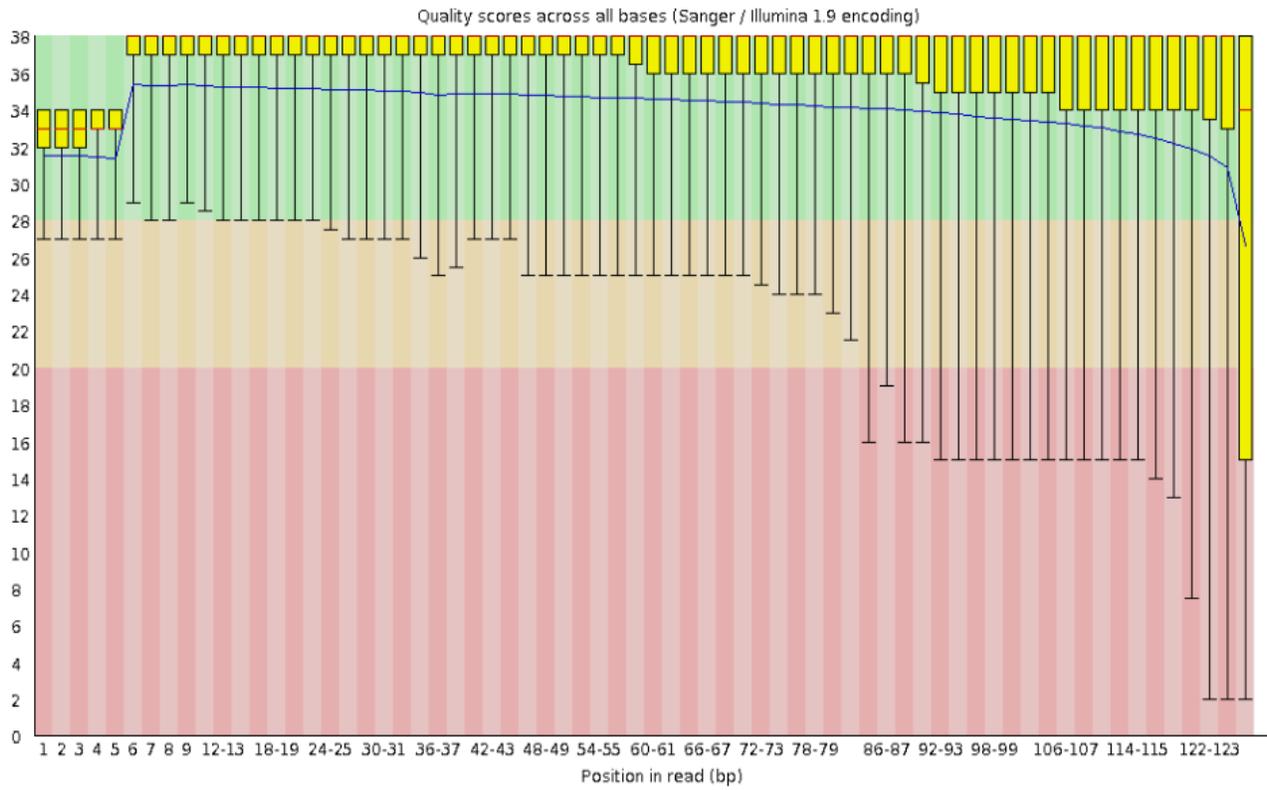
### TBM control D5386: Per base sequence quality



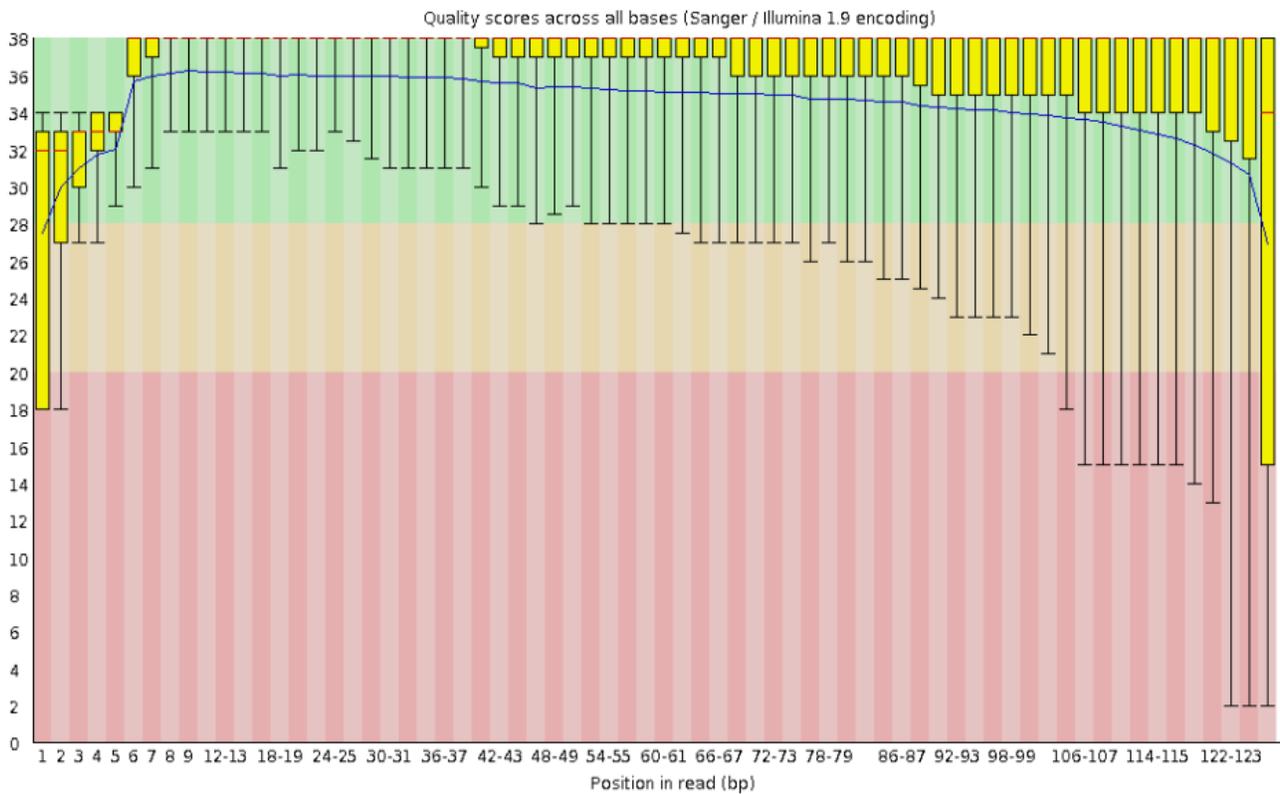
### TBM control D5387: Per base sequence quality



### TBM control D5388: Per base sequence quality



### TBM control D5389: Per base sequence quality



APPENDIX III – **Supplementary Table 1: Software packages used for Bioinformatic analysis of exome sequencing data.**

Software packages	Versions	Important parameters	Interpretation
FastQC	0.11.2		
Burrows-Wheeler Aligner (BWA)	0.6	bwa aln: -n 2 -q 15 -l 5000 -t 8 bwa sampe: -a 500	Align sample to reference Maximum edit distance of 2 Parameter of 15 for read trimming Take the first 5000 subsequence as seed Number of threads is 8 Convert to SAM format Maximum insert size of 500 for read pair to be considered being mapped properly
Picard	1.55	FixMateInformation.jar:  VALIDATION_STRINGENCY=LENIENT CREATE_INDEX=true  MarkDuplicates.jar: REMOVE_DUPLICATES=true ASSUME_SORTED=true VALIDATION_STRINGENCY=SILENT  OPTICAL_DUPLICATE_PIXEL_DISTANCE=10 CollectHsMetrics.jar:  VALIDATION_STRINGENCY=LENIENT	Verify mate-pair information between mates and fix if needed Validation stringency for all SAM files is lenient Creates a BAM index when writing a coordinate-sorted BAM file Identifies duplicate reads Removes duplicate reads Sort order in the header file will be ignored Validation stringency for all SAM files is silent, to improve performance when processing BAM file in which variable-length data do not need to be decoded Maximum offset of 10 between two duplicate clusters to consider them optical duplicates. Calculates set of Hybrid Selection specific metrics from aligned SAM/BAM file. Validation stringency for all SAM files is lenient
SamTools	0.1.19	Mpileup:  -q 15 -Q 20  View: -b  -q 1	Generate VCF, BCF or pileup for one or multiple BAM files. Alignment records are grouped by sample identifiers in @RG header lines. Minimum mapping quality of 15 for an alignment to be used Minimum base quality of 20 for a base to be considered  Prints all alignments in specified input alignment file (SAM, BAM, CRAM format) to BAM output Skip alignments with MAPQ smaller than 1
Genome analysis toolkit (GATK)	3.3	Local realignment: -T RealignerTargetCreator	Define intervals to target for local realignment

	<p>-T IndelRealigner</p> <p>Base quality score recalibration:</p> <p>-T BaseRecalibrator</p> <p>-l INFO</p> <p>-T PrintReads</p> <p>-l INFO</p> <p>Variant calling:</p> <p>-T UnifiedGenotyper</p> <p>-stand_call_conf 50.0</p> <p>-stand_emit_conf 10.0</p> <p>-dcov 1000</p> <p>-mbq 10</p> <p>-glm BOTH</p> <p>-nt 4</p> <p>Separate SNPs and InDels:</p> <p>-T SelectVariants</p> <p>-selectType SNP</p> <p>-selectType INDEL</p> <p>SNP Recalibration:</p> <p>(1) Generate tranches file, recal file and plots for snps:</p> <p>-T VariantRecalibrator</p> <p>-mode SNP</p> <p>--maxGaussians 4</p> <p>--percentBadVariants 0.05</p> <p>-an QD</p> <p>-an HaplotypeScore</p> <p>-an MQRankSum</p> <p>-an ReadPosRankSum</p> <p>-an FS</p> <p>-an MQ</p> <p>(2) Perform snp recalibration + filter:</p> <p>-T ApplyRecalibration</p>	<p>Perform local realignment of reads around indels</p> <p>Detect systematic errors in base quality scores</p> <p>Set the minimum level of logging. INFO = default value.</p> <p>Manipulate sequencing data in SAM/BAM format.</p> <p>Set the minimum level of logging. INFO = default value.</p> <p>Call SNPs and indels on per-locus basis</p> <p>Minimum phred-scaled confidence threshold of 50 at which variants should be called</p> <p>Minimum phred-scaled confidence threshold of 10 at which variants should be emitted</p> <p>Target coverage threshold of 1000 for downsampling to coverage</p> <p>Minimum base quality of 10 required to consider a base for calling</p> <p>Genotype likelihoods calculation model to employ -- BOTH to call SNPs and INDELS together</p> <p>Number of data threads to allocate to this analysis</p> <p>Select a subset of variants from a larger callset</p> <p>Select only SNPs</p> <p>Select only INDELS</p> <p>Build a recalibration model to score variant quality for filtering purposes</p> <p>Recalibrating SNP</p> <p>Max number of Gaussians for the positive model</p> <p>Percentage of variants used to train the negative model</p> <p>-an indicates all annotations to be used for calculations</p> <p>QD: quality by depth</p> <p>MQ: mapping quality</p> <p>FS: strand bias</p> <p>MQ: mapping quality</p> <p>Apply a score cutoff to filter variants based on</p>
--	--	---

		<p>--ts_filter_level 99.0</p> <p>INDEL Recalibration:  (1) Generate tranches file, recal file and plots for indels:  -T VariantRecalibrator</p> <p>-mode INDEL  --maxGaussians 4</p> <p>-an QD  -an ReadPosRankSum  (2) Perform indel recalibration:  -T ApplyRecalibration</p> <p>--ts_filter_level 99.0</p> <p>(3) Perform indel filtration:  -T VariantFiltration</p> <p>-filter "QD &lt; 2.0"  -filterName "QDFilter"</p>	<p>a recalibration table  The truth sensitivity level at which to start filtering</p> <p>Build a recalibration model to score variant quality for filtering purposes</p> <p>Recalibrating INDELS  Max number of Gaussians for the positive model</p> <p>Use quality by depth and ReadPosRankSum for calculations</p> <p>Apply a score cutoff to filter variants based on a recalibration table  The truth sensitivity level at which to start filtering</p> <p>Filter variant calls based on INFO and FORMAT annotations  Filter out quality by depth &lt;2  Names to use for the list of filters</p>
--	--	---	---

APPENDIX IV – **ANNOVAR explained**

ANNOVAR Annotated File Header	Explanation
Func	Function of the variant -- exonic, intronic, UTR, etc.
Gene	Gene name for variant
ExonicFunc	If the variant is exonic, synonymous, non-synonymous, indel, etc.
AAChange	If exonic, variant change in nucleotide and protein format
Conserved	
SegDup	Indicates if the variant is located in a segmental duplication region
ESP5400_ALL	MAF in Exome Sequencing Project dataset (5,400 exomes) for all populations
1000g2012feb_ALL	MAF in 1000Genomes February 2012 release
dbSNP135	RS# from the dbSNP database
AVSIFT	SIFT Pathogenicity score: closer to 0 is more damaging
LJB_PhyloP	Pathogenicity score from dbNSFP: conserved > 0.95, not conserved < 0.95
LJB_PhyloP_Pred	Pathogenicity call from dbNSFP: C - conserved, N - not conserved
LJB_SIFT	Pathogenicity score from dbNSFP: tolerated < 0.95, deleterious > 0.95
LJB_SIFT_Pred	Pathogenicity call from dbNSFP: T - tolerated, D - deleterious
LJB_PolyPhen2	Pathogenicity score from dbNSFP: probably damaging > 0.85, possibly damaging 0.85-0.15, benign < 0.15
LJB_PolyPhen2_Pred	Pathogenicity call: D - probably damaging, P - possibly damaging, B - benign
LJB_LRT	Pathogenicity probability score from dbNSFP: closer to 1 is more likely to be damaging -- see below
LJB_LRT_Pred	Pathogenicity score from dbNSFP: D - deleterious fulfills the following: (i) from a codon defined by LRT as significantly constrained (LRTorio0.001 and oo1), (ii) from a site with Z10 eutherian mammals alignments, and (iii) the alternative AA is not presented in any of the eutherian mammals N - otherwise neutral
LJB_MutationTaster	Pathogenicity probability score from dbNSFP: closer to 1 is more likely to be damaging -- see below
LJB_MutationTaster_Pred	Pathogenicity score from dbNSFP: automatically calculated categories: "disease_causing_automatic," "disease_causing," "polymorphism," and "polymorphism_automatic," which we coded as "A," "D," "N," and "P,"
LJB_GERP++	Nucleotide conservation score from dbNSFP GERP: Higher number is more conserved, > 0 is generally conserved
Chr	Chromosome of variant
Start	Start coordinate of variant
End	End coordinate of variant
Ref	Reference allele for variant
Obs	Observed allele for variant
Otherinfo	Other information

APPENDIX V – **Reagents****1 % Agarose Gel**

Agarose	1g
SB Buffer (1X)	100mL
Microwave 2-3 minutes	
Add 4µl ethidium bromide (10mg/mL) when temperature reaches ± 55°C	

**10% Cryomedia**

Foetal Bovine Serum	500mL
Dimethyl sulphoxide	50mL

**2% FACS Buffer**

Foetal Bovine Serum	20mL
Phosphate buffered saline (PBS)	980mL

**Bromophenol Blue Loading Dye**

Bromophenol blue 1% (w/v)  
Deionized water to a final volume of 100mL  
Store at room temperature

**CAP Buffer**

60 mM CaCl <sub>2</sub>	2.21 g
15% Glycerol	37.5mL
10mM PIPES	0.76 g
Deionized water to final volume of 250mL	
pH to 7.0	
Store at 4°C	

***E.coli* strain DH5α**

Φ 80d *lacZΔM15* recA1, *endA1*, *Gry A96 thi-1*, *hsdR17 supE44*, *relA1*, *deoR* Δ  
(*lacZYA argF*) u169

**Ethidium bromide stock (10mg/mL)**

Ethidium bromide (Sigma)	500mg
Deionized water	50mL
Stir for 4 hours using magnetic stirrer	
Store in dark container at room temperature	

**Growth Media**

DMEM (4.5g/L glucose, with L-glutamine)	500mL
Foetal Bovine Serum (FBS)	50mL
Penicillin/Streptomycin	5mL
Pre-warm to 37°C before use	

**LB Agar Plates**

Bacto tryptone	5g
Yeast extract	2g
NaCl	5g
Agar	8g
Deionized water to final volume of 500mL	
Autoclave for 20 minutes at 121°C	
Add 5mg/mL Kanamycin to media when the temperature has reached $\pm 55^{\circ}\text{C}$ , prior to pouring plates	

**Luria-Bertani (LB) Media**

Bacto tryptone	5g
Yeast extract	2g
NaCl	5g
Agar	8g
Deionized water to final volume of 500mL	
Autoclave for 20 minutes at 121°C	
Add 5mg/mL Kanamycin to media when the temperature has reached $\pm 55^{\circ}\text{C}$	

**Lysis buffer**

Hepes	10mL
5M NaCl	4mL
0.5M EDTA	4mL
Triton X-100	2mL
10mM Nappi	80mL
1M $\text{Na}_3\text{VO}_4$	400 $\mu\text{l}$
Deionized water to final volume of 200mL	
<ul style="list-style-type: none"> <li>• <math>\frac{1}{4}</math> tablet protease inhibitor freshly added to 5mL lysis buffer before use</li> </ul>	

**Membrane blocking solution**

Bovine serum albumin	25g
Tween-20	5 $\mu\text{l}$
TBST solution	500mL

**PBMC Growth Media**

RPMI	500mL
Foetal Bovine Serum	50mL
Penicillin/Streptomycin	5mL
L-Glutamine, 200mM	500uL

Pre-warm to 37°C before use

**Phosphate buffered saline (PBS)**

1 PBS Tablet	
Deionized water to a final volume of 200mL	

**SB Buffer (20X stock)**

Di-sodium tetraborate decahydrate	38.137g/mol
Deionized water to a final volume of 1L	

**SDS Loading Dye**

LaemmLi sample buffer (Bio-Rad)	950µl
β-mercapto-ethanol	50µl

**SDS-PAGE Running buffer (10X)**

Tris base	30g
Glycine	144g
10% SDS buffer	100mL
Deionized water to a final volume of 1L	

**TBST**

5M NaCl	60mL
1M Tris base	40mL
0.1% Tween-20	2mL
pH to 7.6	
Deionized water to a final volume of 2L	

## APPENDIX VI – Cell Counting with Hemocytometer

### 1. HAEMOCYTOMETER PREPARATION

- Clean haemocytometer and cover slip with 70% Ethanol and a paper towel
- Do not rub too roughly over the counting chamber area as it may scratch the surface
- To apply the cover glass, apply a thin layer of moisture in the mounting support by breathing onto haemocytometer and drying off counting chamber area, while leaving the breath condensate on the cover glass mounting support
- Carefully but firmly slide the cover glass vertically onto haemocytometer
- Ensure cover glass is set in place and does not fall off when tilted

### 2. CELL PREPARATION

- Prepare a 1:10 dilution of cells:trypan blue:
  - Pipet 10 $\mu$ L of cells (re-suspended in cryomedia) into a sterile eppendorf tube
  - Add 80 $\mu$ L of sterile 1 $\times$ PBS
  - Add 10 $\mu$ L of trypan blue
  - Mix thoroughly
- Add 10 $\mu$ L of stained cells to prepared haemocytometer using fine-tipped Pasteur pipette
- Gently touch pipette against side of cover slip where it touches the base of the depression (Figure 1). A drop of liquid should be drawn out of pipette and into counting chamber. If not, gently and slowly squeeze out one drop, which will fill one counting chamber.
- Ensure that no bubbles are present so that cells are distributed as evenly as possible

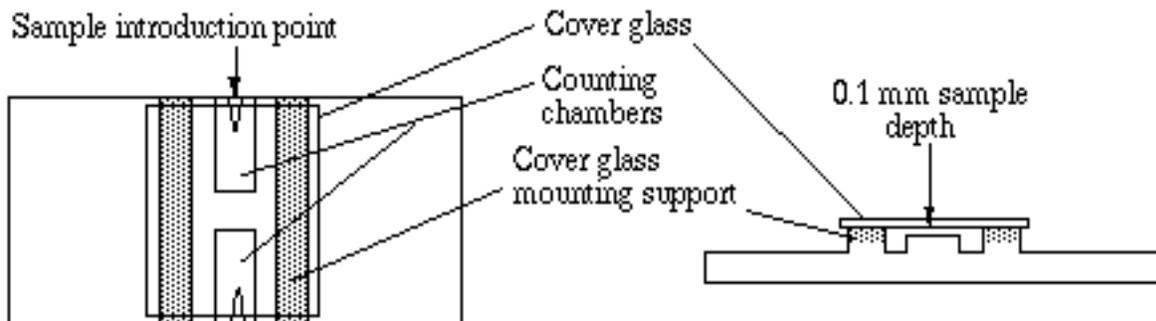
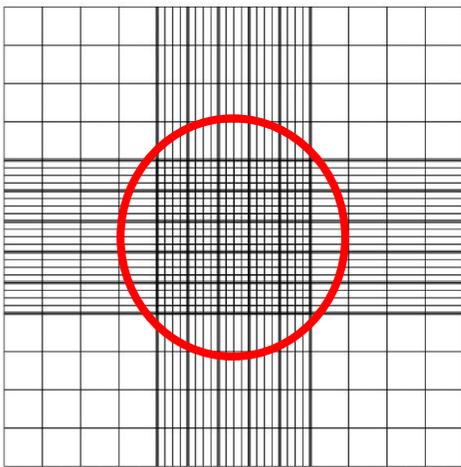


Figure 1: Illustration of Haemocytometer.

### 3. MANUAL CELL COUNTS

- Switch on microscope and place haemocytometer on the stage
- Set microscope to Phase 1 (Ph1) and adjust light source
- Use 10× or 20× objectives to orientate haemocytometer in correct position and to check that there is not too much neutrophil or platelet contamination
- Set microscope to 40× objectives to start cell counting
- Live PBMCs will appear white, round with a 'shiny halo' around periphery
- Dead cells will stain blue (entire cytoplasm will be blue)
- Count cells in the middle grid (Figure 2)
- Count cells touching middle line on top and left and ignore those touching middle line on bottom and right
- Document live and dead cell count and determine cell concentration (total cell count/ml) using standard cell counting formula:

*Concentration (Total cell count/ml) = haemocytometer count  $\times 10^4 \times$  dilution factor (dilution made with Trypan blue)*



**Figure 2: Counting chamber grid.**

APPENDIX VII – **Supplementary Table 2: MSMD-associated genes investigated in TBM patients**

Gene name	Abbreviation
Interferon gamma receptor 1	IFGR1
Interferon gamma receptor 2	IFGR2
Interleukin 12 beta	IL12B
Interleukin 12 receptor beta 1	IL12RB1
Signal transducer and activator of transcription 1	STAT1
Interferon-stimulated genes 15	ISG15
Interferon regulatory factor 8	IRF8
NF-kappa-B essential modulator/ Inhibitor of nuclear factor kappa-B kinase subunit gamma	NEMO/IKBKG
Cytochrome b beta	CYBB
Tyrosine kinase 2	TYK2
Transporter associated with antigen processing 1	TAP1
Mitogen activated protein kinase kinase kinase 14	MAP3K14

APPENDIX VIII – **List of suppliers**

3-Amino-1, 2, 4-triazole	Sigma
Agarose	Whitehead scientific
Ampicillin	Roche
Autoradiography film	Thermo Scientific
β-mercapto-ethanol	Sigma
Bacto Agar	Merck
Bacto tryptone	Fluka
BigDye® Terminator v3.1 Cycle Sequencing kit	Perkin-Elmer
Bradford reagent	Thermo Scientific
Bromophenol blue	Promega
CaCl <sub>2</sub>	Merck
dATP	Promega
dCTP	Promega
dGTP	Promega
Dimethylformamide	Merck
Di-Sodium tetraborate decahydrate	Merck
DMEM	Whitehead scientific
dTTP	Promega
EDTA	Boehringer Mannheim
Ethanol	Sigma
Ethidium bromide	Roche
Foetal bovine serum	The Scientific Group
Glass beads	Amersham pharmacia
Glucose	Kimix
Glycine	Sigma
Greiner tubes	Greiner Bio-one
HCl	Merck
Isopropanol	Merck
KCl	Merck
Leammi sample buffer	Bio-Rad
Lipofectamine™ LTX Reagent	Invitrogen
MgCl <sub>2</sub>	Bioline
Mini-PROTEAN® TGX™ precast polyacrylamide gels	Bio-Rad Laboratories
Molecular size marker (100 bp)	Lasec
Multi-Ethnic Genotyping Array	Illumina
<i>MscI</i>	Promega
NaCl	Sigma
NaHep/EDTA vacutainers	Greiner Bio-One

*NdeI*

Nucleon BACC3 kit

Oligonucleotide Primers

PBS

Penicillin/Streptomycin

Peptone

PIPES

PMSF

Protein G agarose

Protein ladder

PureYield™ Plasmid Miniprep System

Q5® Site-Directed Mutagenesis Kit

SDS

Secondary antibodies

SuperSignal® West Pico Chemiluminescent Substrate kit

Spectra™ Multicolor Broad Range Protein Ladder

T4 DNA Ligase

Taq polymerase

Tris

Tris-HCl

Triton X-100

Trypsin

Tryptone

Tween-20

Universal DNA ladder

Wizard® SV Gel and PCR Clean-up System

X-α-Galactosidase

Yeast extract

Promega

GE Healthcare

Department of Molecular and Cell  
Biology, University of Cape Town  
(UCT) Cape Town, RSA

Sigma

The Scientific Group

Difco

Merck

Roche

Whitehead scientific

Thermo Scientific

Promega

New England BioLabs

Sigma

Santa Cruz Biotechnology

Thermo Scientific

Thermo Scientific

Promega

Bioline

Merck

Merck

Sigma

Whitehead scientific

Fluka

Merck

KAPA Biosystems

Promega

Southern Cross

Difco

APPENDIX IX – **Supplementary Figure 1: Validation of candidate variants through Sanger sequencing**

**(A) *TSPAN33* (G331A)**

Wild-type CCTGCTGCAGCTGGCC**G**CTGGGATCCTGGGCT  
 PID\_012 Forward CCTGCTGCAGCTGGCC**A**CTGGGATCCTGGGCT  
 PID\_012 Reverse CCTGCTGCAGCTGGCC**A**CTGGGATCCTGGGCT



**(B) *JAK3* (G2753C)**

Wild-type ACAATACAGGTGCCCT**G**GTGGCCGTGAAACAG  
 Mother Forward ACAATACAGGTGCCCT**G**GTGGCCGTGAAACAG  
 Mother Reverse ACAATACAGGTGCCCT**G**GTGGCCGTGAAACAG  
 PID\_012 Forward ACAATACAGGTGCCCT**G**GTGGCCGTRAAACAG  
 PID\_012 Reverse ACAATACAGGTGCC**Y**T**G**GTGGCCGTGAAACAG



**(C) *TAPI* (A888G)**

Wild-type .TGTCATTCTCACCAT**A**GCCAGTGCAGTGCTG  
 Mother Forward .TGTCATTCTCACCAT**A**GCCAG : : : GT : CTGG  
 Mother Reverse .TGTCATTCTCACCAT**A**GCCAG : : : GT : CTGG  
 Father Forward .TGTCATTCTCACCAT**A**GCCAG : : : GT : CTGG  
 Father Reverse .TGTCATTCTCACCAT**A**GCCAG : : : GT : CTGG  
 PID\_012 Forward .TGTCATTCTCAC**C**AT**R**GCCAG : : : GT : CTGG  
 PID\_012 Reverse .TGTCATTCTCACCAT**R**GCCAG : : : GT : CTGG



**(D) *MAP3K14* (G1033A)**

Wild-type	TGCTCTGCAAGGCAGC <b>G</b> TGAGCTCAGGCCAGG
Mother Forward	TGCTCTGCAAGGCAGC <b>R</b> TGAGCTCAGGCCAGG
Mother Reverse	TGCT <b>C</b> TGCAAGGCAGC <b>R</b> TGAGCTCAGGCCAGG
Father Forward	TGCTCTGCAAGGCAGC <b>R</b> TGAGCTCAGGCCAGG
Father Reverse	TGCT <b>C</b> TGCAAGGCAGC <b>R</b> TGAGCTCAGGCCAGG
PID_040 Forward	TGCTCTGCAAGGCAGC <b>A</b> TGAGCTCAGGCCAGG
PID_040 Reverse	TGCTCTGCAAGGCAGC <b>A</b> TGAGCTCAGGCCAGG

**(E) *TAPI* (P67S)**

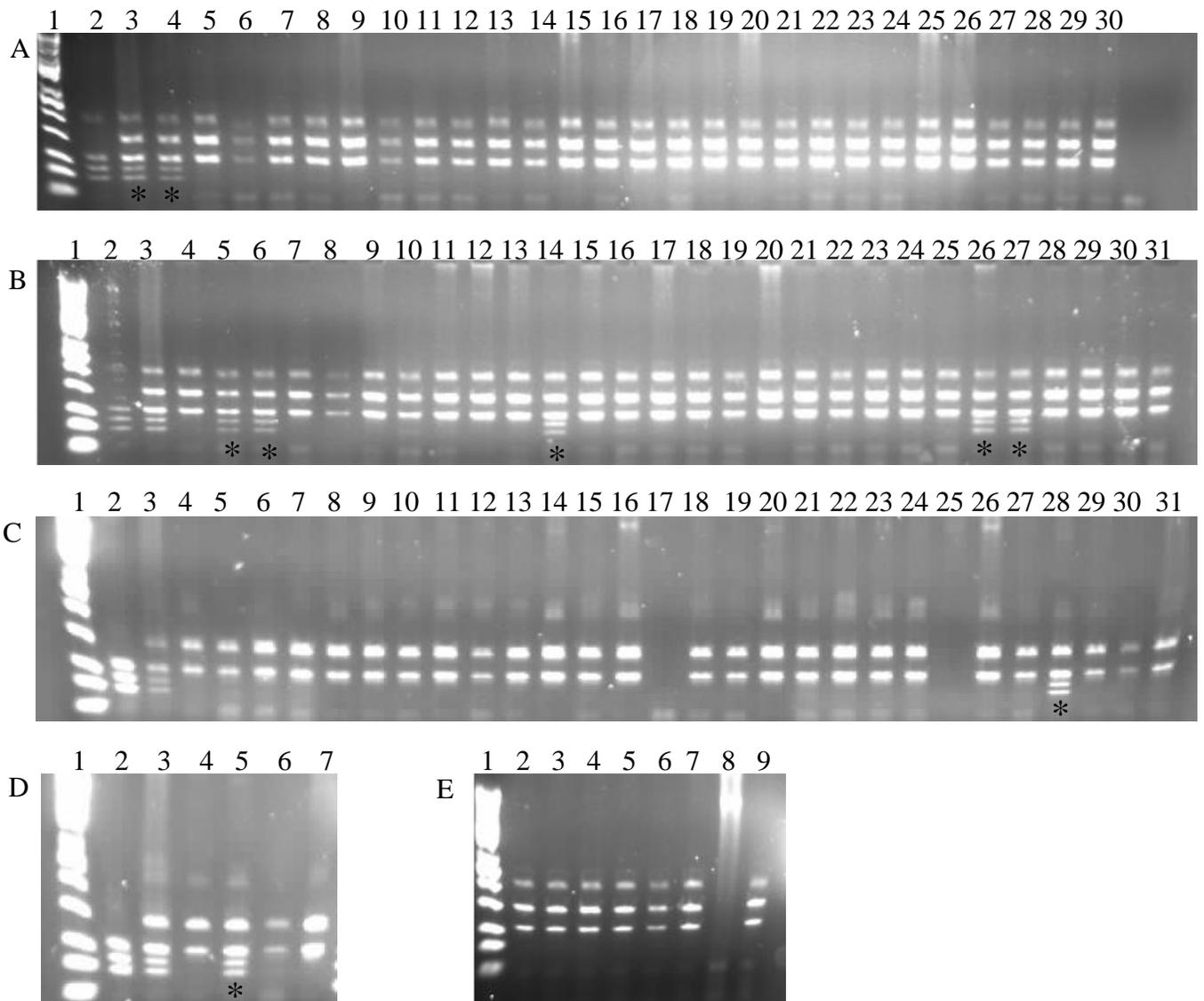
Wild-type	GGCTAGCTCTAGGTGT <b>C</b> CCGCTCCCCGCGGGT
Mother Forward	GGCTAGCTCTAG <b>G</b> TGT <b>Y</b> CC <b>S</b> CTCCCCG <b>S</b> GGGT
Mother Reverse	GGCTAGCTCTAGGTGT <b>Y</b> CCGCTCCCCGCGGGT
Father Forward	GGCTAGCTCTAGGTGT <b>Y</b> CC <b>S</b> CTCCCCG <b>S</b> GGGT
Father Reverse	GGCTAGCT <b>C</b> TAGGTGT <b>Y</b> CCGCTCCCCGCGGGT
PID_060 Forward	GGCTAGCTCTAGGTGT <b>T</b> CC <b>S</b> CTCCCCG <b>S</b> GGGT
PID_060 Reverse	GGCTAGCT <b>C</b> TAGGTGT <b>T</b> CCGCTCCCCGCGGGT



**Supplementary Figure 1: Validation of candidate variants through Sanger sequencing.** The location of the SNP of interest in each case is indicated by a red arrow. The wild-type is the reference sample obtained from NCBI. The sequences of the healthy parents are indicated where available.



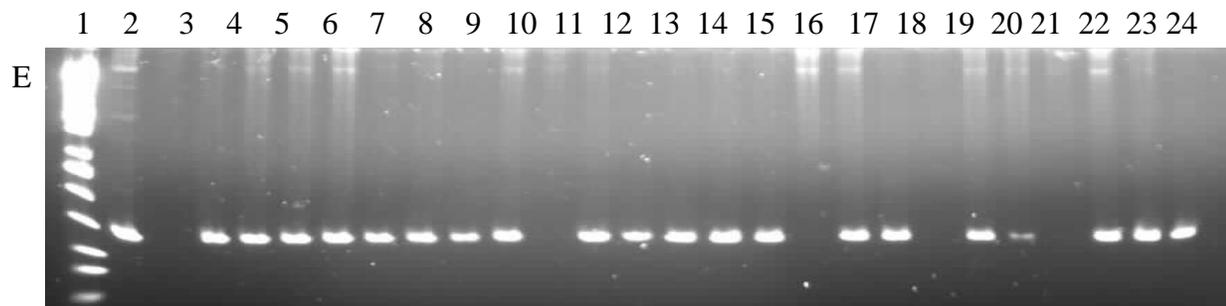
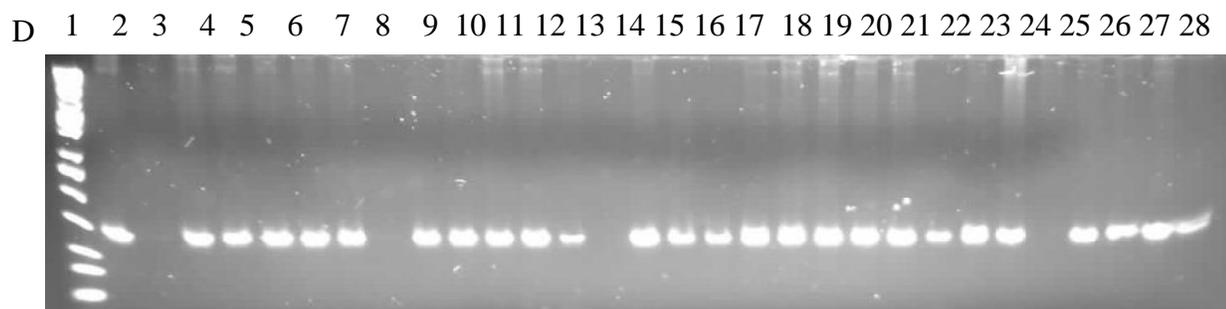
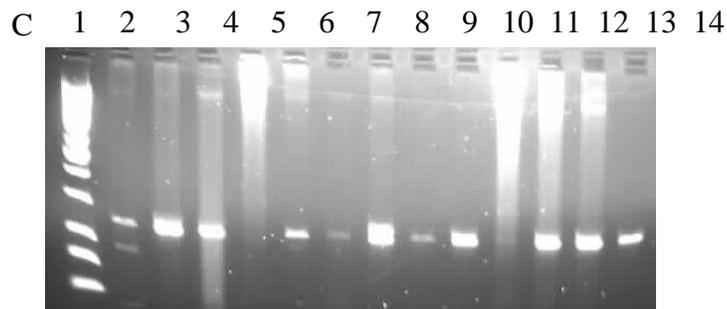
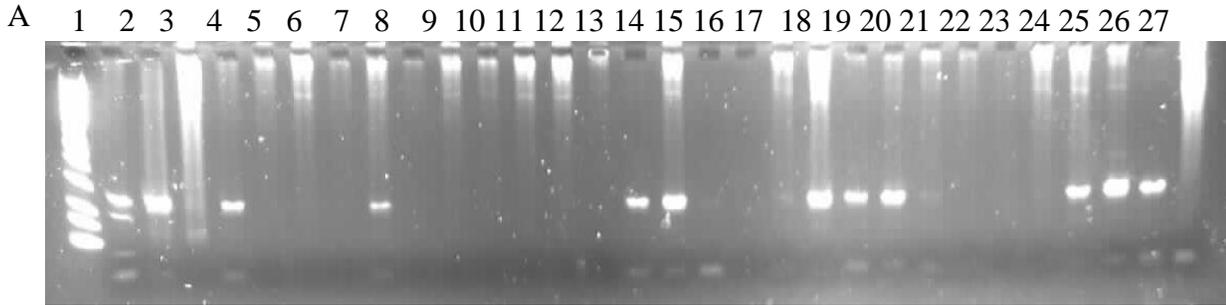
**APPENDIX X – Supplementary Figure 2: Enzyme digest results for *TSPAN33 A111T* identified in PID\_012.**



(A) Lane 1: 100bp DNA ladder; Lane 2: PID\_012; Lane 3: PID\_012's healthy mother; Lanes 4-30: ethnically matched control individuals. (B) Lane 1: 100bp DNA ladder; Lane 2: PID\_012; Lanes 3: PID\_012's healthy mother; Lanes 4-31: ethnically matched control individuals. (C) Lane 1: 100bp DNA ladder; Lane 2: PID\_012; Lane 3: PID\_012's healthy mother; Lanes 4-31: ethnically matched control individuals. (D) Lane 1: 100bp DNA ladder; Lane 2: PID\_012; Lane 3: PID\_012's healthy mother; Lanes 4-7: ethnically matched control individuals. (E) Lane 1: 100bp DNA ladder; Lane 2-9: ethnically matched control individuals.

The mother and eight other control individuals (Lanes A 4; B 5, 6, 14, 26, 27; C 28 and D 5) are heterozygous for the *TSPAN33 A111T* variant, while all other controls are homozygous wild-type. In total, 84 of the 94 controls were homozygous wild-type and 9 were heterozygous for this variant. \* = heterozygous controls.

APPENDIX XI – **Supplementary Figure 3: Enzyme digest results for *TAP1 I296M* identified in PID 012.**



(A) Lane 1: 100bp DNA ladder; Lane 2: PID\_012; Lane 3-30: ethnically matched control individuals. (B) Lane 1: 100bp DNA ladder; Lane 2: PID\_012; Lanes 3-31: ethnically matched control individuals. (C) Lane 1: 100bp DNA ladder; Lane 2: PID\_012; Lane 3-14: ethnically matched control individuals. (D) Lane 1: 100bp DNA ladder; Lane 2-31: ethnically matched control individuals. (E) Lane 1: 100bp DNA ladder; Lane 2: PID\_012's healthy mother; Lanes 3-27: ethnically matched control individuals. The mother and all other control individuals, 92 in total, are homozygous wild-type for the *TAP1 I296M* variant. Only PID\_012 is heterozygous for the variant.

**APPENDIX XII – Case-control association study results.****Supplementary Table 3: All SNPs in *TAP1* and *MAP3K14* investigated for PTB association.**

SNP	Rs ID	Minor	Major	AFF	AFF MAF	UNAFF	UNAFF MAF	Uncorrected	BONF	OR	SE	95% CI	Power
6:32813214	rs3198005	T	C	0/10/372	0.0131	0/8/381	0.0103	0.5161	1	1.4060	0.5246	0.5028-3.931	0.199
6:32813279	rs1057373	T	G	3/43/336	0.0641	3/45/341	0.0656	0.4556	1	0.8497	0.2183	0.554-1.303	0.637
6:32813768	rs116827595	A	G	0/19/363	0.0249	1/20/368	0.0283	0.9948	1	0.9977	0.3463	0.5061-1.967	0.326
6:32814902	rs41551515	T	C	0/24/358	0.0314	0/19/370	0.0244	0.3629	1	1.3640	0.3411	0.6989-2.662	0.400
6:32816216	rs77287921	C	T	2/21/359	0.0327	1/23/365	0.0321	0.9039	1	1.0380	0.3097	0.5658-1.905	0.412
6:32816443	rs41561219	T	C	0/25/357	0.0327	0/20/369	0.0257	0.3174	1	1.3950	0.3333	0.7261-2.682	0.412
6:32816889	rs2228110	A	C	2/21/359	0.0327	1/23/365	0.0321	0.9039	1	1.0380	0.3097	0.5658-1.905	0.412
6:32817774	rs2395269	G	T	6/90/283	0.1346	9/98/282	0.1491	0.7277	1	0.9458	0.1599	0.6913-1.294	0.904
6:32817797	rs2395270	T	G	0/27/355	0.0353	0/23/366	0.0296	0.3286	1	1.3670	0.3199	0.7302-2.559	0.438
6:32818010	rs241423	T	C	1/13/368	0.0196	0/10/379	0.0129	0.4364	1	1.4200	0.4507	0.5871-3.436	0.274
6:32818031	rs2127680	C	T	0/27/355	0.0353	0/23/366	0.0296	0.3286	1	1.3670	0.3199	0.7302-2.559	0.438
6:32818236	rs2127679	A	G	0/25/357	0.0327	0/20/369	0.0257	0.3174	1	1.3950	0.3333	0.7261-2.682	0.412
6:32818432	rs4148881	T	G	0/15/367	0.0196	0/14/375	0.0180	0.9748	1	1.0130	0.4098	0.4538-2.262	0.274
6:32818678	rs2071538	T	C	0/39/343	0.0511	4/43/342	0.0656	0.7873	1	0.9369	0.2416	0.5835-1.504	0.573
6:32819478	rs4148879	T	C	5/68/309	0.1021	7/77/305	0.1170	0.6279	1	0.9188	0.1747	0.6525-1.294	0.830
6:32819865	rs4148879	G	A	5/68/309	0.1021	7/78/304	0.1183	0.6161	1	0.9162	0.1747	0.6506-1.29	0.830
6:32819968	rs2071481	A	G	0/27/355	0.0353	0/23/366	0.0296	0.3286	1	1.3670	0.3199	0.7302-2.559	0.438
6:32820991	rs55702652	G	A	0/27/355	0.0353	0/23/366	0.0296	0.3286	1	1.3670	0.3199	0.7302-2.559	0.438
6:32821365	rs57640466	G	C	0/16/366	0.0209	0/14/375	0.0180	0.6862	1	1.1840	0.4178	0.522-2.685	0.289
17:43340936	rs386607283	C	T	89/184/109	0.4738	81/201/107	0.4666	0.6564	1	1.0510	0.1119	0.844-1.309	0.993
17:43342141	rs1047833	G	C	17/127/235	0.2124	15/132/242	0.2082	0.4306	1	1.1150	0.1381	0.8506-1.461	0.970
17:43358418	rs16939948	C	T	0/22/360	0.0288	0/21/368	0.0270	0.4520	1	1.3000	0.3491	0.656-2.578	0.373
17:43373400	rs28735460	T	C	0/28/354	0.0367	0/29/360	0.0373	0.7777	1	0.9167	0.3081	0.5012-1.677	0.450
17:43378213	rs66512538	T	A	54/202/126	0.4058	75/168/146	0.4087	0.4194	1	1.0970	0.1145	0.8764-1.373	0.993
17:43385349	rs12449740	G	A	75/202/105	0.4607	91/177/121	0.4614	0.5752	1	1.0650	0.1121	0.8548-1.326	0.993

17:43388413	rs17686001	A	G	2/38/342	0.0550	2/50/337	0.0694	0.7701	1	0.9346	0.2314	0.5938-1.471	0.602
17:43364070	rs113278485	G	T	0/14/368	0.0183	0/11/378	0.0141	0.8620	1	1.0800	0.4433	0.453-2.575	0.260
17:43364737	rs17846855	A	C	0/27/355	0.0353	0/33/356	0.0424	0.9993	1	0.9998	0.2953	0.5604-1.784	0.438
17:43364914	rs2074289	C	T	10/71/301	0.1191	1/79/309	0.1041	0.8099	1	1.0430	0.1743	0.7411-1.467	0.874
17:43367653	rs7222094	T	C	41/183/158	0.3469	59/171/159	0.3715	0.8404	1	0.9767	0.1170	0.7766-1.228	0.991
17:43371295	rs73303268	T	G	4/76/302	0.1099	5/63/321	0.0938	0.8050	1	0.9554	0.1849	0.6649-1.373	0.852
17:43391328	rs4792848	T	G	1/80/301	0.1073	3/70/316	0.0977	0.8227	1	1.0440	0.1901	0.719-1.515	0.845

Chromosome 6 is where TAP1 is situated and chromosome 17 is the location of MAP3K14. The AFF and UNAFF columns give three values. 1st = amount of homozygous minor allele genotypes; 2nd = amount of heterozygous genotypes; 3rd = amount of homozygous major allele genotypes observed in each of the cases (AFF) and controls (UNAFF). SNP = single nucleotide polymorphism; AFF = affected PTB cases; UNAFF = unaffected controls; BONF = Bonferroni corrected p-value; OR = odds ratio; SE = standard error; 95% CI = 95% confidence interval. GAS Power Calculator was used to determine the power scores.

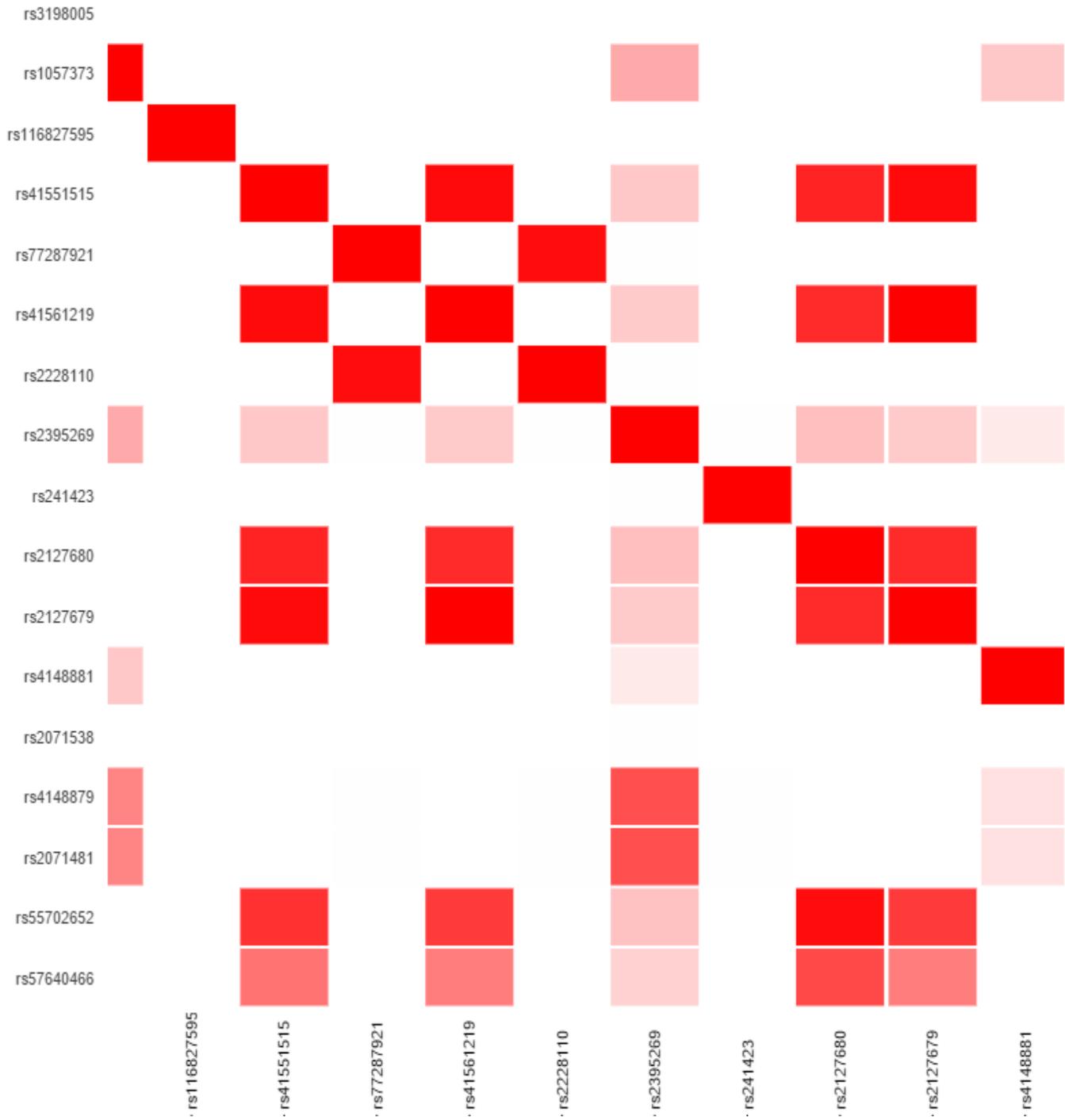
**Supplementary Table 4: R<sup>2</sup> values for SNPs in *TAP1*.**

	rs3198005	rs1057373	rs116827595	rs41551515	rs77287921	rs41561219	rs2228110	rs2395269	rs241423
rs3198005	1.0	0.0	0.0	0.0	0.0	0.0	0.001	0.001	0.00
rs1057373	0.0	1.0	0.001	0.002	0.003	0.002	0.004	0.334	0.003
rs116827595	0.0	0.001	1.0	0.001	0.001	0.001	0.002	0.003	0.001
rs41551515	0.0	0.002	0.001	1.0	0.002	0.961	0.002	0.219	0.002
rs77287921	0.0	0.003	0.001	0.002	1.0	0.002	0.949	0.009	0.003
rs41561219	0.0	0.002	0.001	0.961	0.002	1.0	0.002	0.211	0.002
rs2228110	0.001	0.004	0.002	0.002	0.949	0.002	1.0	0.009	0.004
rs2395269	0.001	0.334	0.003	0.219	0.009	0.211	0.009	1.0	0.008
rs241423	0.0	0.003	0.001	0.002	0.003	0.002	0.004	0.008	1.0
rs2127680	0.0	0.002	0.001	0.865	0.002	0.831	0.002	0.253	0.002
rs2127679	0.0	0.002	0.001	0.961	0.002	1.0	0.002	0.211	0.002
rs4148881	0.0	0.219	0.0	0.0	0.001	0.0	0.001	0.085	0.001
rs2071538	0.209	0.002	0.001	0.001	0.002	0.001	0.003	0.006	0.002
rs4148879	0.001	0.485	0.002	0.003	0.006	0.003	0.006	0.692	0.006
rs2071481	0.001	0.485	0.002	0.003	0.006	0.003	0.006	0.692	0.006
rs55702652	0.0	0.002	0.001	0.807	0.002	0.772	0.002	0.241	0.002
rs57640466	0.0	0.002	0.001	0.551	0.002	0.511	0.002	0.18	0.002
	rs2127680	rs2127679	rs4148881	rs2071538	rs4148879	rs2071481	rs55702652	rs57640466	
rs3198005	0.0	0.0	0.0	0.209	0.001	0.001	0.0	0.0	
rs1057373	0.002	0.002	0.219	0.002	0.485	0.485	0.002	0.002	
rs116827595	0.001	0.001	0.0	0.001	0.002	0.002	0.001	0.001	
rs41551515	0.865	0.961	0.0	0.001	0.003	0.003	0.807	0.551	
rs77287921	0.002	0.002	0.001	0.002	0.006	0.006	0.002	0.002	
rs41561219	0.831	1.0	0.0	0.001	0.003	0.003	0.772	0.511	
rs2228110	0.002	0.002	0.001	0.003	0.006	0.006	0.002	0.002	
rs2395269	0.253	0.211	0.085	0.006	0.692	0.692	0.241	0.18	
rs241423	0.002	0.002	0.001	0.002	0.006	0.006	0.002	0.002	
rs2127680	1.0	0.831	0.0	0.001	0.004	0.004	0.949	0.714	
rs2127679	0.831	1.0	0.0	0.001	0.003	0.003	0.772	0.511	
rs4148881	0.0	0.0	1.0	0.0	0.123	0.123	0.0	0.0	
rs2071538	0.001	0.001	0.0	1.0	0.004	0.004	0.001	0.001	
rs4148879	0.004	0.003	0.123	0.004	1.0	1.0	0.003	0.003	
rs2071481	0.004	0.003	0.123	0.004	1.0	1.0	0.003	0.003	
rs55702652	0.949	0.772	0.0	0.001	0.003	0.003	1.0	0.752	
rs57640466	0.714	0.511	0.0	0.001	0.003	0.003	0.752	1.0	

**Supplementary Table 5: D' values for SNPs in *TAP1*.**

	<b>rs3198005</b>	<b>rs1057373</b>	<b>rs116827595</b>	<b>rs41551515</b>	<b>rs77287921</b>	<b>rs41561219</b>	<b>rs2228110</b>	<b>rs2395269</b>	<b>rs241423</b>
rs3198005	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs1057373	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.925	1.0
rs116827595	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs41551515	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs77287921	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs41561219	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs2228110	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs2395269	1.0	0.925	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs241423	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs2127680	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs2127679	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs4148881	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs2071538	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs4148879	1.0	0.928	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs2071481	1.0	0.928	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs55702652	1.0	1.0	1.0	0.941	1.0	0.939	1.0	1.0	1.0
rs57640466	1.0	1.0	1.0	0.798	1.0	0.754	1.0	0.975	1.0
	<b>rs2127680</b>	<b>rs2127679</b>	<b>rs4148881</b>	<b>rs2071538</b>	<b>rs4148879</b>	<b>rs2071481</b>	<b>rs55702652</b>	<b>rs57640466</b>	
rs3198005	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
rs1057373	1.0	1.0	1.0	1.0	0.928	0.928	1.0	1.0	
rs116827595	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
rs41551515	1.0	1.0	1.0	1.0	1.0	1.0	0.941	0.798	
rs77287921	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
rs41561219	1.0	1.0	1.0	1.0	1.0	1.0	0.939	0.754	
rs2228110	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
rs2395269	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.975	
rs241423	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
rs2127680	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.977	
rs2127679	1.0	1.0	1.0	1.0	1.0	1.0	0.939	0.754	
rs4148881	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
rs2071538	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
rs4148879	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
rs2071481	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
rs55702652	1.0	0.939	1.0	1.0	1.0	1.0	1.0	0.978	
rs57640466	0.977	0.754	1.0	1.0	1.0	1.0	0.978	1.0	

**Supplementary Figure 4: Linkage disequilibrium plot of known SNPs in *TAPI*.**



**Supplementary Table 6: R<sup>2</sup> values for SNPs in *MAP3K14*.**

	<b>rs708563</b>	<b>rs1047833</b>	<b>rs16939948</b>	<b>rs11327848 5</b>	<b>rs17846855</b>	<b>rs2074289</b>	<b>rs7222094</b>
rs708563	1.0	0.25	0.019	0.002	0.015	0.008	0.012
rs1047833	0.25	1.0	0.007	0.001	0.003	0.002	0.005
rs16939948	0.019	0.007	1.0	0.0	0.586	0.003	0.011
rs113278485	0.002	0.001	0.0	1.0	0.0	0.0	0.001
rs17846855	0.015	0.003	0.586	0.0	1.0	0.004	0.02
rs2074289	0.008	0.002	0.003	0.0	0.004	1.0	0.039
rs7222094	0.012	0.005	0.011	0.001	0.02	0.039	1.0
rs73303268	0.023	0.008	0.002	0.0	0.001	0.288	0.057
rs28735460	0.079	0.118	0.002	0.0	0.004	0.0	0.07
rs144161003	0.001	0.021	0.038	0.001	0.035	0.002	0.229
rs12449740	0.045	0.0	0.022	0.0	0.036	0.001	0.525
rs17686001	0.035	0.011	0.001	0.0	0.001	0.005	0.057
rs4792848	0.029	0.003	0.003	0.0	0.004	0.049	0.065
	<b>rs73303268</b>	<b>rs28735460</b>	<b>rs144161003</b>	<b>rs12449740</b>	<b>rs17686001</b>	<b>rs4792848</b>	
rs708563	0.023	0.079	0.001	0.045	0.035	0.029	
rs1047833	0.008	0.118	0.021	0.0	0.011	0.003	
rs16939948	0.002	0.002	0.038	0.022	0.001	0.003	
rs113278485	0.0	0.0	0.001	0.0	0.0	0.0	
rs17846855	0.001	0.004	0.035	0.036	0.001	0.004	
rs2074289	0.288	0.0	0.002	0.001	0.005	0.049	
rs7222094	0.057	0.07	0.229	0.525	0.057	0.065	
rs73303268	1.0	0.0	0.052	0.03	0.003	0.012	
rs28735460	0.0	1.0	0.041	0.126	0.004	0.014	
rs144161003	0.052	0.041	1.0	0.448	0.014	0.021	
rs12449740	0.03	0.126	0.448	1.0	0.031	0.064	
rs17686001	0.003	0.004	0.014	0.031	1.0	0.004	
rs4792848	0.012	0.014	0.021	0.064	0.004	1.0	

**Supplementary Table 7: D' values for SNPs in *MAP3K14*.**

	rs708563	rs1047833	rs16939948	rs113278485	rs17846855	rs2074289	rs7222094
rs708563	1.0	0.963	1.0	1.0	0.722	0.258	0.134
rs1047833	0.963	1.0	1.0	1.0	0.528	0.191	0.1
rs16939948	1.0	1.0	1.0	1.0	0.943	1.0	0.927
rs113278485	1.0	1.0	1.0	1.0	1.0	1.0	1.0
rs17846855	0.722	0.528	0.943	1.0	1.0	1.0	1.0
rs2074289	0.258	0.191	1.0	1.0	1.0	1.0	0.691
rs7222094	0.134	0.1	0.927	1.0	1.0	0.691	1.0
rs73303268	0.521	0.525	1.0	1.0	0.046	0.634	0.983
rs28735460	0.891	0.564	1.0	1.0	1.0	0.144	1.0
rs144161003	0.059	0.415	0.844	0.533	0.663	0.084	0.968
rs12449740	0.24	0.009	0.948	0.371	1.0	0.077	0.984
rs17686001	0.935	1.0	1.0	1.0	1.0	1.0	1.0
rs4792848	0.454	0.28	1.0	1.0	1.0	0.231	0.93
	rs73303268	rs28735460	rs144161003	rs12449740	rs17686001	rs4792848	
rs708563	0.521	0.891	0.059	0.24	0.935	0.454	
rs1047833	0.525	0.564	0.415	0.009	1.0	0.28	
rs16939948	1.0	1.0	0.844	0.948	1.0	1.0	
rs113278485	1.0	1.0	0.533	0.371	1.0	1.0	
rs17846855	0.046	1.0	0.663	1.0	1.0	1.0	
rs2074289	0.634	0.144	0.084	0.077	1.0	0.231	
rs7222094	0.983	1.0	0.968	0.984	1.0	0.93	
rs73303268	1.0	0.018	0.465	0.528	1.0	1.0	
rs28735460	0.018	1.0	0.941	0.989	1.0	1.0	
rs144161003	0.465	0.941	1.0	0.996	1.0	0.649	
rs12449740	0.528	0.989	0.996	1.0	1.0	0.764	
rs17686001	1.0	1.0	1.0	1.0	1.0	1.0	
rs4792848	1.0	1.0	0.649	0.764	1.0	1.0	

**Supplementary Figure 5: Linkage disequilibrium plot of known SNPs in *MAP3K14*.**

