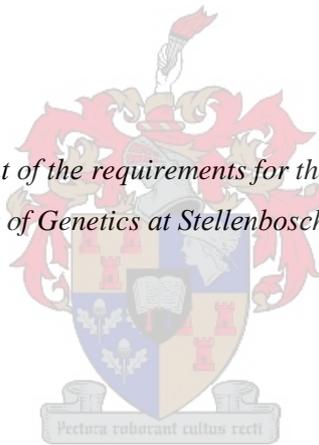


**Mutation analysis of the promoter region of  
*CYBRD1, HFE, LTF, HAMP* and *SLC40A1* in a  
Tuberculosis cohort**

**By: Margarete Sittmann**

*Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in the  
Department of Genetics at Stellenbosch University*



**Supervisor: Prof MG Zaahl**

**April 2014**

## **DECLARATION**

By submitting this thesis electronically, 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:.....

## SUMMARY

Tuberculosis (TB) is an epidemic disease characterised by wet, persistent coughing, weight loss, fever, fatigue, and blood in the sputum. It has been reported that one in every three individuals is currently infected with *Mycobacterium tuberculosis*, the causative agent of TB, and that 10% of them will develop the active disease. The high prevalence and low penetrance of this disease has resulted in increased research performed to ascertain what factors play a role in susceptibility to *M. tuberculosis* infection. Some factors known to play a role in a minority of cases may include: HIV infection, diabetes, alcohol abuse and old age, but racial differences in susceptibility have also been observed. However, the influence of genetic factors is gaining popularity in current research.

*M. tuberculosis* requires iron to proliferate, which it must appropriate from its host. For this reason the genes involved in the metabolism of iron in the human host are of particular interest when considering susceptibility to *M. tuberculosis* infection. In order to determine whether the expression of these genes may influence disease susceptibility, the promoter region of the *CYBRD1*, *HAMP*, *HFE*, *LTF* and *SLC40A1* genes have been targeted for investigation. The aim of this study was to determine whether DNA variation in the promoter region of these genes involved in iron metabolism is associated with *M. tuberculosis* susceptibility.

The study cohort consisted of 49 TB patients and 51 healthy, unrelated, population-matched controls, all of whom were Black, Xhosa-speaking individuals. Initially, 15 patient samples were randomly selected for exploratory screening, utilising semi-automated bi-directional sequencing analysis. In this manner, 40 variants were identified of which 30 were previously described. The novel variants included ***CYBRD1***: -849 C/G, -492 A/G, -454 C/T, -397 A/C; ***HAMP***: -323 C/T; ***HFE***: -561 A/G, -331 A/C; and ***LTF***: -1377 T/G, -457 T/C, -437 C/G. A number of loci demonstrated a statistically significant difference in allele and genotype frequencies, and in iron parameter levels when comparing the patient and control groups and for each variant separately. *In silico* analyses revealed the creation and/or abolishment of several transcription factor binding sites (TFBSs) due to the presence or absence of certain identified variants. The change in the composition of TFBSs in the promoter region may lead to differential expression of the gene.

This study served as a pilot investigation to identify promoter region variants in the candidate genes involved in iron metabolism, in TB patients. The results presented here indicate that the identified variants (-1813 C/T, -1459 T/C, -238 A/G, -166 C/G [*CYBRDI*]; -561 A/G [*HFE*]; -1470 C/T, -1355 G/C and -1098 G/A [*SLC40A1*]) could possibly contribute to the increased absorption of iron in the TB patient group, which could subsequently increase the occurrence of pathogenic infection. The findings of this study could further aid in the elucidation of the exact mechanism(s) by which iron, its metabolism, and the genes involved affect disease susceptibility, more specifically, *M. tuberculosis* susceptibility.

## OPSOMMING

Tuberkulose (TB) is 'n epidemiese siekte gekarakteriseer deur nat, aanhoudende hoes, gewigsverlies, moegheid, en bloed in die speeksel. Dit is gerapporteer dat een in elke drie individue tans geïnfecteer is met *Mycobacterium tuberculosis*, die veroorsakende agent van TB, en dat 10% van dié individue die aktiewe vorm van die siekte sal ontwikkel. Die hoë voorkoms en lae effek van hierdie siekte het daartoe gelei dat meer navorsing mettertyd gedoen is om die faktore wat 'n rol mag speel in *M. tuberculosis* infeksie, te bepaal. Sommige faktore bekend vir hul rol in 'n minderheid van gevalle sluit in: MIV infeksie, diabetes, alkoholmisbruik en bejaardheid, maar etniese verskille in vatbaarheid vir die siekte is ook al waargeneem. Die waarskynlikheid van genetiese invloed op die ontwikkeling van TB word ook meer deur navorsers ondersoek.

*M. tuberculosis* benodig yster om te vermeerder, wat dit moet bekom vanaf die gasheer. Vir hierdie rede is die gene betrokke by yster metabolisme in die menslike gasheer veral van belang vir die oorweging van vatbaarheid vir *M. tuberculosis*. Om te bepaal of die uitdrukking van hierdie gene moontlik 'n invloed het op vatbaarheid vir die siekte, was die promoter areas van die *CYBRD1*, *HAMP*, *HFE*, *LTF* en *SLC40A1* gene geteiken. Die doel van hierdie studie was om te bepaal of DNS variasie in die promoter area van hierdie gene betrokke in yster metabolisme moontlik verband kan hou met vatbaarheid vir *M. tuberculosis*.

Die studie kohort het uit 49 TB pasiënte en 51 gesonde, onverwante, populasie-gepaarde kontroles, waarvan almal Swart, Xhosa-sprekende individue was, bestaan. Aanvanklik was 15 pasiënt monsters lukraak gekies vir ondersoekende sifting, deur die gebruik van semi-outomatiese twee-rigting volgordebepalings. Op hierdie manier is 40 variante geïdentifiseer waarvan 30 voorheen beskryf is. Die nuwe variante sluit in *CYBRD1*: -849 C/G, -492 A/G, -454 C/T, -397 A/C; *HAMP*: -323 C/T; *HFE*: -561 A/G, -331 A/C; en *LTF*: -1377 T/G, -457 T/C, -437 C/G. 'n Aantal loci het statisties betekenisvolle verskille getoon in alleel en genotipe frekwensies, en in yster parameter vlakke met die vergelyking van die pasiënt groep met die kontrole groep. *In silico* analise het verder die skepping en/of afskaffing van verskeie transkripsiefaktor bindingsetels (TFBSs), as gevolg van die teenwoordigheid of afwesigheid van sekere variante, getoon. Die verandering in die samestelling van TFBSs in die promoter area kan lei tot differensiële uitdrukking van die geen.

Dié studie het gedien as 'n voorlopige ondersoek om te bepaal of promoter area variante, geïdentifiseer in kandidaat gene betrokke by yster metabolisme, 'n invloed het in die ontwikkeling van TB. Die resultate wat hier gewys word dui aan dat die geïdentifiseerde variante (-1813 C/T, -1459 T/C, -238 A/G, -166 C/G [*CYBRDI*]; -561 A/G [*HFE*]; -1470 C/T, -1355 G/C and -1098 G/A [*SLC40A1*]) moontlik die verhoogde opname van yster kan veroorsaak, wat later die toename van die patogeniese infeksie kan veroorsaak. Die bevindinge van hierdie studie kan moontlik bydra tot die toeligting van die presiese meganisme(s) waardeur yster, yster metabolisme, en die betrokke gene vatbaarheid vir siekte, meer spesifiek *M. tuberculosis* vatbaarheid, beïnvloed.

# TABLE OF CONTENTS

## Contents

DECLARATION .....	i
SUMMARY .....	ii
OPSOMMING .....	iv
TABLE OF CONTENTS.....	vi
LIST OF ABBREVIATIONS AND SYMBOLS .....	x
LIST OF FIGURES .....	xvii
LIST OF TABLES .....	xx
ACKNOWLEDGEMENTS .....	xxi
Chapter 1 .....	1
1. LITERATURE REVIEW .....	2
1.1 TUBERCULOSIS.....	2
1.1.1 Disease Penetrance and Prevalence .....	3
1.1.2 The Causative Agent – <i>Mycobacterium tuberculosis</i> .....	3
1.1.2.1 Diagnosis.....	4
1.1.2.2 Disease Progression and Treatment .....	6
1.2 IRON AND TUBERCULOSIS – ADAPTATION TO HUMAN INFECTION.....	8
1.3 IRON IN THE HUMAN BODY .....	9
1.3.1 Overview.....	9
1.3.2 Iron Absorption and Storage.....	11
1.3.3 Iron Transport .....	13
1.4 GENES UNDER INVESTIGATION .....	14
1.4.1 Cytochrome b reductase 1 ( <i>CYBRD1</i> ) .....	16
1.4.2 Heparin Antimicrobial Peptide ( <i>HAMP</i> ).....	17

1.4.3 HFE (Haemochromatosis gene).....	18
1.4.4 Ferroportin (SLC40A1) .....	19
1.4.5 Lactotransferrin (LTF) .....	20
1.5 TRANSCRIPTIONAL REGULATION OF GENES INVOLVED IN IRON METABOLISM .....	21
1.5.1 Overview .....	21
1.5.2 Iron Response Elements (IREs) .....	22
1.5.3 Other Regulatory Elements .....	24
1.6 AIM AND OBJECTIVES.....	25
Chapter 2.....	26
2. DETAILED EXPERIMENTAL PROCEDURES .....	27
2.1 COHORT SAMPLING.....	27
2.1.1 Patient and Control Groups.....	27
2.1.2 Iron Parameter Data .....	27
2.1.3 DNA Extraction .....	28
2.2 POLYMERASE CHAIN REACTION (PCR) AMPLIFICATION.....	29
2.2.1 Oligonucleotide Primers .....	29
2.2.2 Amplification of Fragments .....	34
2.2.2.1 Amplification of PCR Fragments .....	34
2.2.2.2 PCR Reaction Reagents and Amplification Cycles .....	34
2.2.3 Agarose Gel Electrophoresis.....	36
2.3 MUTATION DETECTION.....	37
2.3.1 Semi-automated bi-directional DNA sequencing analysis .....	37
2.3.2 Restriction Fragment Length Polymorphism (RFLP) Analysis.....	39
2.3.3 Heteroduplex Single-Strand Conformation Polymorphism (HEX-SSCP) Analysis.....	40

2.5 <i>IN SILICO</i> ANALYSES .....	41
2.6 STATISTICAL ANALYSIS .....	41
2.6.1 Iron parameters .....	41
2.6.2 Genotyping.....	42
2.6.3 TB association.....	43
Chapter 3.....	44
3. RESULTS & DISCUSSION.....	45
3.1 ABSTRACT.....	45
3.2 INTRODUCTION .....	46
3.3 MATERIALS AND METHODS.....	48
3.4 RESULTS .....	48
3.4.1 Mutation Detection .....	48
3.4.1.1 Variant List .....	48
3.4.1.2 HWE Analysis .....	52
3.4.1.3 Genotype and Allele Frequency Analysis.....	53
3.4.2 <i>In Silico</i> Analysis .....	58
3.4.3 Iron Parameter Analysis.....	59
3.4.3.1 Descriptive Statistics.....	59
3.4.3.2 Pearson r Correlation .....	60
3.4.3.3 Normality Distribution and Data Transformation.....	62
3.4.3.4 T-tests (Parametric and Non-parametric).....	63
3.5 DISCUSSION.....	67
3.4.1 <i>CYBRD1</i> .....	71
3.4.2 <i>HFE</i> .....	78
Chapter 4.....	80
4.1 CONCLUDING REMARKS AND FUTURE PROSPECTS.....	81

Chapter 5.....	84
5. REFERENCES .....	85
5.1 GENERAL REFERENCES.....	85
5.2 ELECTRONIC REFERENCES .....	101
Addendum A.....	102
ADDENDUM A .....	103
A.1 PROMOTER REGION SEQUENCES OF <i>CYBRD1</i> , <i>HAMP</i> , <i>HFE</i> , <i>LTF</i> AND <i>SLC40A1</i> .....	103
Addendum B .....	109
ADDENDUM B.....	110
B.1 VARIANT IDENTIFICATION .....	110
B.1.1 Chromatograms .....	110
B.1.2 RFLP Gel Images .....	122
B.1.3 HEX-SSCP Gel Images.....	123
Addendum C .....	125
ADDENDUM C.....	126
C.1 <i>T</i> -TESTS FOR IRON PARAMETER LEVEL COMPARISONS.....	126

**LIST OF ABBREVIATIONS AND SYMBOLS**

<b>3'</b>	3-prime
<b>5'</b>	5-prime
<b><math>\alpha</math></b>	alpha
<b>~</b>	approximately
<b><math>\beta</math></b>	beta
<b>°C</b>	degrees Celsius
<b><math>\gamma</math></b>	gamma
<b>%</b>	Percentage
<b><math>\mu\text{g/ml}</math></b>	microgram per millilitre
<b><math>\mu\text{g/l}</math></b>	microgram per litre
<b><math>\mu\text{l}</math></b>	microlitre
<b><math>\mu\text{mol/l}</math></b>	micromole per litre
<b>A</b>	adenosine
<b>AA</b>	acrylamide
<b>APS</b>	ammonium persulphate
<b>BAA</b>	bisacrylamide
<b>BCE</b>	before current era
<b>BCG</b>	bacillus Calmette-Guerin
<b>bp</b>	base-pair
<b>BR</b>	broad distribution promoter
<b>BRE</b>	broad distribution element
<b>BSA</b>	bovine serum albumin

<b>C</b>	cytosine
<b>CACCC-bi</b>	CACCC-binding
<b>C/EBP</b>	CACAAT/enhancer binding protein
<b>C<sub>4</sub>H<sub>11</sub>NO<sub>3</sub></b>	tris-acetate
<b>CLB</b>	cell lysis buffer
<b>c-Myb</b>	cellular myeloblastosis transcription factor
<b>c-Myc</b>	cellular myelocytomatosis transcription factor
<b><i>CYBRD1</i></b>	cytochrome b reductase 1 gene
<b>dATP</b>	deoxyadenosine triphosphate
<b>dbSNP</b>	SNP database
<b>dCTP</b>	deoxycytidine triphosphate
<b>DCYTB</b>	duodenal cytochrome B protein
<b>del</b>	deletion
<b>dGTP</b>	deoxyguanosine triphosphate
<b>dH<sub>2</sub>O</b>	distilled water
<b>DMT1</b>	divalent metal transporter 1
<b>DNA</b>	deoxyribonucleic acid
<b>dNTP</b>	deoxyribonucleoside triphosphate
<b>DPE</b>	downstream promoter element
<b>dTTP</b>	deoxythymidine triphosphate
<b>E1</b>	enhancer 1
<b>E2F-1</b>	E2 promoter binding factor 1
<b><i>E. coli</i></b>	Escherichia coli
<b>EDTA</b>	ethylenediaminetetraacetic acid
<b>EtBr</b>	ethidium bromide

<b>ETF</b>	embryonic TEA domain-containing factor
<b>EtOH</b>	ethanol
<b>F</b>	forward primer
<b>Fe</b>	iron
<b>Fe<sup>2+</sup></b>	ferrous iron
<b>Fe<sup>3+</sup></b>	ferric iron
<b>FOG</b>	friend of GATA
<b>FPN</b>	ferroportin protein
<b>g</b>	gram
<b>G</b>	guanosine
<b>g/l</b>	gram per litre
<b><i>HAMP</i></b>	hepcidin antimicrobial peptide gene
<b>Hb</b>	hunchback
<b>H<sub>3</sub>BO<sub>3</sub></b>	boric acid
<b>HCl</b>	hydrogen chloride
<b>HCP1</b>	haem carrier protein 1
<b>HEP</b>	hepcidin protein
<b>HEPH</b>	hephaestin protein
<b>HEX-SSCP</b>	heteroduplex single-strand conformation polymorphism
<b><i>HFE</i></b>	haemochromatosis gene
<b>HFE</b>	haemochromatosis protein
<b>HH</b>	hereditary haemochromatosis
<b>HIV</b>	human immunodeficiency virus
<b>HNF</b>	hepatocyte nuclear factor
<b>HWD</b>	Hardy-Weinberg disequilibrium

<b>HWE</b>	Hardy-Weinberg equilibrium
<b>IGRA</b>	interferon- $\gamma$ (gamma) release assay
<b>Inc.</b>	<b>incorporated</b>
<b>ins</b>	insertion
<b>IRP1</b>	iron regulatory protein 1
<b>IRP2</b>	iron regulatory protein 2
<b>IRE</b>	iron response element
<b>kb</b>	kilobase
<b>KCH<sub>3</sub>CO<sub>2</sub></b>	potassium acetate
<b>KCl</b>	potassium chloride
<b>KHCO<sub>3</sub></b>	potassium hydrogen carbonate
<b>KH<sub>2</sub>PO<sub>4</sub></b>	potassium phosphate
<b>Knox-20</b>	knotted-like homeobox 20
<b>LD</b>	linkage disequilibrium
<b>LEF-1</b>	lymphoid enhancer-binding factor 1
<b>LTF</b>	lactotransferrin gene
<b>LTF</b>	lactotransferrin protein
<b>M</b>	moles per litre
<b>Max</b>	maximum
<b>MEF-2A</b>	myocyte enhancer factor 2A
<b>mg</b>	milligram
<b>Mg<sup>2+</sup></b>	magnesium
<b>Mg(CH<sub>3</sub>CO<sub>2</sub>)<sub>2</sub></b>	magnesium acetate
<b>MgCl<sub>2</sub></b>	magnesium chloride
<b>mg/ml</b>	milligrams per millilitre

<b>MgSO<sub>4</sub></b>	magnesium sulphate
<b>MHC</b>	major histocompatibility complex
<b>Min</b>	minimum
<b>ml</b>	millilitre
<b>mM</b>	millimoles per litre
<b>mRNA</b>	messenger RNA
<b><i>M. tuberculosis</i></b>	<i>Mycobacterium tuberculosis</i>
<b><i>n</i></b>	number of individuals
<b>NADPH</b>	nicotinamide adenine dinucleotide phosphate
<b>Na<sub>2</sub>HPO<sub>4</sub></b>	disodium phosphate
<b>NaCl</b>	sodium chloride
<b>NF1</b>	nuclear factor 1
<b>NFI/CTF</b>	nuclear factor 1/ CCAAT-binding transcription factor
<b>NH<sub>4</sub>CL</b>	ammonium chloride
<b>NHLS</b>	national health laboratory service
<b>(NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub></b>	ammonium sulphate
<b>ng</b>	nanogram
<b>ng/μl</b>	nanogram per microlitre
<b>NLB</b>	nuclear lysis buffer
<b>NRAMP2</b>	natural resistance-associated macrophage protein 2
<b>ns</b>	not significant
<b>p</b>	short arm of chromosome
<b><i>P</i></b>	probability
<b>PAA</b>	polyacrylamide
<b>Pax-5</b>	paired box 5

<b>PCR</b>	polymerase chain reaction
<b>pmol</b>	picomole
<b>PBS</b>	phosphate buffered saline
<b>PR</b>	progesterone receptor
<b>q</b>	long arm of chromosome
<b>R</b>	reverse primer
<b>RE</b>	restriction endonuclease
<b>RFLP</b>	restriction fragment length polymorphism
<b>RNA</b>	ribonucleic acid
<b>ROI</b>	reactive oxygen intermediates
<b>ROS</b>	reactive oxygen species
<b>RNI</b>	reactive nitrogen intermediates
<b>rpm</b>	revolutions per minute
<b>rs</b>	reference sequence
<b>RXR-alpha</b>	retinoid X-receptor alpha
<b>S.D.</b>	standard deviation
<b>SDS</b>	sodium dodecyl sulphate
<b><i>SLC40A1</i></b>	solute carrier family 40 member 1 gene
<b>SNP</b>	single nucleotide polymorphism
<b>SP1</b>	specificity protein 1
<b>T</b>	thymidine
<b>T<sub>A</sub></b>	annealing temperature
<b>TB</b>	tuberculosis
<b>TBE</b>	tris-borate/EDTA
<b>TEMED</b>	N, N, N' N',-tetramethylethylenediamine

<b><i>TF</i></b>	transferrin gene
<b>TF</b>	transferrin protein
<b>TFII-I</b>	transcription factor II-I
<b>TFBS</b>	transcription factor binding site
<b>TFPGA</b>	tools for population genetics analysis
<b>TFR</b>	transferrin receptor protein
<b>T<sub>M</sub></b>	melting temperature
<b>Tris-HCL</b>	tris hydrochloride
<b>UTR</b>	untranslated region
<b>UV</b>	ultraviolet
<b>viz.</b>	visually
<b>v/v</b>	volume per volume
<b>WHO</b>	World Health Organisation
<b>w/v</b>	weight per volume
<b>XBP-1</b>	x-box binding protein 1

---

## LIST OF FIGURES

<b>Figure 1.1</b> The most common testing methods for the diagnosis of <i>M. tuberculosis</i> infection. .....	5
<b>Figure 1.2</b> Standard and drug-resistant TB treatment regimens. ....	7
<b>Figure 1.3</b> Distribution of iron in the human body.. ....	10
<b>Figure 1.4</b> The absorption, storage and transport of iron.....	12
<b>Figure 1.5</b> The interaction of IRE and IRP in mRNA regulation .....	23
<b>Figure 2.1</b> Work flow for the identification and subsequent genotyping of variants .....	38
<b>Figure 3.1</b> Graph depicting the correlation between serum iron and transferrin saturation levels for the entire cohort .....	60
<b>Figure 3.2</b> Graph depicting the correlation between serum iron and transferrin saturation levels in the patient group .....	61
<b>Figure 3.3</b> Graph depicting the correlation between serum iron and transferrin saturation levels in the control group.....	61
<b>Figure 3.4</b> Graph depicting the correlation between serum ferritin and serum transferrin in the patient group .....	62
<b>Figure 3.5</b> Comparison of iron parameters between the patient and control groups .....	63

---

<b>Figure 3.6</b> Comparison of iron parameters to determine the effect of the variant -1470 C/T in the promoter region of <i>SLC40A1</i> .....	64
<b>Figure 3.7</b> Comparison of iron parameters to determine the effect of the variant -1355 G/C in the promoter region of <i>SLC40A1</i> .....	65
<b>Figure 3.8</b> Comparison of iron parameters to determine the effect of the variant -1098 G/A in the promoter region of <i>SLC40A1</i> .....	66
<b>Figure A.1</b> The promoter region investigated upstream of the <i>CYBRD1</i> gene.....	104
<b>Figure A.2</b> The promoter region investigated upstream of the <i>HAMP</i> gene .....	105
<b>Figure A.3</b> The promoter region investigated upstream of the <i>HFE</i> gene.....	106
<b>Figure A.4</b> The promoter region investigated upstream of the <i>LTF</i> gene.....	107
<b>Figure A.5</b> The promoter region investigated upstream of the <i>SLC40A1</i> gene .....	108
<b>Figure C.1</b> Comparison of iron parameters to determine the effect of the variant -1631 C/T in the promoter region of <i>HAMP</i> .....	126
<b>Figure C.2</b> Comparison of iron parameters to determine the effect of the variant -582 A/G in the promoter region of <i>HAMP</i> .....	127
<b>Figure C.3</b> Comparison of iron parameters to determine the effect of the variant -1168 A/G in the promoter region of <i>HFE</i> .....	128
<b>Figure C.4</b> Comparison of iron parameters to determine the effect of the variant -561 A/G in the promoter region of <i>HFE</i> .....	129

**Figure C.5** Comparison of iron parameters to determine the effect of the variant -467 G/C in the promoter region of *HFE*..... 129

**Figure C.6** Comparison of iron parameters to determine the effect of the variant -1461 T/C in the promoter region of *SLC40A1* ..... 130

**Figure C.7** Comparison of iron parameters to determine the effect of the variant -1399 G/A in the promoter region of *SLC40A1* ..... 131

## LIST OF TABLES

<b>Table 1.1</b> General information for genes under investigation.....	15
<b>Table 2.1</b> Oligonucleotide primers and PCR conditions for amplification of DNA for semi-automated bi-directional sequencing .....	30
<b>Table 2.2</b> Oligonucleotide primers and PCR conditions for amplification of genotyping fragments.....	31
<b>Table 2.3</b> Oligonucleotide primers and PCR conditions for amplification of genotyping fragments.....	32
<b>Table 3.1</b> General information regarding variants, including HWE analysis .....	50
<b>Table 3.2</b> Genotype and allele frequencies for variants identified.....	54
<b>Table 3.3</b> Putative TFBSs predicted to be created or abolished by the presence of the variants .....	58
<b>Table 3.4</b> Descriptive statistics for iron parameter levels compared between patient and control groups .....	59
<b>Table 3.5</b> Default settings for the Exact test used in TFPGA and Genepop.....	67

## ACKNOWLEDGEMENTS

I would like to thank the NRF for funding of this project, and Stellenbosch University for the use of its facilities and support in my studies.

*Prof MG Zaahl*, my supervisor, for her continued assistance throughout the project and whose knowledge, vision and insight made this project a success.

*Natalie Strickland, Danielle da Silva and Michelle-Angelique Hallendorff* for their assistance with regards to procedures used during the project and critical reading of my thesis.

*Danielle Cowie and Megan Anne Coomer* for critical reading, discussions and their friendship.

My parents, for their support and understanding.

Finally, and maybe most importantly, *David Landman*, for his support, love, friendship, patience and copy-editing skills.

# Chapter 1

## LITERATURE REVIEW

---

# 1. LITERATURE REVIEW

## 1.1 TUBERCULOSIS

Tuberculosis (TB) is an infectious disease affecting, and taking, the lives of many people worldwide. TB can affect almost all parts of the body but the most common site of infection is the lungs. The general symptoms include fever, night sweats, a wet and persistent cough occasionally accompanied by blood in the sputum, weight-loss, and fatigue (Perez-Osorio *et al.*, 2012).

It has been postulated that this disease has existed for approximately 15 000 to 20 000 years but the oldest evidence of TB infection in humans has been found in bone fossils which date back to 8 000 BCE (Herzog, 1998). Further research suggests that 25 of the world's major diseases caused by pathogens, of which TB is one, originated in what is known as the Old World (Africa, Asia and Europe) (Wolfe *et al.*, 2007). The exact date of the first human infection with TB is unclear and molecular methods to genetically determine this are inadequate as they only take into account the genetic information for the current strains of the infectious pathogen (Smith *et al.*, 2009). These methods do not take into account strains that no longer exist, from which current strains may have originated; the genetic information from the original ancestor is important as it could provide resolution as to the changes to the genome which facilitated human infection. The discovery of the precise origin of this pathogen-derived disease, both geographical and temporal, and the determination of its lineage, could provide insight in terms of control, treatment and eradication.

TB is mainly found in temperate regions and is considered to be a “crowd epidemic disease” which is defined as “occurring locally as a brief epidemic and capable of persisting regionally only in large human populations” (Wolfe *et al.*, 2007). In present times, a decreased number of infections is observed in the Caucasian population group when compared to other ethnic groups, such as the Sub-Saharan Africans. This could be explained by prolonged exposure to TB during times of industrialization, from the seventeenth century onwards, when regions such as Europe and North America experienced population booms and TB epidemics were common (Lipsitch & Sousa, 2002). The prolonged exposure of these Caucasians to the disease may have exerted a selective pressure on this ethnic group to select for resistant individuals. This is supported by Fares (2012) in a review article which reports that Black individuals have a higher prevalence of TB infection when compared to Caucasians.

---

The basis for this racial difference in susceptibility to TB is as yet unclear but socio-economic, immunologic, environmental and genetic factors have been implicated (Fares, 2012).

The following sub-sections will expound on the disease and will present information regarding the prevalence, aetiology, diagnosis, and treatment thereof.

### **1.1.1 Disease Penetrance and Prevalence**

The African continent was ranked as the second highest contributor to the world's TB burden in 2010, contributing 26% of TB cases worldwide (WHO, 2011). For the year 2012, the World Health Organisation (WHO) reported that South Africa had approximately 390 000 reported cases of active TB and close to 25 000 patients succumbed to the disease (WHO, 2012). According to this report, South Africa also has the highest number of TB cases co-infected with HIV when compared to other countries worldwide. The high prevalence of TB cases in this country is surprising when one considers the low penetrance of this disease: it has been estimated that at least one in every three individuals in the world is infected with the bacterium that causes TB but that only 5 to 10% of these cases will present with active TB (Dye *et al.*, 1999; Herzog, 1998). The high prevalence of this low penetrating disease could be attributed to a number of factors influencing susceptibility. Some of these factors including environment, socio-economic status and immunology, could explain why South Africa contributes such a large proportion to the world's TB burden but genetics has specifically been implicated as playing a role in susceptibility to TB among Africans (Bellamy *et al.*, 2000). There is much support for a genetic influence on susceptibility to TB: one of the best examples for this support comes from twin studies (Lipsitch & Sousa, 2002). Monozygotic twins have been shown to have a higher level of concordance in TB morbidity than dizygotic twins, and this concordance decreased as the blood relationship distance increased (Comstock, 1978).

### **1.1.2 The Causative Agent – *Mycobacterium tuberculosis***

*Mycobacterium tuberculosis* (*M. tuberculosis*) is the aetiological agent of most forms of TB. Since its discovery in 1882 by Robert Koch, much research has been done on the bacterium

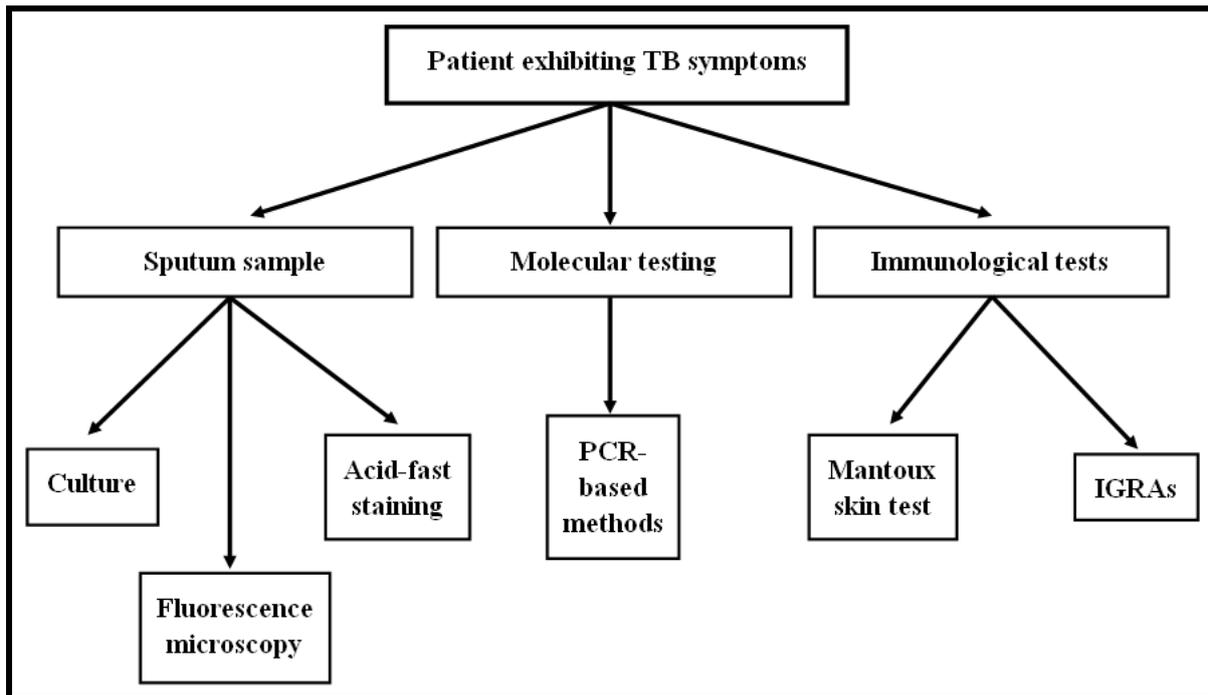
---

and the causative disease. The full genome sequence was determined in 1998 and was shown to be 4 411 529 base pairs (bp) in length containing approximately 4 000 genes (Cole *et al.*, 1998; Camus *et al.*, 2002). This bacterium is a rod-shaped, gram-positive, acid-fast obligate aerobe which is non-motile. The bacterium's genome is GC-rich and it has a dividing time of approximately 24 hours which sets it apart from most other pathogenic bacteria which can divide in less than one hour (Cole *et al.*, 1998). The cell wall of *M. tuberculosis* contains high levels of lipids and mycolic acid, which only weakly retains crystal violet dye, if at all; thus, from an empirical stand-point, it is not technically gram-positive. It is classified as such because, like other gram-positive bacteria, it does not have an outer cell membrane (Gregersen, 1978). The bacterium can withstand dilute disinfectants and has the ability to survive in a dry state for several weeks, which is most likely due to its unique cell wall contents (Madison, 2001).

The bacterium most commonly infects the lungs; consequently transmission from an infected individual to a healthy, uninfected individual relies on the aerosolization of the bacteria. (Dillon *et al.*, 2000; Nicas *et al.*, 2005). It is important to note that only people with active TB, who have developed symptoms due to the bacteria escaping the confinement of a granuloma, are contagious, whereas passive TB patients, who have not developed any active symptoms, are considered unable to transmit the bacteria to other healthy individuals (Kumar *et al.*, 2007).

### ***1.1.2.1 Diagnosis***

There are many methods for the diagnosis of TB but the method used most often relies on the socio-economic status of the country where tested (Metcalf *et al.*, 2011). Figure 1.1 illustrates the most common methods of testing for *M. tuberculosis* infection.



**Figure 1.1** The most common testing methods for the diagnosis of *M. tuberculosis* infection. Testing begins when a patient displays TB symptoms following which a sputum or biopsy sample may be collected for testing. The most popular tests used include sputum culturing, acid-fast staining and PCR-based methods.

If *M. tuberculosis* infection is suspected (the patient displays all the symptoms and is at high risk for infection) then a sputum sample or tissue biopsy may be taken for analysis. Fluorescence microscopy may be performed on the sample in order to determine the morphology of any infecting bacteria present using an acid-fast fluorochrome dye, such as auramine-O, to determine whether it is a bacillus infection. Fluorescence microscopy is an expensive method to use for testing and often in low- to middle-income countries conventional light microscopy using the Ziehl-Neelsen acid-fast stain is used. However, conventional light microscopy has been shown to be less sensitive than fluorescence microscopy.

The sample may also be sent to be cultured on selective medium but as the bacterium has a slow generation time the culture period may take as long as 6 to 8 weeks (Wilson *et al.*, 1997). In developed countries where funding is more readily available, polymerase chain reaction (PCR) based techniques are being used to identify whether a patient is infected with *M. tuberculosis* and kits for these purposes have been available since the 1990s, although sputum culture still remains the primary technique performed to verify *M. tuberculosis* infection (Causse *et al.*, 2011).

---

In some instances a tuberculin skin test can be performed to test for infection. The most commonly used tuberculin skin test is the Mantoux test (Rothel & Andersen, 2005). This test involves the injection of the tuberculin antigen into the flexor surface of the forearm where a reaction will occur with the antibodies formed by someone infected with *M. tuberculosis* (Huebner *et al.*, 1993). The site of injection will show a delayed-type hypersensitivity response. This response reaction needs to be larger than a certain diameter before a positive result can be established.

Vaccines have been developed for the prevention of TB infection and the most widely used is the bacillus Calmette-Guerin (BCG) vaccine (Rothel & Andersen, 2005). This vaccine is intended to prevent infection by bacillus bacteria and is administered to people in regions of high risk for infection. In developing countries most of the population receives the BCG vaccine which causes the formation of antibodies which may interfere with the Mantoux test, where a false-positive might be observed. For this reason the Mantoux test is not generally used in developing countries.

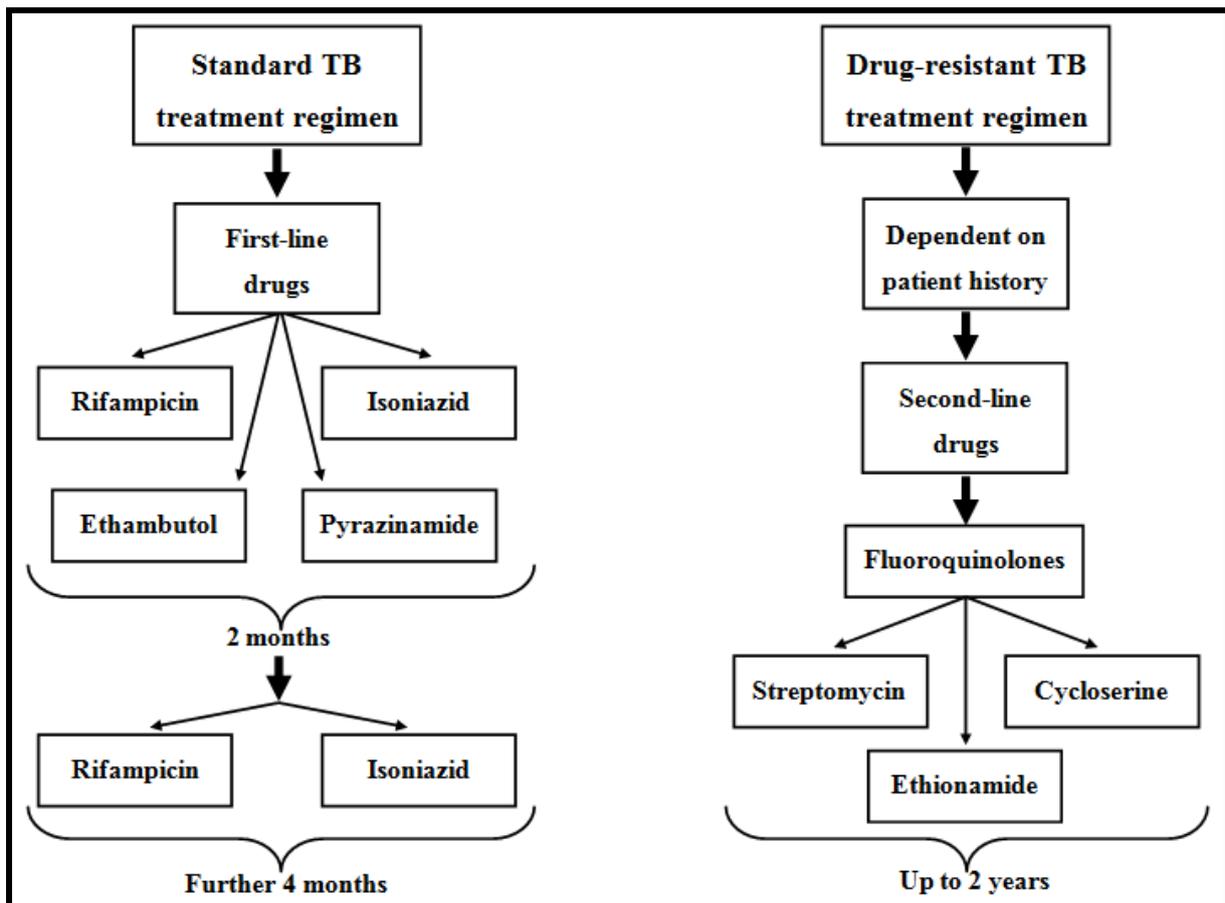
In developed countries, however, the BCG vaccine is not commonly administered unless the individual is at high risk of infection. The Mantoux test is more commonly performed in these regions (Teo & Shingadia, 2006). Interferon- $\gamma$  release assays (IGRAs) are assays that measure the release of interferon- $\gamma$  which is secreted by the body in response to inflammation and infection, and so can also be used to show infection by *M. tuberculosis*. It too may present false-positive results and cannot distinguish between latent, active or past infection with *M. tuberculosis* (Sester *et al.*, 2011). Chest x-rays are sometimes performed on patients with a chronic cough but, although the x-ray may show regions of granulomatous tissue, they are not specific enough to confirm tuberculosis and will rather aid in the determination of the patient's prognosis (Herzog, 1998).

### ***1.1.2.2 Disease Progression and Treatment***

As mentioned previously, the bacterium can be contracted from airborne, aerosolised particles produced by infected individuals. Once these particles are inhaled and enter the alveoli they have access to their primary site of infection, which are reticuloendothelial macrophages (Wagner *et al.*, 2005). Infection is initiated when mycobacteria enter the endosome of the macrophage, where they can grow and divide without the probability of

fusing with lysosomes and being digested. The exact mechanism by which the bacterium escapes the fusion of the endosome to the lysosome is poorly understood. The host system responds by encasing the invading pathogen with varying types of white blood cells forming a granuloma (Saunders & Britton, 2007). On the periphery of a granuloma are T cells, macrophages (both infected and uninfected), foamy giant cells and, at the centre, a firm caseous core (Stewart *et al.*, 2003). The entire granuloma or tubercle is isolated from the surrounding tissue by a fibrotic capsule. This asymptomatic stage is known as latent infection. Once the infectious bacteria overwhelm the immune system's defences, the granuloma will perforate and allow for the bacteria to spread and symptoms will begin to appear. This stage is termed active infection.

The treatment regimen for TB is personalised for each patient but Figure 1.2 illustrates a general strategy for the treatment of both drug-susceptible and drug-resistant TB.



**Figure 1.2** Standard and drug-resistant TB treatment regimens. Standard, non-drug-resistant TB treatment comprises of a 6 month period where a combination of at least 6 antibiotics are used. In contrast, drug-resistant TB treatment may last as long as 2 years involving multiple antibiotic treatments in varying stages.

---

First-line drugs are used for a period of six months for cases of drug-susceptible TB (Caminero *et al.*, 2010). *M. tuberculosis* is considered to be constantly mutating and has gained resistance to many antibiotics which impedes attempts to eradicate the disease. With the increasing incidence of drug-resistant *M. tuberculosis* being seen in Europe, Asia and sub-Saharan Africa new treatment regimens have been developed (Lawn & Zumla, 2011). The exact combination of drugs used is dependent on the patient's history of previous *M. tuberculosis* infection, adherence to previous regimens and other factors. In most cases of drug-resistant TB, secondary anti-tuberculosis drugs, such as fluoroquinolones, need to be incorporated (Caminero *et al.*, 2010).

## 1.2 IRON AND TUBERCULOSIS – ADAPTATION TO HUMAN INFECTION

Iron is a vital element required for the proliferation of *M. tuberculosis*. For example, iron is an obligate co-factor for 40 different enzymes encoded by the genome of *M. tuberculosis* (Cole *et al.*, 1998). Since *M. tuberculosis* cannot produce iron it must sequester this essential element from its host. Over time the bacterium has developed methods for outcompeting its host for its iron. One of the methods by which *M. tuberculosis* can obtain iron from its host is through the development of siderophores, which are iron-binding molecules (Schaible & Kaufmann, 2004). *M. tuberculosis* produces two types of siderophores, namely carboxymycobactins and exochelins. The highest concentrations of siderophores are produced when the bacterium is grown in iron-limited conditions, as in the case of TB infection when the host limits the bacterium's interaction with its iron stores (Ratledge, 2004).

Mycobactins, of which carboxymycobactins are a variant containing a carboxylic acid group, have a high affinity for iron, as do exochelins. The term exochelin is a dated term still used in some studies as a collective term for all extracellular siderophores which are found in mycobacteria (Ratledge, 2004). With the development of these siderophores, *M. tuberculosis* is able to sequester iron from host cells despite having to compete with host iron-binding proteins such as transferrin (TF) and lactotransferrin (LTF) (Olayanmi *et al.*, 2007). Apart from the development of siderophores, mycobacteria have the ability to block the maturation of the phagosome and its subsequent phago-lysosomal fusion within host macrophages (Schaible & Kaufmann, 2004; Saunders & Britton, 2007). The exact mechanism by which the

---

bacterium does this is unclear but can at least partly be attributed to the ability of *M. tuberculosis* to hydrolyze phosphatidylinositol 3-phosphate, which is a lipid required for phago-lysosomal fusion (Vergne *et al.*, 2005). Since the phagosome does not fuse with a lysosome and the pH environment of the phagosome does not become acidic, the bacteria can proliferate freely without being destroyed.

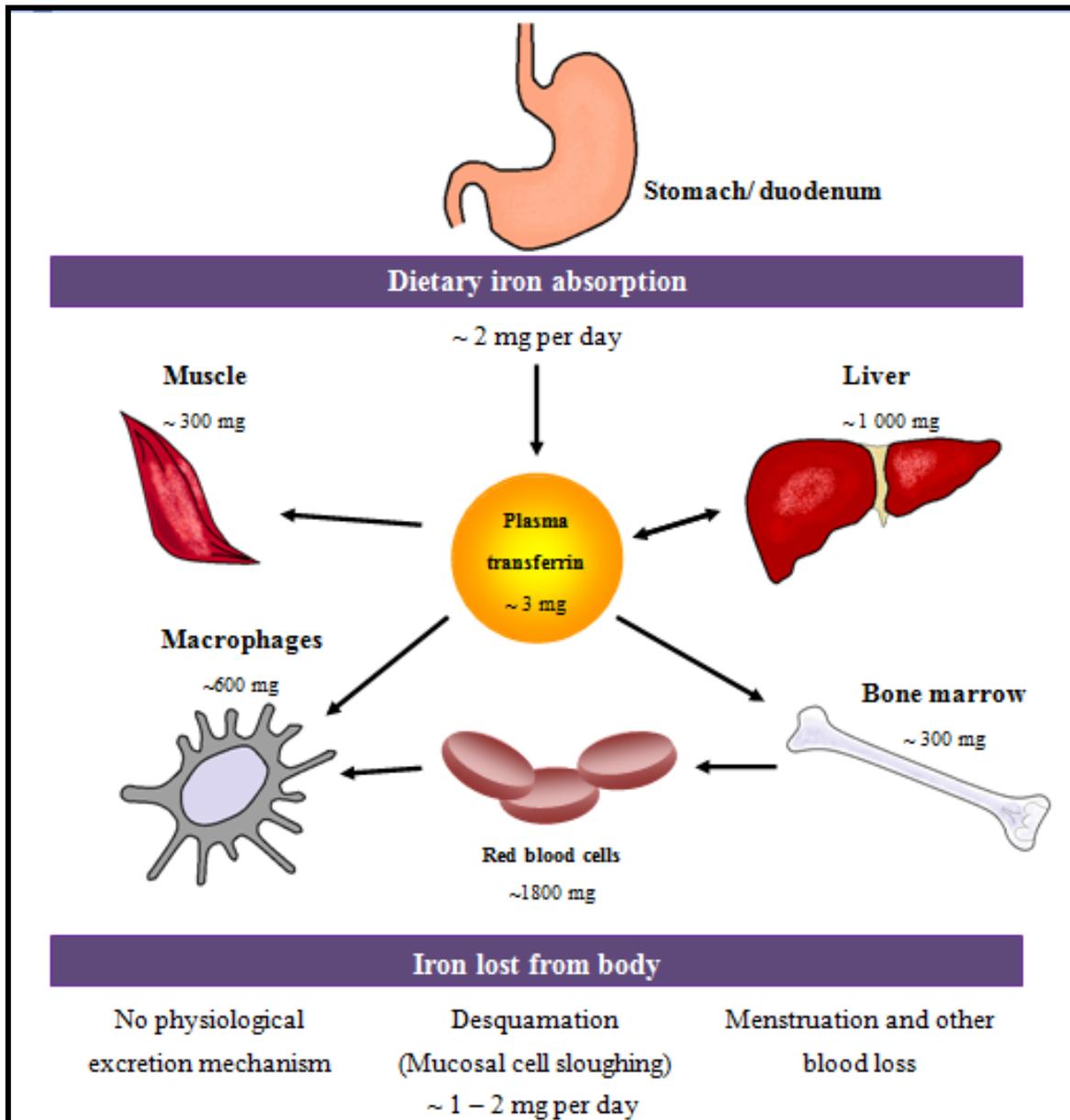
The relationship between iron and pathogenic infection is being widely explored. Schaible and Kaufmann (2004) previously observed that individuals with higher iron levels in their system are more susceptible to pathogenic infection, not only from *M. tuberculosis* but from a number of other bacteria. Furthermore, Gordeuk *et al.* (1996), Schaible *et al.* (2002), and Cronjé *et al.* (2005) have suggested that there is an association between the metabolism of iron – and more specifically iron overload – and *M. tuberculosis* susceptibility. Lounis *et al.* (2001) provided unequivocal evidence that *M. tuberculosis* infections are enhanced by increased iron levels in mice. These researchers go on to suggest that further study on the relationship between iron and *M. tuberculosis* is merited.

### **1.3 IRON IN THE HUMAN BODY**

#### **1.3.1 Overview**

Iron is not only important for bacterial proliferation but is an essential element required by almost all forms of life, except perhaps lactic acid bacteria, where it acts as both an electron acceptor and donor in many crucial reduction-oxidation (redox) reactions (Ong *et al.*, 2006). Since iron has this property, it plays a critical role in many physiological processes such as DNA biosynthesis, gene regulation, oxygen transport, respiration, and the citric acid cycle (for energy production), and for the incorporation into proteins as a co-factor; all of which are necessary for the existence of many organisms.

The human body contains approximately 3 to 5 g of iron at any one time, depending on the weight and gender of an individual (Papanikolaou & Pantopoulos, 2005). An example of the difference that gender makes is that women tend to have slightly less iron in their bodies compared to men, which is explained by iron lost through blood loss during menstruation.



**Figure 1.3** Distribution of iron in the human body. Iron is absorbed in the duodenum by the brush border epithelial cells. Iron contained within the muscles is primarily found within myoglobin while the iron within red blood cells and bone marrow is primarily found in haemoglobin. Liver iron is mostly stored within ferritin where it serves as the body's main storage site for iron. Redrawn from Papanikolaou & Pantopoulos (2005).

The daily diet of an average individual contains roughly 15 mg of iron of which only 1 to 2 mg is absorbed (Ganz, 2007). Iron is mainly stored in haem and forms part of haemoglobin, which is used to transport oxygen in red blood cells via the bloodstream, and myoglobin, which is stored in muscle tissue. The other primary repository for iron is within the ferritin protein found in the liver (the major iron storage organ). Ferritin is a protein with the unique ability to form a hollow core that allows for the storage of up to 4 500 iron ions at a time

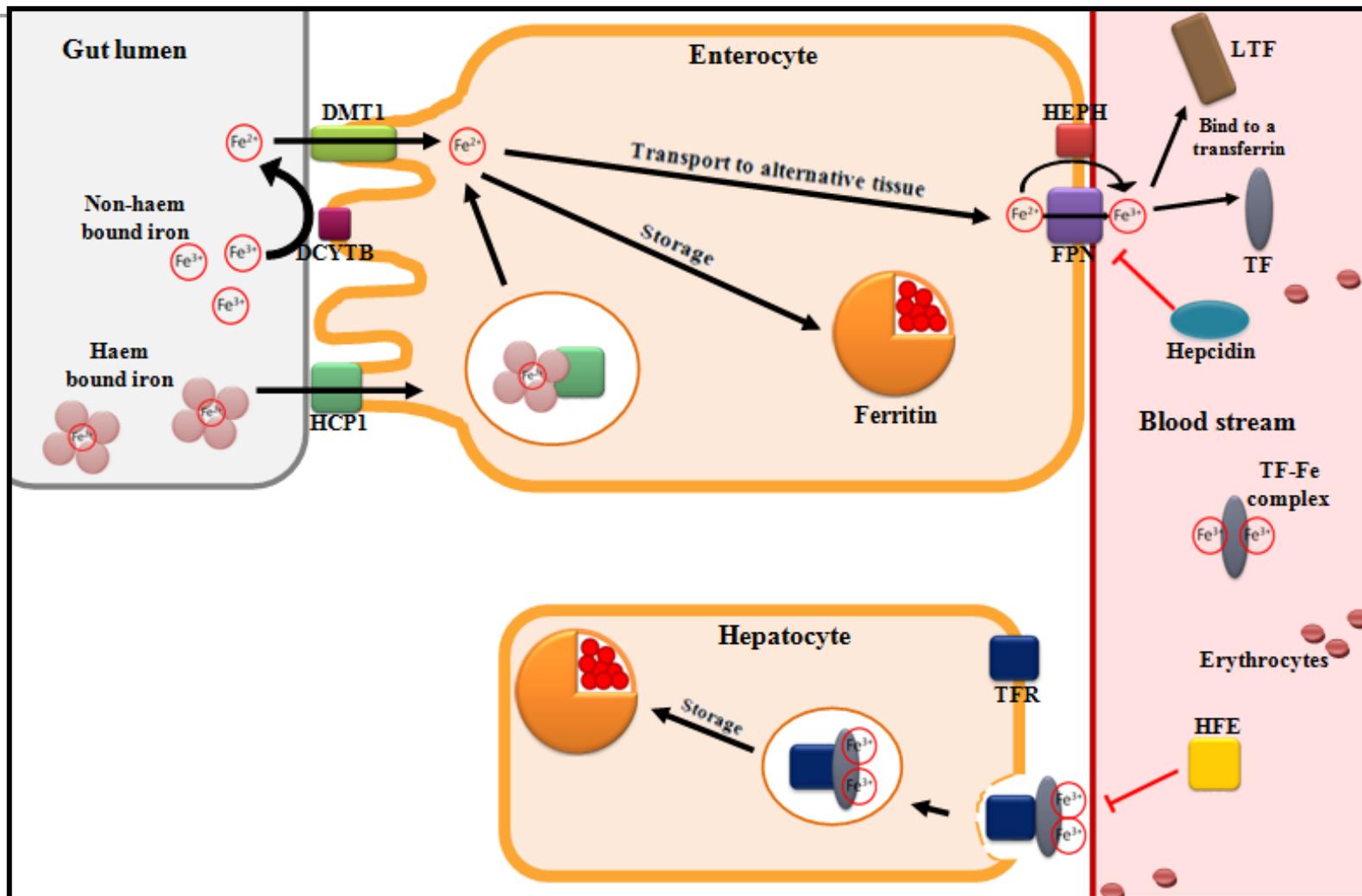
---

(Ong *et al.*, 2006). The remaining distribution of iron in the body can be seen in Figure 1.2. The largest portion of iron within the body originates from erythrocytes (Papanikolaou & Pantopoulos, 2005). Large amounts of iron are required for erythropoiesis and is supplied through the recycling of senescent erythrocytes by the reticuloendothelial macrophages. The recycled iron is then returned to the plasma to be bound to TF.

Too much iron can be detrimental for the body and can lead to disorders such as hereditary haemochromatosis (iron overload), while too little iron is also harmful and can lead to anaemia (Schaible & Kaufmann, 2004). Since there is no physiological excretion mechanism for iron other than being lost through sloughing of mucosal cells, menstruation, and other blood loss, and since so little is absorbed daily from food intake, it is imperative to tightly regulate absorption, storage and transport of iron throughout the body in order to maintain iron homeostasis (Papanikolaou & Pantopoulos, 2005). These processes are briefly discussed in the sections to follow and Figure 1.4 illustrates these processes.

### **1.3.2 Iron Absorption and Storage**

Once food in the form of the bolus enters the duodenum, the brush border epithelial cells absorb iron and other nutrients required by the body (McKie *et al.*, 2001). The iron that is absorbed is either non-haem bound or haem bound in nature. In the alkaline pH environment of the duodenum (pH 7.0-8.5), non-haem bound iron is predominantly found in the ferric ( $\text{Fe}^{3+}$ ) form, which is highly insoluble (Miret *et al.*, 2003). In order to be absorbed, it is necessary for the iron to be reduced from ferric to ferrous ( $\text{Fe}^{2+}$ ) iron, which is the soluble form. This reduction is facilitated by the duodenal cytochrome B (DCYTB) protein, as illustrated in Figure 1.4.



**Figure 1.4** The absorption, storage and transport of iron. Iron is absorbed from the gut lumen in either its haem bound or non-haem bound state. Non-haem bound iron is absorbed by divalent metal transporter 1 (DMT1) in conjunction with duodenal cytochrome b (DCYTB) which reduces the iron to its  $\text{Fe}^{2+}$  form. Haem-bound iron is absorbed by haem carrier protein 1 (HCP1). Iron is then either stored in ferritin or transported out of the cell to an alternative tissue. Iron is exported from the cell by the interaction of ferroportin (FPN), an iron transporter, and hephaestin (HEPH), a ferroxidase, so that it is present in the bloodstream in its  $\text{Fe}^{3+}$  form. The proteins transferrin (TF) and lactotransferrin (LTF) bind iron in the bloodstream in order to facilitate its transport to cells that require it. TF binds to transferrin receptor (TFR) and the entire complex is internalized in order to transport iron into a cell. There are protein regulators in the bloodstream in the form of hepcidin (HEP), which regulates FPN, and haemochromatosis (HFE), which regulates TF.

---

The reduction of the iron allows for it to be transported into the cell by the divalent metal transporter 1 (DMT1) protein (Papanikolaou & Pantopoulos, 2005). Haem bound iron is absorbed into the cell by a protein known as haem carrier protein 1 (HCP1) and upon entry to the cell the iron is dissociated from the haem by a haem oxygenase (Raffin *et al.*, 1974). Once  $\text{Fe}^{2+}$  enters the cell it is either exported into the bloodstream for use by the body or stored by ferritin binding. This is because this form of iron is highly unstable when unbound and may cause the formation of reactive oxygen or nitrogen intermediates (ROIs and RNIs respectively). These reactive intermediates can damage the cell during oxidative bursts and can lead to the formation of free radicals, which have been shown under some conditions to cause the development of cancer (Dreher & Junod, 1996). The storage site for iron within the cell is within the protein ferritin, which is primarily found in the liver (Papanikolaou & Pantopoulos, 2005). When iron is required elsewhere in the body it is exported into the bloodstream where it can be transported to the tissue where it is required.

### 1.3.3 Iron Transport

Iron is exported from the cell by the only known iron transport protein known as ferroportin (FPN). In favourable conditions, the pH of blood is neutral (pH 7.0-7.4), which favours the formation of  $\text{Fe}^{3+}$  (Vaupel *et al.*, 1989). However, for iron to be bound in the bloodstream it needs to be oxidised into its more stable and insoluble  $\text{Fe}^{3+}$  form. A ferroxidase, such as hephaestin (membrane-bound) or ceruloplasmin (unbound in the bloodstream), facilitates the oxidation of  $\text{Fe}^{2+}$  into  $\text{Fe}^{3+}$  and is then exported from the cell into the bloodstream (Ganz, 2007; Geissler & Singh, 2011). Once in the bloodstream,  $\text{Fe}^{3+}$  binds to either TF or its homolog LTF. The transferrin family (including TF and LTF) can bind two  $\text{Fe}^{3+}$  ions at a time and these globular proteins allow for the movement of iron throughout the body (Papanikolaou & Pantopoulos, 2005).

When iron bound to TF ( $\text{TF-Fe}^{3+}$ ) needs to be imported into an awaiting cell, this complex must first bind to a cell surface receptor known as the transferrin receptor (TFR) (Levy *et al.*, 1999; Ponka *et al.*, 1998). Once the  $\text{TF-Fe}^{3+}$  molecule is bound to TFR, the complex is internalised via endocytosis and subsequently iron is imported into the cell. Unlike TF, LTF is not required to bind to TFR in order to import iron into the cell, but the exact mechanism by which it does this is poorly understood, and it is thought LTF receptors may be involved (Papanikolaou & Pantopoulos, 2005).

#### **1.4 GENES UNDER INVESTIGATION**

The scope of this thesis is to investigate the promoter region of genes involved in the iron metabolism pathway and whether there is a correlation with *M. tuberculosis* susceptibility. Five genes were chosen to be investigated in this study based on their importance in iron metabolism, namely in absorption (*CYBRD1*), regulation (*HAMP* and *HFE*), transport (*LTF*), and export (*SLC40A1*). Information on these genes regarding their structure, length, protein product, orientation, and general function are given in Table 1.1.

**Table 1.1 General information for genes under investigation**

GENE INFORMATION						PROTEIN INFORMATION			
Name (symbol)	Encode accession number	Length (bp)	Genome position (orientation)	Chromosome position	Number of exons	Product	Length (amino acid residues)	Weight (kilo Daltons)	General function
<i>Cytochrome B Reductase 1 (CYBRD1)</i>	ENSG00000071967	35 887	172 378 757 – 172 414 643 (plus strand)	2q31.1	4	Duodenal cytochrome B (DCYTB)	286	32	Reduces Fe <sup>3+</sup> to Fe <sup>2+</sup>
<i>Hepcidin Anti-Microbial Peptide (HAMP)</i>	ENSG00000105697	2 637	35 773 410 – 35 776 046 (plus strand)	19q13.1	3	Hepcidin	84 (25 <sup>*</sup> )	9	Binds to FPN which results in FPN degradation
<i>Haemochromatosis (HFE)</i>	ENSG00000010704	11 124	26 087 448 – 26 098 571 (plus strand)	6p21.3	7	Haemochromatosis (HFE)	348	40	Binds to TFR; inhibits TF binding
<i>Lactotransferrin (LTF)</i>	ENSG00000012223	49 589	46 477 136 – 46 526 724 (minus strand)	3p21.31	17	Lactotransferrin (LTF)	710	78	Iron transport in macrophages
<i>Solute Carrier Family 40 Member A1 (SLC40A1)</i>	ENSG00000138449	23 180	190 425 305 – 190 448 484 (minus strand)	2q32.2	8	Ferroportin (FPN)	571	63	Iron export from cell

\*The peptide length given is for that of the prepropeptide whereas the most common form of the mature peptide is indicated in brackets. Abbreviations: bp = base pairs; *CYBRD1* = cytochrome B reductase 1 gene; DCYTB = duodenal cytochrome B protein; FPN = ferroportin protein; *HAMP* = hepcidin antimicrobial peptide gene; *HFE* = Haemochromatosis gene; HFE = Haemochromatosis protein; *LTF* = lactotransferrin gene; LTF = lactotransferrin protein; *SLC40A1* = solute carrier family 40 member A1 gene

### 1.4.1 Cytochrome b reductase 1 (*CYBRD1*)

At physiological pH (ideally pH 7.4), iron is predominantly found in the  $\text{Fe}^{3+}$  insoluble form (Miret *et al.*, 2003). In order for a cell to take up iron, the iron needs to be in the  $\text{Fe}^{2+}$  form, which is soluble. However, the  $\text{Fe}^{2+}$  form is highly unstable and can facilitate the formation of ROIs and so will quickly be oxidised into the  $\text{Fe}^{3+}$  form. Although these reactive oxygen species (ROS) have a potentially anti-microbial function, the presence of too many may lead to the formation of free radicals and have a cytotoxic effect on the cell (Bjorstad, 2009; Martin *et al.*, 2004). In order for the iron to be reduced to its ferrous form the body has to produce iron reductases to facilitate the reaction (McKie *et al.*, 2001). One such iron reductase is DCYTB, which is produced by the gene *CYBRD1*. This membrane protein is highly expressed in the duodenal brush border membrane, i.e., the enterocytes, and also operates in the airway epithelial cells (Oakhill *et al.*, 2008; Collins, 2008). Since it is only possible to take up iron in its soluble form, ferric reductases are crucial proteins involved in the absorption of iron by cells in the body. DCYTB is not the only ferric reductase present in the body but it is one of the most important because it aids in the regulation of iron absorption from the diet. DCYTB reduces iron so that the DMT1 protein (also known as NRAMP2) can import the iron into the cell (Papanikolaou & Pantopoulos, 2005).

Mutations in this gene have been associated with primary iron overload and anaemia, although the mechanism by which this occurs requires further elucidation (Zaahl *et al.*, 2004). The sequence of this gene is classified as belonging to the cytochrome  $b_{561}$  family (Oakhill *et al.*, 2008). The cytochromes  $b_{561}$  are a class of intrinsic membrane proteins that play a role in ascorbate (vitamin C) regeneration (Verelst & Asard, 2003). These are multi-pass membrane proteins that non-covalently bind 2 haem molecules. It is likely that the structural features of the cytochrome  $b_{561}$  family members, for example that they are highly hydrophobic and contain 6 transmembrane helices, play an important role in their functioning. It is thought that prokaryotes do not have cytochromes  $b_{561}$  and that their function in eukaryotes must be essential as they are relatively conserved between species in terms of the inner 4 helices found in the protein.

It has been postulated that DCYTB plays a role as a co-factor in the NADPH oxidase dependent respiratory burst (Collins, 2008). One hypothesis, put forward by Collins (2008), is that DCYTB levels are elevated in individuals experiencing microbial infection because the

---

protein allows for an ample supply of ferrous iron to be available to act as a co-factor for NADPH oxidase-mediated anti-microbial functions; this is supported by unpublished observations in which Dcytb-deficient mice showed reduced production of ROS.

#### 1.4.2 Heparin Antimicrobial Peptide (*HAMP*)

Heparin is considered a major regulator of iron levels in the bloodstream since it inhibits the export of iron from the cell. It does this by causing the internalisation and eventual degradation of the FPN protein responsible for exporting iron out of the cell (Collins, 2008). In order to perform this function, heparin has specific domains that recognise FPN and allow the two proteins to interact. The heparin protein is encoded for by the *HAMP* gene.

Mutations in this gene have been associated with Haemochromatosis type 2B, also known as Juvenile Haemochromatosis (Papanikolaou *et al.*, 2005). This disorder affects iron metabolism by causing excess iron to be deposited in tissue, which results in typical Haemochromatosis symptoms, including bronzing of the skin, cirrhosis, diabetes mellitus and cardiomyopathy (Nandar & Connor, 2011). It is called Juvenile Haemochromatosis because it presents at an early age unlike type 1 Haemochromatosis, which only presents later in life.

This prepropeptide is secreted by the liver to aid in the regulation of iron metabolism and is cleaved into different forms, ranging from 20 to 25 amino acids in length; the predominant form found in humans is that of heparin-25 (25 amino acids long) as indicated in Table 1.1 (Ganz, 2003).

Heparin is produced in response to high iron levels and inflammation and has also been shown to be produced in response to bacterial infection since it has antimicrobial activity and inhibits pathogen proliferation (Collins, 2008; De Domenico *et al.*, 2007; Ganz, 2003). The interaction between FPN and heparin inhibits iron from leaving the cell and in this way it prohibits bacterial growth by reducing iron availability to pathogens that sequester iron from the bloodstream (Ward *et al.*, 2011). Such pathogens include those that infect the gut, e.g., *E. coli* (Park *et al.*, 2001). Although the activity of heparin does not allow for the proliferation of pathogens that reproduce in the bloodstream, it creates the ideal environment for pathogens that infect from within cells, such as the infection of macrophages by *M. tuberculosis*, since it causes large amounts of iron to collect within the cell (Collins, 2008).

---

### 1.4.3 HFE (Haemochromatosis gene)

The other protein considered to be a major regulator of iron homeostasis is HFE. The HFE protein is encoded for by the gene *HFE*. This protein controls iron absorption by regulating the interaction between the TFR and TF. In order for iron to move into a cell it must be bound to TF in the blood. The TF-iron complex then binds to the surface protein TFR and this whole complex is then internalised, thereby moving iron into the cell (Nandar & Connor, 2011). The HFE protein binds to TFR and reduces TFR's affinity for TF. Since less TF-iron complexes are able to bind to TFR, the amount of iron that can be given by TF to a cell that requires it is reduced. This theory is under debate as some research has shown that the HFE protein can inhibit iron from entering the cell even when it cannot bind to TFR (Zhang *et al.*, 2003). Zhang and his colleagues (2003) used a mutated form of HFE that has a 5 000-fold decrease in affinity for TFR but could still bind beta-2-microglobulin and was not misfolded. They showed that cells with the mutated as well as the wild type HFE accumulated 30% less iron than cells with a non-functional form of HFE. This suggests that HFE might interact with a number of proteins which could affect the binding of the TF-iron complex to TFR in a cascade type manner. Mutations in this gene have most commonly been linked to the genetic iron overload disorder Hereditary Haemochromatosis. This disease is characterised by the accumulation of iron in tissues such as the liver, heart, and brain, and can result in symptoms such as cirrhosis, cardiomyopathy, diabetes mellitus, arthritis and, more speculatively, neurodegenerative disorders such as Alzheimer's disease (Allen *et al.*, 2008; Nandar & Connor, 2011)

The protein product is a membrane protein similar to major histocompatibility complex (MHC) class-I type proteins (Feder *et al.*, 1996). This protein consists of two alpha extracellular domains and a third alpha domain which spans the cell membrane. This third alpha region binds to beta-2-microglobulin which is involved in the innate immune system (Lebron *et al.*, 1998). Notably, this protein contains a histidine cluster, which is thought to be a binding site for iron as it is similar to some iron-binding sites seen in other proteins. Although a direct link has yet to be shown between the HFE protein and the host immune system, MHC class-I type proteins have previously been shown to play a role in the immune system and specifically in the cytoplasm during *M. tuberculosis* infection (Fairbairn, 2004) and thus it could be proposed that the HFE protein also plays a role in immunity. The HFE protein also contains a C1-set domain, which is almost exclusively found in proteins involved

---

in the immune system (Pamer & Cresswell, 1998). This further strengthens the hypothesis that HFE plays a role in the immune system.

#### 1.4.4 Ferroportin (SLC40A1)

Once iron is absorbed from the gut by the lining epithelial cells it has to be directed to the bloodstream to reach tissues in the body that require it. For this to occur the iron needs to be oxidised to its ferric form so that it can be bound to TF once in the bloodstream. *SLC40A1* is the iron gene which encodes for the only known FPN responsible for the regulation of iron efflux from duodenal epithelial cells (Collins, 2008). The FPN protein mediates the efflux of iron in the presence of a ferroxidase, such as hephaestin (membrane-bound) or ceruloplasmin (free in the plasma), which is required to oxidise the iron to  $\text{Fe}^{3+}$  in order for its movement out of the cell to occur (Geissler & Singh, 2011). Mutations in this gene have most commonly been linked to Hereditary Haemochromatosis type 4, or FPN disease, which is inherited in a dominant negative fashion (Pietrangelo, 2004). This disease can present in two manners. The mutated protein is changed so that hepcidin can no longer bind and regulate the protein thus resulting in increased transferrin saturation as more iron is present in the bloodstream. Alternatively, the mutated protein loses its function and iron accumulates in the cell because it cannot be exported to the bloodstream resulting in decreased transferrin saturation.

FPN is a multi-pass transmembrane protein which contains anywhere from 9 to 12 transmembrane domains depending on the researchers (Wallace *et al.*, 2010). Most researchers agree, however, that it is more likely that there are 12 transmembrane domains in the FPN protein (Liu *et al.*, 2005). This protein also carries a domain for the specific recognition of the hepcidin protein (Wallace *et al.*, 2010).

FPN is regulated by the protein hepcidin in order to control the efflux of iron from the cell (Collins, 2008). During bacterial infection, the body produces more hepcidin in order to degrade FPN and thus inhibits iron export out of the cell. In this way bacterial infection is supposed to inhibit bacteria that sequester their iron from the bloodstream. In the case of bacteria such as *M. tuberculosis*, the iron is sequestered from within the cell – in this instance the macrophage – and so the degradation of FPN actually allows for the proliferation of these pathogens. Recently researchers have shown that 4 single nucleotide polymorphisms (SNPs) in the *SLC40A1* gene have been associated with an increased risk of development of

---

tuberculosis and 1 SNP in this gene has been associated with a decreased risk for the development of the disease (Baker *et al.*, 2012).

#### 1.4.5 Lactotransferrin (LTF)

Since iron is required by many tissues in the body, a transport mechanism is required to deliver it to its destination. This transport is facilitated by either the TF protein or its closely homologous relative, LTF. The main difference between TF and LTF is that TF must bind to TFR in order for iron to move into the cell whereas LTF does not require this interaction. Another difference is that TF is the main iron binding transport protein in the serum and is found in all regions where iron transport is required, whereas LTF is mainly found at mucosal membranes (Olahanmi *et al.*, 2007). Mutations in this gene have been associated with a number of diseases, such as cystic fibrosis, Sjogren's syndrome and Alzheimer's disease (Moskwa *et al.*, 2007; Ohashi *et al.*, 2003; Kawamata *et al.*, 1993).

LTF is part of the transferrin family (a major iron-binding protein) and is predominantly found in secondary granules of neutrophils within milk and bodily secretions, and it functions in close association with the body's immune system as it has antimicrobial activity (Singh *et al.*, 2002; Orsi, 2004). This globular glycoprotein has been implicated in the transferred immunity from mother to infant via breast milk, which contains a high concentration of LTF. It is thought to aid in strengthening the infant's immune system against pathogenic infection (Wheeler *et al.*, 2007). The LTF protein is folded into two lobes, viz. the carboxyl (C-) and the amino (N-) terminal halves, which are joined by an  $\alpha$ -helix (Baker & Baker, 2005; Ward *et al.*, 1996). The lobes are further sub-divided into two  $\alpha/\beta$  domains designated N1 and N2, or C1 and C2, depending on the lobe in question, in which a deep cleft can be found, which is the binding site for iron (Baker & Baker, 2005). These lobes have a high affinity for iron and reversibly bind this molecule when required.

LTF is produced by the macrophages in response to high serum iron concentrations and pathogen infection (Ward *et al.*, 2011). LTF's main function is to bind iron and transport it to the small intestine and other parts of the body. LTF also regulates iron homeostasis as part of the non-specific immune system, and by doing so regulates cell growth and differentiation (Horiuchi *et al.*, 2009). TFs can bind two  $\text{Fe}^{3+}$  ions in association with an anion, such as bicarbonate, and by doing so facilitate the transport of iron throughout the body. Unlike TF,

---

LTF has antimicrobial activity that acts in two ways: first, a bacteriostatic activity that inhibits bacteria from sequestering iron by competing with them; and second, a bacteriocidal activity by which LTF binds to the outer membrane of bacteria with high lipopolysaccharide contents, thereby increasing bacterial membrane permeability by causing the release of polysaccharides (Orsi, 2004). By increasing the bacterial membrane's permeability, the pathogen becomes more vulnerable to breakdown by phagosomes within the macrophage (Ellison *et al.*, 1988). LTF can also stimulate the production of apoptotic signals within an infected cell so that it is destroyed before bacterial infection can spread (Orsi, 2004). *M. tuberculosis* has been shown to have the ability to acquire iron from both the cytoplasm of the macrophage (endogenous), as well as from TF and LTF found extracellularly (exogenous) (Olanmi *et al.*, 2007). However, it has also been shown that *M. tuberculosis* acquires two- to three-fold more iron from LTF than from TF. This phenomenon most likely occurs because iron acquisition from TF involves TFR-mediated endocytosis whereas LTF does not and thus the pathogen has less competition for binding of the iron; and because LTF is the major iron-binding protein in the airway of humans and so would be more prevalent in this region than TF (Olanmi *et al.*, 2007; Finkelstein *et al.*, 1983; Quinlan *et al.*, 2002; Richardson & Ponka, 1997).

## 1.5 TRANSCRIPTIONAL REGULATION OF GENES INVOLVED IN IRON METABOLISM

### 1.5.1 Overview

Iron homeostasis is necessary for the normal functioning of the body and thus the tight regulation of the genes that maintain iron is very important. One of the mechanisms for controlling these genes is through transcriptional regulation, which, as the name suggests, occurs during transcription. Another mechanism is through post-transcriptional regulation. In this way the mRNA produced through transcription is either stabilised or degraded as required (Papanikolaou & Pantopoulos, 2005).

Promoters are regions of DNA comprising *cis*-regulatory elements because they regulate the expression of DNA adjacent to them and are generally found upstream of a gene (Sandelin *et al.*, 2007). Promoters may be composed of varying combinations of recognition sequences

---

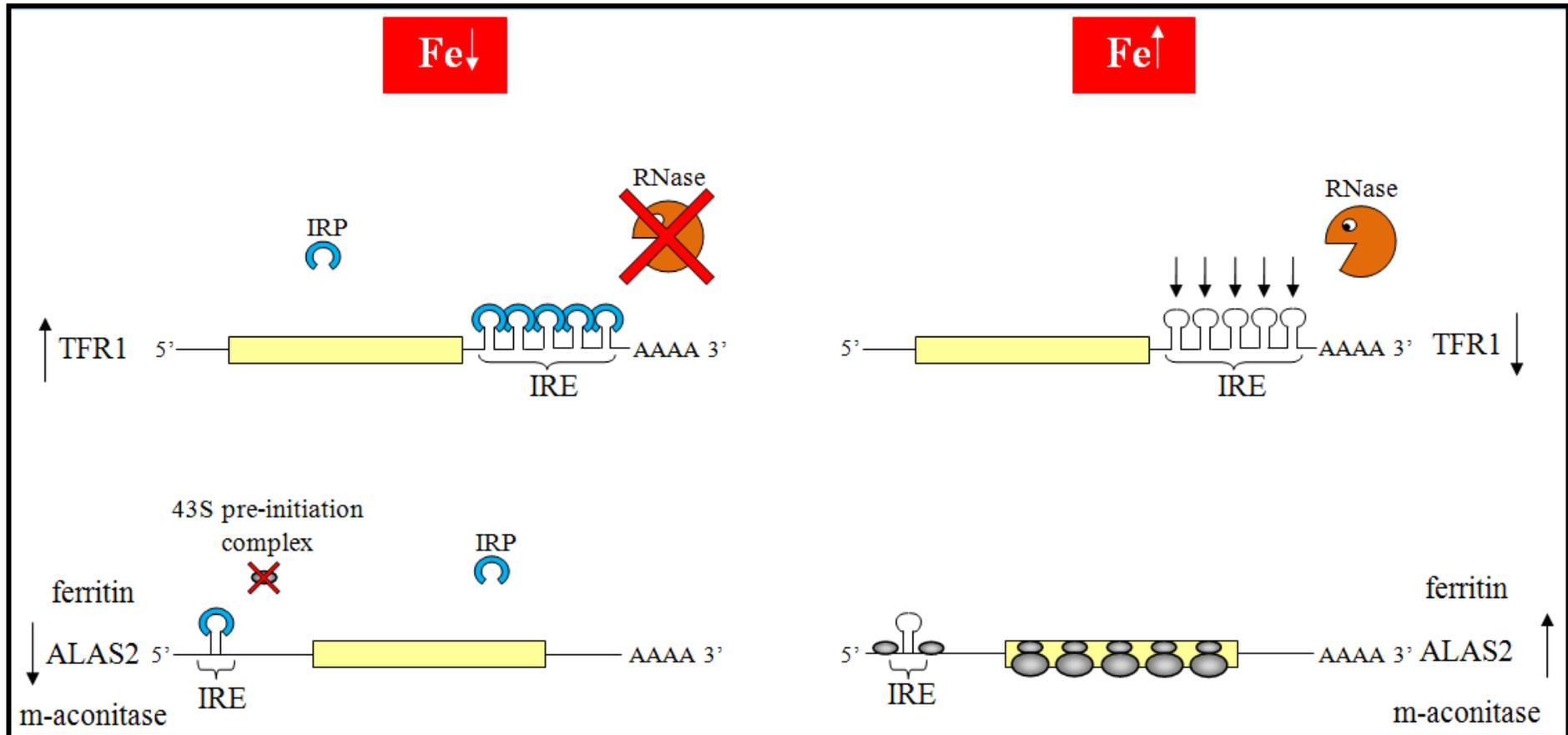
necessary for the binding of RNA polymerase II, the various transcription factors required for transcription and enhancers, but the exact composition of these promoters differs between genes.

Two forms of transcription initiation exist, namely focused and dispersed transcription initiation (Juven-Gershon & Kadonaga, 2010). During focused initiation, a single nucleotide, or small areas of several nucleotides, is responsible for the start of transcription. In contrast, during dispersed initiation, many weak start sites over a broad region are responsible for transcription to initiate. This type of promoter has also been referred to as a broad distribution (BR) promoter (Carninci *et al.*, 2006). BR promoters have been found to be more abundant than single start site promoters and typically occur in CpG islands (Juven-Gershon & Kadonaga, 2010). BR promoters tend to lack the BRE, TATA and DPE motifs commonly found in focussed initiation promoters (Sandelin *et al.*, 2007; Carninci *et al.*, 2006).

The promoter motifs discussed previously represent the most studied and simplest forms and do not encompass all known promoters. Furthermore, promoters are not the only motifs or regions of DNA that control gene expression (Juven-Gershon & Kadonaga, 2010). The subtle control of the interactions between both *cis*- and *trans*-regulatory elements is crucial to the precise expression of genes in terms of both spatial and temporal requirements in the body, and this is especially true for genes involved in the metabolism of a nutrient as important as iron. The regulatory elements involved in the expression of genes involved in iron metabolism are currently poorly understood; some of the known elements are described in the following sections.

### **1.5.2 Iron Response Elements (IREs)**

Iron response elements (IREs) are hairpin structures that serve as transcription factor binding sites located in either the five prime (5') or three prime (3') untranslated regions (UTRs) of genes known to be involved in iron metabolism (Hentze & Kuhnt, 1996). The IREs are recognised by a specific protein named the iron-regulatory protein (IRP) of which two homologs exist, IRP1 and IRP2 (Hentze & Kuhnt, 1996; Papanikolaou & Pantopoulos, 2005). This interaction is illustrated in Figure 1.3.



**Figure 1.5** The interaction of IRE and IRP in mRNA regulation. In conditions of low iron availability the IRP/IRE interaction allows for the increased expression of the transferrin receptor (TFR1) and decreased expression of ferritin, ALAS2 and m-aconitase. In this manner more iron will be able to enter the cell and less will be stored increasing the overall iron pool within the cell. Conversely, in high iron conditions the IRP protein is either degraded or inactive and cannot bind to IREs. This results in the destabilization of TFR mRNA and the simultaneous initiation of the binding of the small ribosomal subunit to the mRNA of ferritin, ALAS2 and m-aconitase which allows for translation to take place. In this way TFR expression is reduced and less iron can enter the cell. More iron is now stored in ferritin as more ferritin is being produced and this decreases the overall iron pool within the cell. Redrawn from Papanikolaou & Pantopoulos, 2005.

---

A common example of an IRE occurring in the 3'UTR of mRNA is that of the mRNA of TFR. During times of low iron levels, the IRP protein is active and binds to the IRE of the mRNA of TFR in its 3'UTR. The binding of IRP to the IRE in this region stabilises the normally unstable TFR mRNA and less degradation occurs by RNase. This results in increased expression of TFR and thus more iron is taken up into the cell (Papanikolaou & Pantopoulos, 2005). Conversely, when the cell is experiencing high iron levels, the IRP proteins are either inactivated (IRP1) or degraded (IRP2) and thus do not bind to TFR's mRNA. This causes the destabilization of the mRNA, which results in the decreased expression of TFR and less iron taken into the cell (Hentze & Kuhnt, 1996).

An example of an IRE occurring in the 5'UTR is that of ferritin (but it is also known to occur in the ALAS2 and m-aconitase genes) (Papanikolaou & Pantopoulos, 2005). In circumstances when the internal levels of iron are high within the cell more ferritin needs to be produced in order to store the iron and negate any negative effect it might have while unbound in the cell. During this time, the IRP protein is inactive or degraded and thus cannot bind to the IRE in the 5'UTR of the ferritin mRNA. This results in the normal translation of the mRNA and thus the increased expression of ferritin, which eventually leads to the decrease in intracellular iron. When the cell is depleted of iron the IRP proteins are active and can then bind to the IRE within ferritin's mRNA. This binding of IRP to the IRE prevents the 43S pre-initiation complex (the small ribosomal subunit) which facilitates the initiation of translation from binding and in this way inhibits the expression of ferritin (Hentze & Kuhnt, 1996). This will result in less ferritin being formed and less iron being stored and the eventual increase of free iron within the cell.

### 1.5.3 Other Regulatory Elements

The IRE transcription factor binding site is possibly the best characterised of all the binding sites involved in genes regulating iron homeostasis but is not the only binding site known. Other binding sites and their interacting proteins include: the GATA-box and its interaction with GATA- and its co-factor Friend of GATA (FOG), which have been shown to regulate the expression of the *HFE* gene; the CCAAT/enhancer-binding protein site, cyclic AMP response element-binding protein site, bone morphogenetic protein-response element site, and signal transducer and activator of transcription 3-response element site which are all

---

known to play a role in *HAMP* regulation; and the known interaction of binding factors such as SP1, SP3, IKLF with the promoter region of the *LTF* gene, which also contains response elements such as the TATA box and GATAAA (Bagu & Santos, 2011; Teng, 2002). Although there are many known interactions that regulate the expression of genes involved in iron homeostasis, the exact mechanism by which this regulation occurs and why there are differences in expression during developmental stages of these genes is still poorly understood, emphasizing the importance of research into these fields (Hentze & Kuhnt, 1996).

## 1.6 AIM AND OBJECTIVES

The aim of this study is to elucidate the role of five genes involved in iron metabolism – *CYBRD1*, *HAMP*, *HFE*, *LTF* and *SLC40A1* – in relation to *M. tuberculosis* susceptibility.

The key objectives identified in order to achieve this aim are:

- 1) to identify a cohort suitable for this study consisting of TB patients and healthy population-matched controls;
- 2) to determine the iron parameters (including serum-iron, serum-ferritin, serum-transferrin and transferrin saturation percentage) for the entire cohort;
- 3) to genotype the cohort for the promoter region of the five genes for novel and/or known variants;
- 4) to determine allele and genotype frequencies for the data;
- 5) to determine the effect of each identified variant through *in silico* analyses; and
- 6) to perform statistical analyses to determine whether there are any correlations between the identified variants, iron parameters measured and the susceptibility to *M. tuberculosis*.

## **Chapter 2**

### **DETAILED EXPERIMENTAL PROCEDURES**

## 2. DETAILED EXPERIMENTAL PROCEDURES

### 2.1 COHORT SAMPLING

#### 2.1.1 Patient and Control Groups

This project has been granted ethical approval from the Research and Ethics/Biosafety Committee, Faculty of Health Sciences, University of Stellenbosch (N10/06/194). The cohort consisted of Black, Xhosa-speaking individuals. The patient group consisted of 49 TB patients (29 males; 20 females). Whole blood samples from the patients were obtained from the Ikhwezi Clinic in Somerset West. The controls group consisted of 51 individuals in total. This group was further split for genotyping purposes depending on their ability to amplify and quality of genotyping. The first control group consisted of 29 healthy individuals (11 males; 18 females). For the patient group the mean age is 31.6 (standard deviation [S.D.]  $\pm 7.21$ ) with an age range of 21-48. For the control group the mean age is 46.5 (S.D.  $\pm 12.17$ ) with an age range of 25-66. This group was used to genotype the promoter region of all five genes under investigation.

For the *CYBRDI* genotyping comparisons the above-mentioned controls could not all be amplified and so a second control group was added for comparison. The samples from this control group originated from a previous study and as such do not have iron parameter data available. This group consisted of 22 healthy individuals (11 males; 11 females). The mean age for this group is 58.8 (S.D.  $\pm 14.2$ ) with an age range of 40-89.

#### 2.1.2 Iron Parameter Data

BD Vacutainer® SST™ Serum Separation Tubes were used for blood collection from patient and control groups to determine iron parameters. The following parameters, with reference ranges shown in brackets, were measured: serum ferritin (reference range: 15-322  $\mu\text{g/l}$ ), serum iron (reference range: 9-30  $\mu\text{mol/l}$ ), serum transferrin (reference range: 1.74-3.60 g/l) and transferrin saturation percentage (%) (reference range: 20-55%). The data for C-Reactive Protein levels were also included (reference range: 0-10 mg/l). The WHO (2004) has stressed

the importance of the establishment of standard reference ranges for iron parameters and so, at their suggestion, reference ranges for these parameters were obtained from these laboratories where the assays were performed (Davies Pathology and the NHLS).

### 2.1.3 DNA Extraction

Ethylenediaminetetraacetic acid (EDTA) tubes were used for the collection of blood for DNA extraction. DNA was extracted from whole blood samples using an adaptation of the Miller method (Miller *et al.*, 1988). Whole blood (10 ml) was transferred to a 50 ml polypropylene tube. Cold Cell Lysis Buffer (CLB) (final concentration : 116.25 mM NH<sub>4</sub>Cl; 0.75 mM KHCO<sub>3</sub>; 0.075 mM EDTA) was added and this was mixed by inversion. This solution was then placed on ice for 15-30 minutes (mixing by inversion being performed every 5 minutes). The tubes were centrifuged for 15 minutes at 1 500 rpm on a Hermle Z200A centrifuge (Labnet) and the supernatant discarded thereafter.

Subsequently, 10 ml of cold Phosphate Buffer Saline (PBS) (final concentration: 0.027 M KCl; 137 mM NaCl; 8 mM Na<sub>2</sub>HPO<sub>4</sub>; 1.5 mM KH<sub>2</sub>PO<sub>4</sub>) was added to the pellet which was then mixed and centrifuged for 10 minutes at 1 500 rpm (Hermle Z200A centrifuge) (Labnet). The supernatant was, once again, discarded and the pellet dissolved in Nuclear Lysis Buffer (NLB) (final concentration: 9 mM Tris-Cl; 0.3604 M NaCl; 1.8 mM EDTA; solution set to pH 8.2 using concentrated HCl), Proteinase K (final concentration: 0.09 mg/ml) (Sigma) and 0.909% (w/v) Sodium Dodecyl Sulphate (SDS). This solution was mixed and incubated overnight at 56°C in a water bath. The following day NaCl (final concentration: 1.3857 M) was added to the solution and shaken vigorously for 1 minute. This solution was centrifuged for 20 minutes at 2 500 rpm on a Hermle Z200A centrifuge (Labnet) at room temperature. Thereafter the supernatant was transferred to a clean 50 ml polypropylene tube.

Three volumes of ice cold ±99.9% (v/v) ethanol (EtOH) were added to one volume of the supernatant. The resulting contracted DNA bundle formed was removed with a needle and carefully placed into a microcentrifuge tube containing 70% (v/v) EtOH. This mixture was centrifuged at 4°C for 5 minutes at 14 000 rpm on a Spectrafuge 16M centrifuge (Labnet). The ethanol was subsequently discarded and the pellet allowed to air dry. The pellet was resuspended in 100-500 µl of dH<sub>2</sub>O depending on the size of the pellet. Once the DNA was

resuspended the concentration and purity of each sample was determined through the use of a Nanodrop® ND-1000 Spectrophotometer v3.0.1 (NanoDrop Technologies Inc.).

## 2.2 POLYMERASE CHAIN REACTION (PCR) AMPLIFICATION

### 2.2.1 Oligonucleotide Primers

Primers were designed to amplify the promoter region of each gene (up to 2 000 bp upstream of the first coding ATG). This region encompasses most of the transcription factor binding sites (TFBSs) necessary for gene expression as previously confirmed (Strickland, 2013). The reference sequences for *CYBRD1*, *HAMP*, *HFE*, *LTF* and *SLC40A1* were obtained from Ensembl ([www.ensembl.org](http://www.ensembl.org)). Two primer pairs were used to amplify two overlapping fragments for each gene promoter as indicated in Table 2.1. These fragments were subjected to semi-automated bi-directional sequencing analysis in order to identify novel/known variants. Tables 2.2 and 2.3 show the primer pairs used to amplify fragments used for genotyping the remaining samples either through Heteroduplex Single-Strand Conformation Polymorphism (HEX-SSCP) analysis (-1631 C/T, -323 C/T, -188 C/T [*HAMP*]; -1168 A/G, -331 A/C [*HFE*]; -1377 T/G, -746 del AAA/ins AA, -457 T/C, -437 C/G [*LTF*]; -750 G/A, -662 C/T, -501 T/C [*SLC40A1*]), Restriction Fragment Length Polymorphism (RFLP) analysis (-582 A/G [*HAMP*]; -561 A/G, -467 G/C [*HFE*]) or semi-automated bi-directional sequencing analysis (-1849 T/G, -1844 C/G, -1813 C/T, -1749 del (T)<sub>6</sub>G, -1540 G/A, -1477 G/A, -1459 T/C, -1346 T/C, -1272 T/C, -1011 T/A, -849 C/G, -627 T/C, -624 G/A, -492 A/G, -454 C/T, -449 C/G, -399 T/G, -397 A/C, -238 A/G, -166 C/G [*CYBRD1*]; -1470 C/T, -1461 T/C, -1399 G/A, -1355 G/C, -1098 G/A [*SLC40A1*]).

Primers were designed using PrimerQuest (Integrated DNA Technologies Inc.). Primers indicated by an asterisk in the tables were previously designed also using PrimerQuest (Integrated DNA Technologies) and were made available for use in this study.

## Chapter 2: Detailed Experimental Procedures

**Table 2.1 Oligonucleotide primers and PCR conditions for amplification of DNA for semi-automated bi-directional sequencing**

Gene Name	Ensembl Accession Number	Fragment	Primer Name	Primer Sequence (5' - 3')	Product Size (bp)	T <sub>M</sub> (°C)	T <sub>A</sub> (°C)	Mg <sup>2+</sup> (mM)	Betaine (%)	PCR amplification cycle	Enzyme used
<i>CYBRD1</i>	ENSG00000071967	1	BP1F Seq	AGCTCAAGTGATTCTCTGCCT	972	56.3	75/68	1.5	0	A	
			BP1R Seq	TAGAGGCTAAACAAGCACTGGC		57.2					
		2	BP2F Seq	GGGTTTCAGCTGGTCAGCATC	855	58.5	61	2	0	B	
			BP2R Seq	ACTTCTTGGGGCTGTCTCCG		56.2					
<i>HAMP</i>	ENSG00000105697	1	PP1F Seq	CTGGTGACTTGGCTGACACT	980	57.2	63	1	12	B	
			*PP2R	TCAAGACTAGCCTGGGCAAC		57.1					
		2	*PP3F	CACGCCTGGCTAAATTGTT	768	54.0	53	1.5	0	C	
			PP2R Seq	CCGAGTGACAGTCGCTTT		55.0					
<i>HFE</i>	ENSG00000010704	1	HP1F Seq	CGCCCGGCTAATTTTGTAT	866	53.6	N/O	N/O	N/O	N/O	
			*HP2R	ATAGGTAAGACCATGCACAG		51.7					
		2	*HP3F	TGAGGTTTGGGCTGTATCTG	919	54.5	66	3	0	B	
			*HP7R	CTCAGGAGATGCCAGTAA		53.6					
<i>LTF</i>	ENSG00000012223	1	*LP (NheI)	AATGCTAGCGCTGAGCCTCTTGGTAGTGG	961	64.9	69	2	0	D	
			LP1R Seq	TGGGCAACATAGGGAAACCCC		59.8					
			LP2F Seq	CACCACCATGCTGGGCTAATT		58.0					
			LP2R Seq	TAGTCTGCAAGCCCCTAGGAG		58.1					
<i>SLC40A1</i>	ENSG00000138449	1	SP1F Seq	GACCCTATCTCCACGGTTAGC	926	57.0	55	1.5	0	C	
			*SP3R	ATGCCACAGAGGCCGCTTTC		60.8					
		2	*SP3F	GGGAGAGGAATGATGGTGA	963	54.1	62/56	1.5	12	A	
			*SP6R	GCTAACACTGTAGCTGAAGTTGG		55.7					

\*Primers previously designed. Abbreviations: 5' = 5-prime; 3' = 3-prime; % = percentage; °C = degrees Celsius; bp = base pairs; *CYBRD1* = cytochrome B reductase 1 gene; F = forward primer; *HAMP* = hepcidin antimicrobial peptide gene; *HFE* = Haemochromatosis gene; *LTF* = lactotransferrin gene; N/O = not optimised; R = reverse primer; *SLC40A1* = solute carrier family 40 member A1 gene; T<sub>A</sub> = annealing temperature; T<sub>M</sub> = melting temperature

## Chapter 2: Detailed Experimental Procedures

**Table 2.2 Oligonucleotide primers and PCR conditions for amplification of genotyping fragments**

Gene Name	Variant	Primer Name	Primer Sequence (5' - 3')	Product Size (bp)	T <sub>M</sub> (°C)	T <sub>A</sub> (°C)	Mg <sup>2+</sup> (mM)	Betaine (%)	PCR amplification cycle	Enzyme used
<i>CYBRD1</i>	-1849 T/G, -1844 C/G, -1813 C/T, -1749 del (T) <sub>6</sub> G, -1540 G/A, -1477 G/A, -1459 T/C, -1346 T/C, -1272 T/C, -1011 T/A	BP1F Seq	AGCTCAAGTGATTCTCTTGCCT	972	56.3	75/68	1.5	0	A	
		BP1R Seq	TAGAGGCTAAACAAGCACTGGC		57.2					
		BP2F Seq	GGGTTTCAGCTGGTCAGCATC	855	58.5	61	2	0	B	
		BP2R Seq	ACTTCTGGGGCTGTCTCCG		56.2					
<i>HAMP</i>	-1631 C/T	PP1F Seq	CTGGTGACTTGGCTGACACT	153	57.2	55	1.5	0	E	
		PP R rs66868858	CTGCAGTGCCTTCATTCTTTC		56.6					
	-582 A/G	*PP3F	CACGCCTGGCTAAATTTGTT	326	54.0	55	1.5	0	E	
		*PP3R	CACCACACGTGCATAGGTTC		56.2					
		*PP5F	AAGTGAGTGGAGGAGAGCG		106					
*PP4R	CCATCACGATGTCATTCTGC	53.8								
-323 C/T	PP F rs116178041	CAGAATGACATCGTGATGGG	198	53.3	48	1.5	0	F		
-188 C/T	*PP5R	CTTTGCTCTGTCTCATTTC		51.2						

\*Primers previously designed. Abbreviations: 5' = 5-prime; 3' = 3-prime; °C = degrees Celsius; bp = base pairs; *CYBRD1* = cytochrome B reductase 1 gene; F = forward primer; *HAMP* = Hepcidin antimicrobial peptide gene; *HFE* = Haemochromatosis gene; *LTF* = lactotransferrin gene; R = reverse primer; *SLC40A1* = solute carrier family 40 member A1 gene; T<sub>A</sub> = annealing temperature; T<sub>M</sub> = melting temperature

## Chapter 2: Detailed Experimental Procedures

**Table 2.3 Oligonucleotide primers and PCR conditions for amplification of genotyping fragments**

Gene Name	Variant	Primer Name	Primer Sequence (5' - 3')	Product Size (bp)	T <sub>M</sub> (°C)	T <sub>A</sub> (°C)	Mg <sup>2+</sup> (mM)	Betaine (%)	PCR amplification cycle	Enzyme used
<i>HFE</i>	-1168 A/G	HP1F HP1R	GGTTGGAGCAAATCAGGTG GTGCTGAGTTCACCTTCGCAG	335	54.3 56.4	57	1.5	0	E	
	-561 A/G	*HP5F HP R novel -561	TTAGTGACAGCCTTTCCGCT AGTATTGGGATTATCTGGAGAATGA	145	56.5 53.3	48	1.5	0	F	
	-467 G/C	HP F rs2794720 HP R rs2794720	GCCTCAATATGTTAATCATCTCC ATCTCAACCTTAGACCAACTTATG	213	52.1 52.3	55	1.5	0	E	
	-331 A/C	HP F novel -331 *HP6R	GACATAAGTTGGTCTAAGGTTGAG GGGAAAGTAGCTTCGCAATG	115	53.1 54.1	48	1.5	0	F	
	-1377 T/G	LP1F LP1R	GCCAAGAAAGGTGTTTCAGGT GTGGCAGTGAGCCATGAT	153	55.5 55.3	62	1.5	0	E	
<i>LTF</i>	-746 del AAA/ ins AA	LP2F LP2R	ACCATTTCTACTGAATACTATCAACAC CCACCCAAGGCTGGTTC	150	54.7 56.3	60	1.5	0	E	
	-457 T/C	LP3F LP3R	GGGCAGGGAGATAACCCAT GCCAGTTAGGGCAGGTG	163	59.0 56.0	55	1.5	0	E	
	-437 C/G	LP4F LP2R Seq	CACCTGCCCTAACTGGC TAGTCTGCAAGCCCCTAGGAG	182	56.0 58.1	55	1.5	0	E	
	-1470 C/T, -1461 T/C, -1399 G/A, -1355 G/C, -1098 G/A	SP1F SP3R	GTAGACCTTTGGGGCTCCTG ATGCCACAGAGGCCGCTTTC	738	57.5 60.8	55	1.5	12	B	
	-750 G/A	SP4F SP4R	GAAGCCCTGCTATGCAGTCC GTCAGGTGCTGGAAGAAAGC	255	58.3 56.3	57	1.5	0	E	
<i>SLC40A1</i>	-662 C/T	SP F rs12693542 SP R rs12693542	CAGCACCTGACGCTTAGTTT TTTCTGCTCTCCGGGAA	130	55.4 56.0	N/O	N/O	N/O	N/O	N/O
	-501 T/C	SP6AF SP5R	TCCCGGAGAGCAGGAAAAC CCGTCCCCGCGCCCTTGC	135	57.2 68.0	N/O	N/O	N/O	N/O	N/O

\*Primers previously designed. Abbreviations: 5' = 5-prime; 3' = 3-prime; °C = degrees Celsius; bp = base pairs; *CYBRD1* = cytochrome B reductase 1 gene; F = forward primer; *HAMP* = Hepcidin antimicrobial peptide gene; *HFE* = Haemochromatosis gene; *LTF* = lactotransferrin gene; N/O = not optimised; R = reverse primer; *SLC40A1* = solute carrier family 40 member A1 gene; T<sub>A</sub> = annealing temperature; T<sub>M</sub> = melting temperature

Key to Tables 2.1 to 2.3

**Yellow** = *Pfu* enzyme reaction

**Green** = *Taq* enzyme from Thermo Fisher Scientific reaction

**Red** = *Taq* enzyme from KAPABiosystems reaction

**Orange** = Primers used for semi-automated bi-directional sequencing but not in PCR amplification

**Purple** = Fragments intended to be genotyped through semi-automated bi-directional sequencing analysis

**Blue** = Fragments intended to be genotyped through RFLP analysis

**Grey** = Fragments intended to be genotyped through HEX-SSCP analysis

## 2.2.2 Amplification of Fragments

### 2.2.2.1 Amplification of PCR Fragments

As already mentioned, fifteen randomly selected samples were subjected to semi-automated bi-directional sequencing analysis. Reaction conditions for each promoter region were optimised and the PCR conditions used are stipulated in Table 2.1. An Applied Biosystems PCR thermocycler (Applied Biosystems model 2720 Thermal Cycler) was used to perform all PCR reactions. To confirm the presence of the target DNA, all PCR products were electrophoresed on a 0.8% (w/v) to 3% (w/v) horizontal agarose gel at 120 volts (see section 2.2.3). The first fragment of *HFE* could not be optimised and a second set of primers (HP1F and HP1R as seen in Table 2.3) was used to amplify a smaller region to determine whether any variants were present. This fragment was analysed using HEX-SSCP analysis resulting in the discovery of a variant (-1168 A/G). The remaining population was thus genotyped for this variant.

Table 2.1 shows the conditions for the initial screening of the population for variants. Upon genotyping these 15 samples, 40 variants were identified. The fragments chosen to genotype these variants and their optimised PCR conditions are shown in Tables 2.2 and 2.3. The remaining patient and control samples for the *CYBRD1* promoter were amplified in the same manner as the original subset of fifteen and subsequently subjected to semi-automated bi-directional sequencing analysis. A region of the *SLC40A1* promoter containing five variants was amplified and also subjected to semi-automated bi-directional sequencing analysis.

Three fragments were amplified for subsequent RFLP analysis to genotype three separate variants (Tables 2.2 and 2.3). Twelve fragments were amplified in order to be genotyped by HEX-SSCP analysis so as to genotype 12 separate variants (Tables 2.2 and 2.3). Two variants in the *SLC40A1* promoter could not be optimised and were excluded from further analyses.

### 2.2.2.2 PCR Reaction Reagents and Amplification Cycles

The PCR reagents consisted of either a *Pfu* enzyme reaction, a *Taq* enzyme reaction from Thermo Fisher Scientific or a high fidelity *Taq* enzyme reaction from KAPABiosystems

---

depending on the fragment amplified. Each 25  $\mu$ l *Pfu* enzyme reaction contained 50 ng of DNA, 1 X *Pfu*-buffer [20 mM Tris-HCL, 10 mM ammonium sulphate ((NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>), 10 mM KCl, 0.1% (v/v) Triton X-100, 0.1 mg/ml bovine serum albumin (BSA)] (Thermo Fisher Scientific Inc), 10 pmol of each primer (Integrated DNA Technologies), magnesium sulphate (MgSO<sub>4</sub>) (indicated in Tables 2.1 to 2.3) (Thermo Fisher Scientific Inc), betaine solution (indicated in Tables 2.1 to 2.3) (Sigma-Aldrich), 0.1 mM of each 2'-deoxynucleotide (dNTP) (dATP, dCTP, dGTP, dTTP) (Thermo Fisher Scientific Inc), 0.25 units (U) *Taq* polymerase (Thermo Fisher Scientific Inc) and 0.25 U *Pfu* polymerase (Thermo Fisher Scientific Inc).

Each 25  $\mu$ l *Taq* enzyme reaction from Thermo Fisher Scientific contained 50 ng of DNA, 1 X *Taq*-buffer [75 mM Tris-HCL, 20 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.01% (v/v) Tween 20] (Thermo Fisher Scientific Inc), 10 pmol of each primer (Integrated DNA Technologies), magnesium chloride (MgCl<sub>2</sub>) (indicated in Tables 2.1 to 2.3) (Thermo Fisher Scientific Inc), betaine solution (indicated in Tables 2.1 to 2.3) (Sigma-Aldrich), 0.1 mM of each dNTP (dATP, dCTP, dGTP, dTTP) (Thermo Fisher Scientific Inc) and 0.5 units (U) *Taq* polymerase (Thermo Fisher Scientific Inc).

Each 25  $\mu$ l *Taq* enzyme from KAPABiosystems reaction contained 50 ng of DNA, 1 X KAPA HiFi HotStart ReadyMix buffer (25 mM Mg<sup>2+</sup>) (KAPABiosystems), 10 pmol of each primer (Integrated DNA Technologies), betaine solution (indicated in Table 2.1, Table 2.2 and Table 2.3) (Sigma-Aldrich) and 0.5 units of high fidelity *Taq* enzyme (KAPABiosystems).

The cycles used for amplification are described below.

**Cycle A:** An initial denaturation step was performed at 95°C for 5 minutes, followed by 15 cycles of denaturation at 95°C for 30 seconds, annealing at the temperature (first T<sub>A</sub>) given in Table 2.1 for 45 seconds, and elongation at 72°C for 30 seconds. This was followed by 20 cycles of denaturation at 95°C for 30 seconds, annealing at the temperature (second T<sub>A</sub>) given in Table 2.1 for 45 seconds, and elongation at 72°C for 30 seconds. A final elongation step at 72°C for 10 minutes was performed and then the sample was stored at 4°C.

**Cycle B:** An initial denaturation step was performed at 95°C for 5 minutes. Following this, 35 cycles of denaturation at 95°C for 30 seconds, annealing at the temperature (T<sub>A</sub>) given in Table 2.1 for 1 minute and elongation at 72°C for 2 minutes were performed. Thereafter, a

final elongation step at 72°C for 10 minutes was performed and then the sample was stored at 4°C.

**Cycle C:** An initial denaturation step was performed at 95°C for 5 minutes. Thereafter, 35 cycles of denaturation at 95°C for 30 seconds, annealing at the temperature ( $T_A$ ) given in Table 2.1 for 30 seconds and elongation at 72°C for 2 minutes were performed. Following this, a final elongation step at 72°C for 10 minutes was performed and then the sample was stored at 4°C.

**Cycle D:** An initial denaturation step was performed at 95°C for 2 minutes. Thereafter, 35 cycles of denaturation at 98°C for 20 seconds, annealing at the temperature ( $T_A$ ) given in Table 2.1 for 15 seconds and elongation at 72°C for 2 minutes were performed. Thereafter, a final elongation step at 72°C for 5 minutes was performed and then the sample was stored at 4°C.

**Cycle E:** An initial denaturation step was performed at 95°C for 5 minutes. Thereafter, 35 cycles of denaturation at 95°C for 30 seconds, annealing at the temperature ( $T_A$ ) given in Table 2.1 for 45 seconds and elongation at 72°C for 30 seconds were performed. Following this, a final elongation step at 72°C for 10 minutes was performed and then the sample was stored at 4°C.

**Cycle F:** An initial denaturation step was performed at 95°C for 5 minutes. Thereafter, 40 cycles of denaturation at 95°C for 30 seconds, annealing at the temperature ( $T_A$ ) given in Table 2.1 for 45 seconds and elongation at 72°C for 30 seconds were performed. Following this, a final elongation step at 72°C for 10 minutes was performed and then the sample was stored at 4°C.

Cycle D is specific to the *Taq* enzyme from KAPA Biosystems. The specific enzymes used are indicated in Tables 2.1 to 2.3 (as described in the Key to the Tables).

### 2.2.3 Agarose Gel Electrophoresis

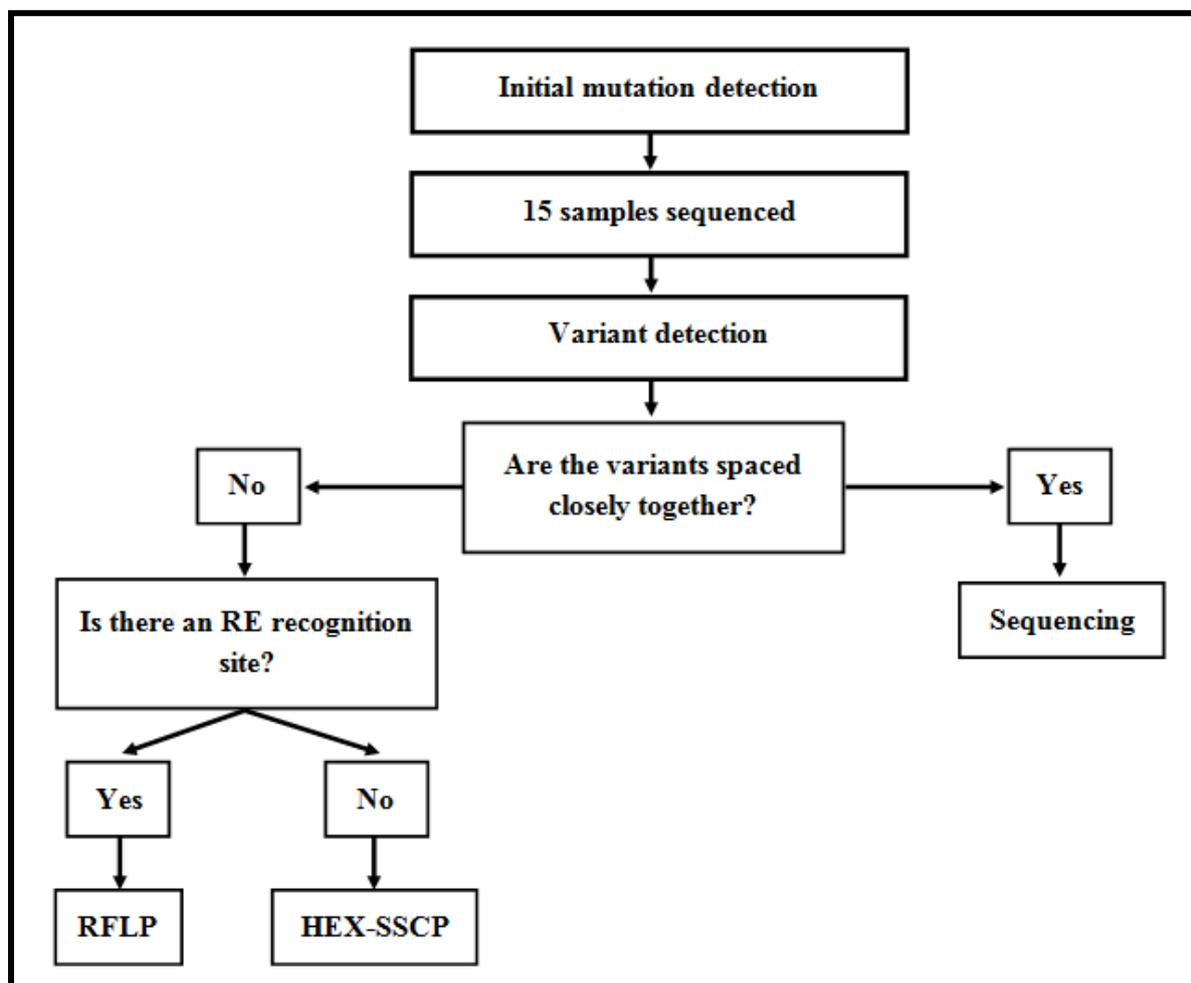
A 2% (w/v) agarose gel consisting of 1X TBE buffer [90 mM Tris ( $C_4H_{11}NO_3$ ), 90 mM boric acid ( $H_3BO_3$ ), 1mM EDTA ( $C_{10}H_{14}N_2Na_2O_8 \cdot 2H_2O$ )] and stained with 0.01% (v/v) ethidium bromide (EtBr) was used to visualise PCR fragments smaller than 1 000 bp. For PCR

fragments larger than 1 000 bp, a 0.8% (w/v) agarose gel was utilised. Five  $\mu$ l of cresol red loading buffer [0.02% (w/v) cresol red, 0.34% (w/v) sucrose] was loaded with 5  $\mu$ l of PCR sample in each well. A 3% (w/v) agarose gel was used to separate RFLP analysed samples. Ten  $\mu$ l of cresol red loading buffer was added directly to the RFLP sample from which 15  $\mu$ l of the sample-cresol red mixture was loaded in each well. The 100 bp O'GeneRuler (Thermo Fisher Scientific Inc) was used as a molecular ladder for comparison for fragments smaller than 1 000 bp. For fragments larger than 1 000 bp, the 1 kb GeneRuler (Thermo Fisher Scientific Inc) was used for size comparison. A sample of 2.5  $\mu$ l of the size standard was electrophoresed with each run of the PCR and RFLP products to determine the size of the products. The electrophoresis was performed for between 45 and 120 minutes at 120 V in 1X TBE buffer. Ultraviolet (UV) light transillumination was used to visualise the electrophoresed PCR products using the Multigenius Bio Imaging System (Syngene).

## 2.3 MUTATION DETECTION

### 2.3.1 Semi-automated bi-directional DNA sequencing analysis

In order to determine whether any known and/or novel variants exist in the promoter region of the five genes under investigation, it was necessary to undertake an exploratory screening of this region. Fifteen samples were randomly selected from the patient group and subjected to semi-automated bi-directional sequencing analysis. All sequencing was performed using the Applied Biosystems 3730xl DNA Analyzer. The resulting chromatograms were compared to the reference sequence for each gene obtained from the Ensembl website (Hubbard *et al.*, 2007; [www.ensembl.org](http://www.ensembl.org)). This comparison was performed using the program Bioedit Sequence Alignment Editor version 7.0.9.0 (Hall, 1999) and variants were subsequently identified.



**Figure 2.1** Work flow for the identification and subsequent genotyping of variants. A random subset of samples was initially submitted to semi-automated bi-directional sequencing analysis to perform an exploratory screen for known and/or novel variants. The results from the exploratory screen were used to design mutation detection analyses based on the distance between variants, and then whether the variants were situated within an RE recognition site.

Sequencing data were available for the control group used for comparison to the *CYBRDI* promoter region but a second program, FinchTV version 1.4.0 ([www.geospize.com](http://www.geospize.com)), had to be used to visualise the chromatograms. The screened region for each gene, all primers used and the identified variants are provided in Addendum A.

Once the presence of variants was confirmed they were named and subsequently verified using the dbSNP (Sherry *et al.*, 2011; <http://www.ncbi.nlm.nih.gov/projects/SNP>) database to determine whether the variant was known and/or novel. The HapMap (The International HapMap Consortium, 2003; [www.hapmap.org](http://www.hapmap.org)) database was utilised to determine genotype and allele frequencies identified within different regions and was thus used as a comparison to genotype and allele frequencies found in this study. In instances where HapMap either did

not recognise a variant or did not have frequency data available for a variant, the dbSNP population diversity data was used for comparison. Based on the variants discovered using this method, the remaining patient and control samples were genotyped using either RFLP, HEX-SSCP or semi-automated bi-directional sequencing analysis (as indicated in Tables 2.2 and 2.3).

The *CYBRD1* and *SLC40A1* genes presented either a large number of variants or variants that are in close proximity to one another. This makes genotyping through RFLP or HEX-SSCP analysis difficult and so the remaining samples were genotyped by subjecting the entire promoter of *CYBRD1* (encompassing 20 variants) and a portion of the promoter of *SLC40A1* (encompassing 5 variants) to semi-automated bi-directional sequencing analysis. The variants of the *SLC40A1* promoter not encompassed by the sequencing reaction were genotyped by HEX-SSCP analysis in the remaining patient and control samples. Chromatograms of all variants are provided in Addendum B.

### 2.3.2 Restriction Fragment Length Polymorphism (RFLP) Analysis

The tool NEBcutter version 2.0 (Vincze *et al.*, 2003; [tools.neb.com/NEBcutter2/](http://tools.neb.com/NEBcutter2/)) was used to determine if any of the identified variants were located within a restriction endonuclease (RE) recognition site and thus whether the variant could be genotyped for the remaining samples using RFLP analysis. Three variants (-582 A/G [*HAMP*]; -561 A/G, -467 G/C [*HFE*]) were situated within RE sites and were subsequently subjected to RFLP analysis. Variant -582 A/G in the promoter region of the *HAMP* gene contained a recognition site for the enzyme BstUI and the enzyme would only digest the DNA fragment in the presence of the variant. Each 20 µl reaction consisted of 10 µl of PCR product (previously amplified under the conditions as described in Table 2.2), 1 X Green buffer (50 mM potassium acetate [KCH<sub>3</sub>CO<sub>2</sub>], 20 mM tris-acetate [C<sub>4</sub>H<sub>11</sub>NO<sub>3</sub>], 10 mM magnesium acetate [Mg(CH<sub>3</sub>CO<sub>2</sub>)<sub>2</sub>], 100 µg/ml BSA, pH 7.9 at 25°C) (New England Biolabs Inc.) and 2 units of BstUI (New England Biolabs Inc.). The reaction mixture was digested at 60°C for 16 hours following which it was electrophoresed on a 3% (w/v) agarose gel as described in section 2.2.3.

Variant -561 A/G in the promoter region of the *HFE* gene contained a recognition site for the enzyme BsrBI and the enzyme would only digest the DNA fragment in the presence of the

variant. Each 20 µl reaction consisted of 10 µl of PCR product (previously amplified under the conditions as described in Table 2.2), 1 X Green buffer (50 mM KCH<sub>3</sub>CO<sub>2</sub>, 20 mM C<sub>4</sub>H<sub>11</sub>NO<sub>3</sub>, 10 mM Mg(CH<sub>3</sub>CO<sub>2</sub>)<sub>2</sub>, 100 µg/ml BSA, pH 7.9 at 25°C) (New England Biolabs Inc.) and 4 units of BsrBI (New England Biolabs Inc.). The reaction mixture was digested at 37°C for 16 hours and then the enzyme was heat inactivated at 80°C for 20 minutes. The reaction was electrophoresed on a 3% (w/v) agarose gel as described in section 2.2.3.

Variant -467 G/C in the promoter region of the *HFE* gene contained a recognition site for the enzyme AflIII and the enzyme is expected to only digest the DNA fragment in the presence of the variant. Each 20 µl reaction consisted of 10 µl of PCR product (previously amplified under the conditions as described in Table 2.2), 1 X Green buffer (100 mM NaCl, 50 mM tris-HCL, 10 mM MgCl<sub>2</sub>, 100 µg/ml BSA, pH 7.9 at 25°C) (New England Biolabs Inc.), 1% BSA and 2 units of AflIII (New England Biolabs Inc.). The reaction mixture was digested at 37°C for 16 hours and then the enzyme was heat inactivated at 80°C for 20 minutes. The reaction was electrophoresed on a 3% (w/v) agarose gel as described in section 2.2.3. Gel images of RFLP analyses are provided in Addendum B.

### **2.3.3 Heteroduplex Single-Strand Conformation Polymorphism (HEX-SSCP) Analysis**

PCR products that were not genotyped using semi-automated bi-directional sequencing analysis were electrophoresed on HEX-SSCP gels using the technique described by Kotze *et al.* (1995). All gels were electrophoresed using the Hoefer vertical gel apparatus. Each PAA gel was supplemented with urea and consisted of 12% (w/v) PAA [1% C of 40% stock (99 acrylamide (AA): 1 bisacrylamide (BAA))], 7.5% (w/v) urea, 1.5 X TBE buffer, 0.1% (w/v) ammonium persulphate (APS), and 0.1% (v/v) tertamethylethylenediamine (TEMED).

Prior to gel loading, 15 µl of loading dye [95% (v/v) formamide, 20mM EDTA, 0.05% (v/v) bromophenol blue, 0.05% (v/v) Xylene cyanol] was added to each PCR product and this solution was denatured at 95°C for 10 minutes and subsequently placed directly onto ice. 20 µl of each denatured product was loaded into a separate well of a gel using a Microlitre™ syringe (Hamilton). All loaded gels were subjected to electrophoresis at 250 volts for ± 18 hours (overnight). PAA gels supplemented with urea were electrophoresed at 4°C while PAA gels supplemented with glycerol were electrophoresed at room temperature. Samples

that had been previously genotyped via semi-automated bi-directional sequencing analysis were electrophoresed alongside the genotyping samples in order to provide comparisons. Any samples that could not be identified through comparison to the sequenced sample were subjected to semi-automated bi-directional sequencing analysis to confirm the presence of variants. All gels were visualised by being subjected to UV transillumination using the MULTI-GENIUS Bio-Imaging System (Syngene).

Gel images of HEX-SSCP analyses are provided in Addendum B.

## **2.5 *IN SILICO* ANALYSES**

*In silico* analyses were performed on variants identified in the promoter region of the five iron genes under investigation. This was done in order to determine if these variants had any effect on putative TFBSs that occurred in these regions. This allows one to determine whether any putative TFBSs are abolished or created due to a specific variant and thus hypothesise the effect of the specific variant(s) on gene expression.

The programs ALIBABA2 (version 2.1), TFblast (version 0.1) and PROMO (version 1.0) from TRANSFAC®7 were used to perform *in silico* analyses (Wingender *et al.*, 2001; <http://gene-regulation.com/pub/database.html#transfac>). These programs screen an input sequence for putative TFBSs found in the TRANSFAC®7 database. Both a graphical and text output are created giving the position and type of TFBS throughout the input sequence. Comparing the outputs from the wild type sequence to those of each variant allows for the determination of any changes to the TFBS composition in the sequence due to the presence or absence of the variant.

## **2.6 STATISTICAL ANALYSIS**

### **2.6.1 Iron parameters**

Descriptive statistics were performed on the iron parameter data using Excel (2007). For normally distributed data parametric tests were performed as follows: F-tests were performed

on the data and if a  $P$ -value of less than 0.05 was returned, the data were subsequently subjected to an unpaired  $t$ -test assuming unequal variances. If the F-test returned a  $P$ -value of greater than 0.05 the data were subsequently subjected to an unpaired  $t$ -test assuming equal variances. The F- and  $t$ -tests were performed for each iron parameter to determine the variance between the patient and control groups. For non-normally distributed data a non-parametric test, the Mann-Whitney test, was performed to compare the data. The F-tests, unpaired  $t$ -tests and Mann-Whitney tests were performed using the GraphPad Prism™ statistical package software (version 5.00) (GraphPad Software, San Diego, CA) ([www.graphpad.com](http://www.graphpad.com)). The Pearson  $r$  correlation test was performed on GraphPad Prism™ with graph outputs being created in Excel to compare each iron parameter to another to determine if any relationship exists between them. These correlations were performed for the entire cohort, as well as within the patient and control groups.

For each variant, the data was further divided into two groups; 1) individuals without the presence of the variant (wild type individuals), and, 2) individuals with the presence of the variant (individuals that were either heterozygous or homozygous for the variant). The iron parameters were compared for these new groupings to determine whether the presence of a variant had any effect on a specific iron parameter. Parametric and non-parametric tests were performed on this data as described above.

### **2.6.2 Genotyping**

Version 1.3 of the program Tools for Population Genetics Analysis (TFPGA) [Miller, 1997] was used to determine the genotype and allele frequencies for all variants discovered and this program performed the chi-squared and Fisher's exact tests to determine whether the data for each group are in Hardy-Weinberg Equilibrium (HWE). The chi-squared test was used in instances where all genotypes had an observed number of more than 5. Fisher's exact test was used in cases where a genotype had less than, or equal to, 5 observed individuals. The Exact test was also used to determine whether the allele frequencies differed statistically significantly between the patient and control groups for each identified variant. TFPGA does not offer an Exact test to compare genotype frequencies and so the program Genepop (version 4.2.1) was used (Rousset, 2008).

### **2.6.3 TB association**

Correlation studies were performed in order to determine whether any relationship exists between the iron parameters measured, genotype and susceptibility to *M. tuberculosis* infection.

## **Chapter 3**

### **RESULTS AND DISCUSSION**

---

## 3. RESULTS & DISCUSSION

### 3.1 ABSTRACT

Tuberculosis (TB) is a serious epidemic in South Africa and many other countries worldwide. Due to the increasing need for more effective treatment regimes and preventative measures the necessity for further study of the interaction between *M. tuberculosis* and the host system has become of paramount importance. Iron metabolism plays a significant role in the development of TB as the causative agent, *Mycobacterium tuberculosis*, requires iron from the host reticuloendothelial macrophages in order to proliferate. It is for this reason that the regulation of iron in the body is crucial and thus the regulation of the genes that regulate the iron metabolism pathway is of paramount importance. One of the mechanisms of controlling the expression of these genes is through transcriptional regulation and therefore the promoter region of genes have been targeted. Five genes were chosen – *CYBRD1*, *HAMP*, *HFE*, *LTF* and *SLC40A1* – for investigation as to the role they play in *M. tuberculosis* susceptibility.

The study cohort consisted of 49 TB patients and 51 healthy, unrelated, randomly selected controls all of whom were Black, Xhosa-speaking individuals. The control group was further divided into two groups: the first group consisted of 22 individuals which were used for comparison to the promoter region of the *CYBRD1* gene, which did not have iron parameter data available; and a second group containing 29 individuals used for analyses of the four remaining genes, which had iron parameter data available. The promoter region of these genes were screened for both known and/or novel variants. In order to assess whether any variants exist within this region, 15 randomly selected patient samples were screened using Polymerase Chain Reaction (PCR) amplification followed by semi-automated bi-directional sequencing. From the initial screening of the first 15 sequenced samples a total of 40 variants were found in the promoter region of the 5 genes under investigation. Of these 40 variants 30 were found to be previously described while 10 variants were novel. Due to technical difficulties, only 31 variants were genotyped for the remainder of the patient and control groups.

Of the 31 variants subjected to mutation detection analyses, 11 variants (-1849 T/G, -1844 C/G, -1813 C/T, -1749 del(T)<sub>6</sub>G, -1477 G/A, -1459 T/C, -1346 T/C, -1272 T/C, -238 A/G, -166 C/G [*CYBRD1*]; -561 A/G [*HFE*]) demonstrated a statistically significant difference

---

when comparing allele frequencies between the patient and control groups. Ten of these variants (-1849 T/G, -1813 C/T, -1749 del(T)<sub>6</sub>G, -1477 G/A, -1459 T/C, -1346 T/C, -1272 T/C, -238 A/G, -166 C/G [*CYBRDI*]; -561 A/G [*HFE*]) demonstrated a statistically significant difference when comparing genotype frequencies between the two groups. The influence of a variant on the iron parameters was determined through the use of the unpaired *t*-test for those genes that had iron parameter data available (*HAMP*, *HFE*, *SLC40A1*). The results showed that a statistically significant difference exists for 3 (-1470 C/T [*HFE*]; -1355 G/C, -1098 G/A [*SLC40A1*]) of the 11 variants compared. A Pearson *r* correlation test was performed to determine if any correlations exist between iron parameters within the cohort as a whole, and within each group separately. A strong positive correlation was demonstrated with serum iron and transferrin saturation levels within both the groups separately as well as when compared in the entire cohort. Interestingly, a statistically significant negative correlation was observed between serum ferritin and serum transferrin levels in the patient group exclusively which could suggest the possible influence of these proteins and their encoding genes on *M. tuberculosis* susceptibility.

*In silico* analyses revealed a number of putative transcription factor binding sites (TFBSs) in the promoter region of the five genes under investigation. It was predicted that the presence of specific variants caused either the abolishment or creation of a total of 41 TFBSs. The elimination or addition of TFBSs may potentially cause changes in the regulation of the affected gene.

The results from this study indicate the possible implication of some of the variants discovered in increasing iron absorption which may induce increased pathogenic infection. From the obtained results it is clear that further investigation is required as to the exact mechanism by which iron, the iron pathway, and the genes involved in this pathway are associated with *M. tuberculosis* susceptibility.

### 3.2 INTRODUCTION

One in every 3 individuals in the world is reported to be infected with *M. tuberculosis* but only 10% of these individuals will ever develop the active disease (Dye *et al.*, 1999; Herzog, 1998). The high prevalence of this low penetrating disease could possibly be attributed to a

---

number of factors influencing susceptibility. Some of the factors known to affect susceptibility in a minority of occurrences include: alcohol abuse, diabetes, corticosteroid use, HIV infection, and old age (Bellamy *et al.*, 2000). However, the remaining cases are affected by the intricate interaction between genetic and environmental factors which lead to the development of active TB. The influence of genetic factors in *M. tuberculosis* susceptibility has been implicated by the discovery of racial differences and through the use of twin studies (Bellamy *et al.*, 2000; Lipsitch & Sousa, 2002; Comstock, 1978).

*M. tuberculosis* requires a plethora of nutrients from its host during infection and of these, iron has been shown to be vital for its growth and proliferation (Cole *et al.*, 1998); this has necessitated the development of methods for outcompeting its host for iron (Olayanmi *et al.*, 2007). For this reason iron metabolism and the genes involved are gaining increasing popularity in research in terms of their possible connection to *M. tuberculosis* susceptibility.

Iron is not only important for bacterial proliferation but is crucial for the survival and proliferation of cells, where it acts as both an electron acceptor and donor in many important redox reactions (Ong *et al.*, 2006). The *CYBRD1* gene produces the protein DCYTB which functions as a ferrireductase reducing  $\text{Fe}^{3+}$  to  $\text{Fe}^{2+}$  in order for iron absorption to occur into duodenal enterocytes (McKie *et al.*, 2001). The *HAMP* gene encodes for the protein product hepcidin which has some antimicrobial activity but mainly functions by binding to FPN and subsequently causing its internalization and degradation therefore decreasing iron efflux from cells (Collins, 2008). The haemochromatosis protein, encoded by the *HFE* gene, regulates absorption of iron by binding to TFR thereby decreasing its affinity for TF (Nandar & Connor, 2011). This interaction leads to the decreased interaction of TF with the cell membrane which in turn leads to less iron being absorbed by the cell from TF. The *LTF* gene encodes for the protein lactotransferrin which is responsible for transportation of iron in the body (Olayanmi *et al.*, 2007). This protein is found mainly at mucosal membranes such as in the lungs and has been shown to have some antimicrobial activity (Orsi, 2004). The only known iron export protein to date is FPN which is encoded by the *SLC40A1* gene which regulates the efflux of iron from cells (Collins, 2008).

Since these genes and their products play an integral role in iron metabolism and potentially *M. tuberculosis* susceptibility it is important to understand the mechanisms by which they are regulated. The aim of this study is thus to elucidate the effect of 5 genes – *CYBRD1*, *HAMP*, *HFE*, *LTF* and *SLC40A1* – involved in iron metabolism on *M. tuberculosis* susceptibility.

---

### 3.3 MATERIALS AND METHODS

Ethical approval was granted for this study from the Research and Ethics/Biosafety Committee, Faculty of Health Sciences, University of Stellenbosch (N10/06/194). The reader is referred to Chapter 2 of this thesis for the detailed experimental procedures performed.

### 3.4 RESULTS

#### 3.4.1 Mutation Detection

##### 3.4.1.1 Variant List

The promoter region of 5 genes (*CYBRD1*, *HAMP*, *HFE*, *LTF* and *SLC40A1*) known to play a role in the metabolism of iron were screened in an initial sample subset as an exploratory venture to determine the presence of known and/or novel variants in this region. The sample subset consisted of 15 randomly selected patient samples. This initial screening yielded a total of 40 variants of which 10 were shown to be novel. Both the remainder of the patient group and the control group were subsequently screened for the variants through either semi-automated bi-directional sequencing, HEX-SSCP or RFLP analysis (as described in Tables 2.2 and 2.3).

If variants were closely spaced, e.g., the variants found in the promoter region of *CYBRD1* (Addendum A), then semi-automated bi-directional sequencing analysis was used as the method to detect genotypes. This method was employed for all variants discovered in the promoter region of the *CYBRD1* gene, and for 5 variants in the promoter region of *SLC40A1* (Tables 2.2 and 2.3). The chromatograms of all 40 variants are provided in Addendum B.

For those variants that were not as closely spaced – where a primer could be placed between variants for amplification of separate fragments containing only one variant – either HEX-SSCP or RFLP analysis was employed as the technique used to genotype the variant. Each variant was examined to determine whether it was suitable for RFLP analysis (whether it was situated in a RE recognition site), and if proven suitable it was genotyped using the enzyme most apt for its digestion (gel images are provided in Addendum B). In most cases, the

---

variants were not suitable for RFLP analysis and as such were genotyped using HEX-SSCP analysis.

In the promoter region of the genes investigated, the following variants were identified; *CYBRDI*: 16 previously described variants (-1849 T/G, -1844 C/G, -1813 C/T, -1749 del (T)<sub>6</sub>G, -1540 G/A, -1477 G/A, -1459 T/C, -1346 T/C, -1272 T/C, -1011 T/A, -627 T/C, -624 G/A, -449 C/G, -399 T/G, -238 A/G, -166 C/G) and four novel variants (-849 C/G, -492 A/G, -454 C/T, -397 A/C); *HAMP*: three previously described variants (-1631 C/T, -582 A/G, -188 C/T) and one novel variant (-323 C/T); *HFE*: two previously described variants (-1168 A/G, -467 G/C) and two novel variants (-561 A/G, -331 A/C); *LTF*: one previously described variant (-746 del [AAA]/ins [AA]) and three novel variants (-1377 T/G, -457 T/C, -437 C/G); and *SLC40A1*: eight previously described variants (-1470 C/T, -1461 T/C, -1399 G/A, -1355 G/A, -1098 G/A, -750 G/A, -662 C/T, -501 T/C). These variants are listed in Table 3.1 which also indicates which variants were excluded, the motivation for exclusion, allele frequency of the variant in both the patient and control groups, and the results of the HWE analysis.

Table 3.1 General information regarding variants, including HWE analysis

Gene	Nucleotide position and change	Reference sequence number	Included/ Excluded from further analyses	Motivation for exclusion	Variant allele frequency		HWE analyses			
					Patients	Controls	Patients		Controls	
							<i>n</i> carriers out of total	<i>P</i> -value	<i>n</i> carriers out of total	<i>P</i> -value
<i>CYBRD1</i>	-1849 T/G	rs12692964	Included		0.26	0.06	15/42	0.00	1/18	0.03
	-1844 C/G	rs7585974	Included		0.92	0.75	41/42	0.24	15/18	0.03
	-1813 C/T	rs12692965	Included		0.31	0.06	18/43	0.00	1/18	0.03
	-1749 del (T) <sub>6</sub> G	rs140246632	Included		0.69	0.39	39/43	1.00	9/18	0.04
	-1540 G/A	rs67339714	Included		0.26	0.11	19/43	1.00	2/18	0.00
	-1477 G/A	rs66572658	Included		0.31	0.06	11/43	0.21	1/18	0.03
	-1459 T/C	rs10199858	Included		0.70	0.39	39/43	1.00	7/18	0.00
	-1346 T/C	rs66818940	Included		0.35	0.06	21/43	0.01	1/18	0.03
	-1272 T/C	rs62183550	Included		0.20	0.44	15/43	0.65	9/18	0.00
	-1011 T/A	rs12621138	Included		0.24	0.26	13/39	0.00	9/21	0.58
	-849 C/G		Included		0.39	0.40	27/47	0.10	18/30	0.36
	-627 T/C	rs10200648	Included		0.09	0.05	6/47	0.02	3/30	1.00
	-624 G/A	rs884408	Included		0.12	0.18	9/47	0.10	8/30	0.03
	-492 A/G		Included		0.15	0.12	14/47	0.57	6/30	0.33
	-454 C/T		Included		0.04	0.00	2/47	0.00	0/30	-
	-449 C/G	rs112362682	Included		0.02	0.02	1/47	0.01	1/30	1.00
	-399 T/G	rs884409	Included		0.13	0.10	7/47	0.00	5/30	0.24
-397 A/C		Included	0.14	0.05	13/47	0.57	3/30	1.00		
-238 A/G	rs868106	Included	0.83	0.53	45/47	0.60	18/29	0.00		
-166 C/G	rs2356782	Included	0.64	0.23	28/43	0.00	8/28	0.00		
<i>HAMP</i>	-1631 C/T	rs66868858	Included		0.28	0.31	25/48	0.30	13/27	0.37
	-582 A/G	rs10421768	Included		0.18	0.16	17/49	1.00	8/29	0.52
	-323 C/T		Included		0.01	0.00	1/49	1.00	0/29	-
	-188 C/T	rs116178041	Excluded		Could not amplify all samples successfully					
<i>HFE</i>	-1168 A/G	rs62625311	Included		0.08	0.04	13/49	1.00	9/28	1.00
	-561 A/G		Included		0.39	0.18	34/45	0.00	8/22	1.00
	-467 G/C	rs2794720	Included		0.15	0.20	8/49	0.29	2/24	0.25
	-331 A/C		Excluded		Genotypes indistinguishable on HEX-SSCP gels					
<i>LTF</i>	-1377 T/G		Excluded							
	-746 delAAA/insAA	rs5848800	Excluded		Genotypes indistinguishable on HEX-SSCP gels					
	-457 T/C		Excluded		All samples were homozygous for the variant					
	-437 C/G		Excluded		Genotypes indistinguishable on HEX-SSCP gels					

Table 3.1 continued

Gene	Nucleotide position and change	Reference sequence number	Included/ Excluded from further analyses	Motivation for exclusion	Variant allele frequency		HWE analyses			
					Patients	Controls	Patients		Controls	
							<i>n</i> carriers out of total	<i>P</i> -value	<i>n</i> carriers out of total	<i>P</i> -value
<i>SLC40A1</i>	-1470 C/T	rs16831699	Included		0.06	0.05	6/49	1.00	1/21	1.00
	-1461 T/C	rs77227487	Included		0.18	0.12	14/49	0.04	3/21	0.01
	-1399 G/A	rs79516140	Included		0.05	0.14	3/49	0.00	3/21	0.00
	-1355 G/C	rs3811621	Included		0.52	0.64	30/49	0.19	18/21	1.00
	-1098 G/A	rs10202029	Included		0.27	0.12	24/49	0.47	5/21	1.00
	-750 G/A	rs13015236	Excluded	Genotypes indistinguishable on HEX-SSCP gels						
	-662 C/T	rs12693542	Excluded	Could not optimise genotyping fragment						
	-501 T/C	rs6728200	Excluded	Could not optimise genotyping fragment						

$P \leq 0.05$  considered to be statistically significant (highlighted in red). Abbreviations: - = no calculation performed (only wild type individuals in group); A = adenine; C = cytosine; *CYBRD1* = cytochrome B reductase 1 gene; del = deletion; G = guanine; *HAMP* = hepcidin antimicrobial peptide gene; *HFE* = Haemochromatosis gene; HWE = Hardy-Weinberg equilibrium; ins = insertion; *n* carriers out of total = number of individuals carrying the variant allele (either heterozygote or homozygote) out of the total number of individuals genotyped for that locus; *P* = probability; rs = reference sequence; *SLC40A1* = solute carrier family 40 member A1 gene; T = thymine

---

As shown in Table 3.1, nine of the 40 variants were excluded from further analysis. The promoter region of the *LTF* gene proved difficult to amplify. The close proximity of the variants warranted genotyping by semi-automated bi-directional sequencing as indicated in section 2.3.1 but a larger fragment encompassing all four variants proved impossible to amplify for all samples. It was then decided to amplify smaller fragments spanning one variant each in order to genotype the variants using conventional methods such as RFLP and HEX-SSCP analysis. However, the genotyping was complicated by the inability to distinguish individual genotypes by these conventional methods. The variant -746 delAAA/insAA proved particularly difficult to genotype since the combination of both a deletion and insertion at the same site produced a number of varying genotypes which were indistinguishable on HEX-SSCP gels. For these reasons the promoter region of the *LTF* gene was excluded from further analyses. The novel variant -457 T/C produced only homozygote individuals, which while interesting, does not allow for analysis through the means employed by this study.

A number of other variants proved difficult to amplify and genotype and were subsequently also excluded from further analyses. These promoter region variants include: rs116178041 in the *HAMP* gene; the novel variant -331 A/C in the *HFE* gene; and variants rs13015236, rs12693542 and rs6728200 in the *SLC40A1* gene (as described in Table 3.1).

#### **3.4.1.2 HWE Analysis**

HWE analyses were performed in TFPGA using the default settings for the exact and the chi-squared tests. Table 3.1 shows the probability (*P*) values obtained. All statistically significant results ( $P < 0.05$ ) have been highlighted in red. For *CYBRD1* the results indicate that 8 of the 20 variants are not in HWE for the patient group; and 12 of the 20 variants are not in HWE for the control group. For *HAMP* all variants were in HWE except for the variant -323 C/T in the control group where the calculation could not be performed as only wild type individuals were seen. For *HFE* the variant -561 A/G was the only variant to deviate from HWE and this deviation was seen for the patient group but not in the control group. For *SLC40A1* the variants -1461 T/C and -1399 G/A were found to deviate from HWE in both the patient and control groups.

### 3.4.1.3 Genotype and Allele Frequency Analysis

Table 3.2 shows the calculated genotype and allele frequencies for all genotyped variants as well as the results of the Exact test. The Exact test was used to determine whether there was a statistically significant difference in allele and genotype frequencies when comparing the patient to the control group for every identified variant. The Exact test for genotypic population differentiation was performed using Genepop whereas the allelic population differentiation was determined using TFGPA. The default settings for the Exact test were used in both programs.

Of the 31 variants subjected to the Exact test for both allele and genotype population differentiation, 11 showed statistically significant differences in allele frequencies between the patient and control groups, and 10 of these variants showed statistically significant differences in genotype frequencies as well. Some variants were shown to be marginally statistically significant when comparing the genotype and allele frequencies. However, for the purpose of this study only those variants which showed a statistically significant difference were subjected to *in silico* analyses to determine whether any putative TFBSs were created or abolished by the presence or absence of these variants.

Table 3.2 Genotype and allele frequencies for variants identified

<i>CYBRD1</i> (ENSG0000007196)											
Nucleotide position and change	Ancestral allele (dbSNP)	Group	n	Genotype Frequency			P-value (Genepop)	2n	Allele Frequency		P-value (TFPGA)
				TT	TG	GG		T	G		
<b>-1849 T/G</b> rs12692964	T	Patients	42	0.64	0.19	0.17	<b>0.03</b>	84	0.74	0.26	<b>0.01</b>
		Controls	18	0.94	0.00	0.06		36	0.94	0.06	
<b>-1844 C/G</b> rs7585974	G	Patients	42	0.02	0.12	0.86	<b>0.08</b>	84	0.08	0.92	<b>0.02</b>
		Controls	18	0.17	0.17	0.66		36	0.25	0.75	
<b>-1813 C/T</b> rs12692965	C	Patients	43	0.58	0.21	0.21	<b>0.01</b>	86	0.69	0.31	<b>0.01</b>
		Controls	18	0.94	0.00	0.06		36	0.94	0.06	
<b>-1749 del (T)<sub>6</sub>G</b> rs104246632	NID	Patients	43	0.09	0.44	0.47	<b>0.01</b>	86	0.31	del(T) <sub>6</sub> G 0.69	<b>0.00</b>
		Controls	18	0.50	0.22	0.28		36	0.61	0.39	
<b>-1540 G/A</b> rs67339714	G	Patients	43	0.56	0.37	0.07	<b>0.13</b>	86	0.74	0.26	<b>0.09</b>
		Controls	18	0.89	0.00	0.11		36	0.89	0.11	
<b>-1477 G/A</b> rs66572658	G	Patients	43	0.51	0.35	0.14	<b>0.01</b>	86	0.69	0.31	<b>0.00</b>
		Controls	18	0.94	0.00	0.06		36	0.94	0.06	
<b>-1459 T/C</b> rs10199858	C	Patients	43	0.09	0.42	0.49	<b>0.01</b>	86	0.30	0.70	<b>0.00</b>
		Controls	18	0.61	0.00	0.39		36	0.61	0.39	
<b>-1346 T/C</b> rs66818940	C	Patients	43	0.51	0.28	0.21	<b>0.00</b>	86	0.65	0.35	<b>0.00</b>
		Controls	18	0.94	0.00	0.06		36	0.94	0.06	
<b>-1272 T/C</b> rs62183550	T	Patients	43	0.65	0.30	0.05	<b>0.02</b>	86	0.80	0.20	<b>0.01</b>
		Controls	18	0.50	0.11	0.39		36	0.56	0.44	

Table 3.2 continued

<i>CYBRD1</i> (ENSG0000007196)											
Nucleotide position and change	Ancestral allele (dbSNP)	Group	n	Genotype Frequency			P-value (Genepop)	2n	Allele Frequency		P-value (TFPGA)
				TT	TA	AA		T	A		
<b>-1011 T/A</b> rs12621138	T	Patients	39	0.67	0.18	0.15	<b>1.00</b>	78	0.76	0.24	<b>0.83</b>
		Controls	21	0.57	0.33	0.10		42	0.74	0.26	
<b>-849 C/G</b> This study	NID	Patients	47	0.43	0.36	0.21	<b>1.00</b>	94	0.61	0.39	<b>1.00</b>
		Controls	30	0.40	0.40	0.20		60	0.60	0.40	
<b>-627 T/C</b> rs10200648	T	Patients	47	0.87	0.09	0.04	<b>0.59</b>	94	0.91	0.09	<b>0.53</b>
		Controls	30	0.90	0.10	<b>0.00</b>		60	0.95	0.05	
<b>-624 G/A</b> rs884408	G	Patients	47	0.81	0.15	0.04	<b>0.42</b>	94	0.88	0.12	<b>0.35</b>
		Controls	30	0.73	0.17	0.10		60	0.82	0.18	
<b>-492 A/G</b> This study	NID	Patients	47	0.70	0.30	<b>0.00</b>	<b>0.63</b>	94	0.85	0.15	<b>0.63</b>
		Controls	30	0.80	0.17	0.03		60	0.88	0.12	
<b>-454 C/T</b> This study	NID	Patients	47	0.96	<b>0.00</b>	0.04	<b>0.52</b>	94	0.96	0.04	<b>0.16</b>
		Controls	30	<b>1.00</b>	<b>0.00</b>	<b>0.00</b>		60	<b>1.00</b>	<b>0.00</b>	
<b>-449 C/G</b> rs112362682	C	Patients	47	0.98	<b>0.00</b>	0.02	<b>1.00</b>	94	0.98	0.02	<b>1.00</b>
		Controls	30	0.97	0.03	<b>0.00</b>		60	0.98	0.02	
<b>-399 T/G</b> rs884409	T	Patients	47	0.85	0.04	0.11	<b>0.70</b>	94	0.87	0.13	<b>0.80</b>
		Controls	30	0.83	0.13	0.04		60	0.90	0.10	
<b>-397 A/C</b> This study	NID	Patients	47	0.72	0.28	<b>0.00</b>	<b>0.08</b>	94	0.86	0.14	<b>0.11</b>
		Controls	30	0.90	0.10	<b>0.00</b>		60	0.95	0.05	
<b>-238 A/G</b> rs868106	G	Patients	47	0.04	0.26	0.70	<b>0.00</b>	94	0.17	0.83	<b>0.00</b>
		Controls	29	0.38	0.17	0.45		58	0.47	0.53	

Table 3.2 continued

<i>CYBRD1</i> (ENSG0000007196)											
Nucleotide position and change	Group	n	Genotype Frequency			P-value (Genepop)	2n	Allele Frequency		P-value (TFPGA)	
<b>-166 C/G</b> rs2356782	C	Patients	43	CC	CG	GG	<b>0.00</b>	86	C	G	<b>0.00</b>
		Controls	28	0.35	0.02	0.63		56	0.36	0.64	
				0.71	0.11	0.18		0.77	0.23		

<i>HAMP</i> (ENSG00000105697)											
Nucleotide position and change	Ancestral allele (dbSNP)	Group	n	Genotype Frequency			P-value (Genepop)	2n	Allele Frequency		P-value (TFPGA)
<b>-1631 C/T</b> rs66868858	C	Patients	48	CC	CT	TT	<b>0.70</b>	96	C	T	<b>0.71</b>
		Controls	27	0.48	0.48	0.04		54	0.72	0.28	
				0.52	0.33	0.15		0.69	0.31		
<b>-582 A/G</b> rs10421768	G	Patients	49	AA	AG	GG	<b>0.67</b>	98	A	G	<b>0.83</b>
		Controls	29	0.65	0.33	0.02		58	0.82	0.18	
				0.72	0.24	0.04		0.84	0.16		
<b>-323 C/T</b> This study	NID	Patients	49	CC	CT	TT	<b>1.00</b>	98	C	T	<b>1.00</b>
		Controls	29	0.98	0.02	0.00		58	0.99	0.01	
				1.00	0.00	0.00		1.00	0.00		

<i>HFE</i> (ENSG00000010704)											
Nucleotide position and change	Ancestral allele (dbSNP)	Group	n	Genotype Frequency			P-value (Genepop)	2n	Allele Frequency		P-value (TFPGA)
<b>-1168 A/G</b> rs62625311	A	Patients	49	AA	AG	GG	<b>0.50</b>	98	A	G	<b>0.51</b>
		Controls	28	0.74	0.22	0.04		56	0.92	0.08	
				0.68	0.25	0.07		0.96	0.04		
<b>-561 A/G</b> This study	NID	Patients	45	AA	AG	GG	<b>0.00</b>	90	A	G	<b>0.02</b>
		Controls	22	0.24	0.73	0.02		44	0.61	0.39	
				0.64	0.36	0.00		0.82	0.18		
<b>-467 G/C</b> rs2794720	C	Patients	49	GG	GA	AA	<b>0.54</b>	98	G	A	<b>0.51</b>
		Controls	24	0.84	0.16	0.00		48	0.85	0.15	
				0.92	0.08	0.00		0.80	0.20		

Table 3.2 continued

<i>SLC40A1</i> (ENSG00000138449)											
Nucleotide position and change	Ancestral allele (dbSNP)	Group	n	Genotype Frequency			P-value (Genpop)	2n	Allele Frequency		P-value (TFPGA)
				CC	CT	TT			C	T	
<b>-1470 C/T</b> rs16831699	G	Patients	49	0.88	0.12	0.00	<b>0.43</b>	98	0.94	0.06	<b>0.68</b>
		Controls	21	0.95	0.05	0.00		42	0.98	0.02	
<b>-1461 T/C</b> rs77227487	T	Patients	49	0.71	0.21	0.08	<b>0.54</b>	98	0.82	0.18	<b>0.44</b>
		Controls	21	0.86	0.05	0.09		42	0.88	0.12	
<b>-1399 G/A</b> rs79516140	C	Patients	49	0.94	0.02	0.04	<b>0.27</b>	98	0.95	0.05	<b>0.08</b>
		Controls	21	0.86	0.00	0.14		42	0.86	0.14	
<b>-1355 G/C</b> rs3811621	C	Patients	49	0.18	0.59	0.23	<b>0.18</b>	98	0.48	0.52	<b>0.21</b>
		Controls	21	0.14	0.43	0.43		42	0.36	0.64	
<b>-1098 G/A</b> rs10202029	T	Patients	49	0.51	0.45	0.04	<b>0.06</b>	98	0.73	0.27	<b>0.08</b>
		Controls	21	0.76	0.24	0.00		42	0.88	0.12	

Abbreviations: 2n = number of alleles; A = adenine; C = cytosine; *CYBRD1* = cytochrome B reductase 1 gene; del = deletion; G = guanine; *HAMP* = hepcidin antimicrobial protein gene; *HFE* = haemochromatosis protein; n = number of individuals; NID = SNP not found in dbSNP database; P = probability; rs = reference sequence; *SLC40A1* = solute carrier family 40 member A1 gene; T = thymine

## Key to Table 3.2

<b>Blue</b>	No homozygous individuals identified	<b>Green</b>	Only wild type individuals identified for the variant
<b>Orange</b>	Marginal statistical significance	<b>Pink</b>	No heterozygote individuals identified for the variant
<b>Red</b>	Statistically significant difference ( $P < 0.05$ )	<b>Yellow</b>	Novel variant identified in this study

### 3.4.2 *In Silico* Analysis

*In silico* analyses were performed to ascertain whether any putative TFBSs are created or abolished (Table 3.3).

**Table 3.3 Putative TFBSs predicted to be created or abolished by the presence of the variants**

Gene	Variant	Programs					
		ALIBABA2		PROMO		TFBlast	
		Created	Abolished	Created	Abolished	Created	Abolished
<i>CYBRDI</i>	- 1849 T/G	-	-	-	-	-	-
	- 1844 C/G	GATA-1	-	-	GR-alpha	-	-
	-1813 C/T	-	SP1	p53, LEF-1	RXR-alpha	-	-
	-1749 del(T)6G	-	HNF-3, Hb	-	-	-	-
	- 1477 G/A	-	E1	HNF-4alpha, C/EPBbeta	MEF-2A, HNF-4alpha	-	-
	- 1459 T/C	-	c-Myc, SP1	Pax-5, p53, HNF-1A	HNF-1A, PR B	-	-
	- 1346 T/C	-	-	-	-	-	-
	- 1272 T/C	NF-1	PR	-	c-Myb	-	-
	- 238 A/G	NF-1	CTF	-	GR-beta, C/EBPbeta, C/EBPalpha, XBP-1	OTX2	-
	- 166 C/G	SP1, ETF	Krox-20, CACCC-bi	E2F-1	NF-1, NFI/CTF	-	-
<i>HFE</i>	-561 A/G	-	-	TFII-I, Pax-5, p53, E2F-1	-	-	-

Abbreviations: A = adenine; C = cytosine; CACCC-bi = CACCC-binding; C/EBP = CCAAT/enhancer binding protein; c-Myb = cellular myeloblastosis transcription factor; c-Myc = cellular myelocytomatosis transcription factor; *CYBRDI* = cytochrome B reductase 1 gene; E1 = enhancer 1; E2F-1 = E2 promoter binding factor 1; ETF = embryonic TEA domain-containing factor; G = guanine; GR = glucocorticoid receptor; Hb = Hunchback; *HFE* = haemochromatosis protein; HNF = hepatocyte nuclear factor; Knox-20 = knotted1-like homeobox 20; LEF-1 = lymphoid enhancer-binding factor 1; MEF-2A = myocyte enhancer factor 2A; NF-1 = nuclear factor 1; NFI/CTF = nuclear factor 1/ CCAAT-binding transcription factor; OTX2 = orthodenticle homolog 2; Pax-5 = paired box 5; PR = progesterone receptor; RXR-alpha = retinoid X-receptor alpha; SP1 = specificity protein 1; T = thymine; TFII-I = transcription factor II-I ; XBP-1 = x-box binding protein 1

A total of 41 sites were identified between the three programs used, of which 18 putative TFBSs were created and 23 were abolished. None of the programs used are in agreement for the majority of the statistically significant variants examined. The exceptions to this are the variants -1849 T/G and -1346 T/C both located in the promoter region of *CYBRDI* as they

were both shown to have no effect on the original position and orientation of any existing TFBSs.

### 3.4.3 Iron Parameter Analysis

#### 3.4.3.1 Descriptive Statistics

Table 3.4 shows the descriptive statistics obtained for the iron parameter levels in the patient and control groups.

**Table 3.4 Descriptive statistics for iron parameter levels compared between patient and control groups**

Iron parameter	Patient group		Control group	
	Untransformed data	Transformed data	Untransformed data	Transformed data
Serum ferritin	Mean: 213.31	Mean: 4.75	Mean: 99.66	Mean: 4.45
	Median: 129	Median: 4.86	Median: 80.7	Median: 4.39
	Mode: 34	Mode: 3.53	Mode: 51.4	Mode: 3.94
	S.D.: 284.20	S.D.: 1.16	S.D.: 59.06	S.D.: 0.57
	Range: 1642	Range: 5.33	Range: 219.2	Range: 2.40
	Min: 8	Min: 2.08	Min: 22	Min: 3.09
	Max: 1650	Max: 7.41	Max: 241.2	Max: 5.49
Serum iron	Mean: 13.92	Mean: 2.30	Mean: 12.63	Mean: 2.37
	Median: 10.9	Median: 2.39	Median: 12.9	Median: 2.56
	Mode: 8.4	Mode: 2.13	Mode: 17	Mode: 2.83
	S.D.: 11.96	S.D.: 0.87	S.D.: 5.31	S.D.: 0.79
	Range: 65.1	Range: 3.79	Range: 24.9	Range: 4.43
	Min: 1.5	Min: 0.41	Min: 0.3	Min: -1.20
	Max: 66.6	Max: 4.20	Max: 25.2	Max: 3.23
Serum transferrin	Mean: 2.37		Mean: 2.03	
	Median: 2.44		Median: 2.11	
	Mode: 1.67		Mode: 0.91	
	S.D.: 0.61		S.D.: 0.57	
	Range: 2.58		Range: 2.09	
	Min: 1.1		Min: 0.91	
		Max: 3		
Transferrin saturation	Mean: 25.58	Mean: 2.96	Mean: 26.14	Mean: 3.09
	Median: 20.3	Median: 3.01	Median: 24	Median: 3.18
	Mode: 30	Mode: 3.40	Mode: 24	Mode: 3.18
	S.D.: 21.41	S.D.: 0.77	S.D.: 14.45	S.D.: 0.70
	Range: 107.4	Range: 3.32	Range: 71.88	Range: 4.02
	Min: 4	Min: 1.39	Min: 1.32	Min: 0.27
	Max: 111.4	Max: 4.71	Max: 73.20	Max: 4.29

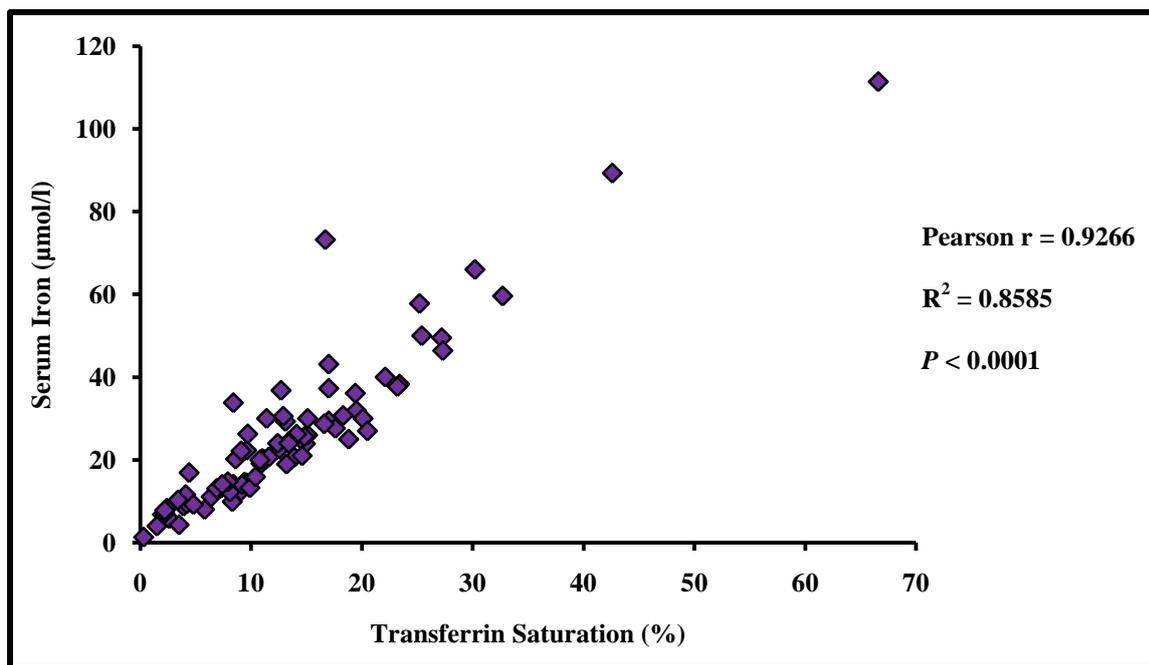
Abbreviations: Max = maximum; Min = minimum; S.D. = standard deviation

A large S.D. is observed for untransformed data in serum ferritin, serum iron and transferrin saturation. For these parameters the mean and median differ greatly which is a sign of potential non-normal distribution. After transformation the S.D. is dramatically decreased and

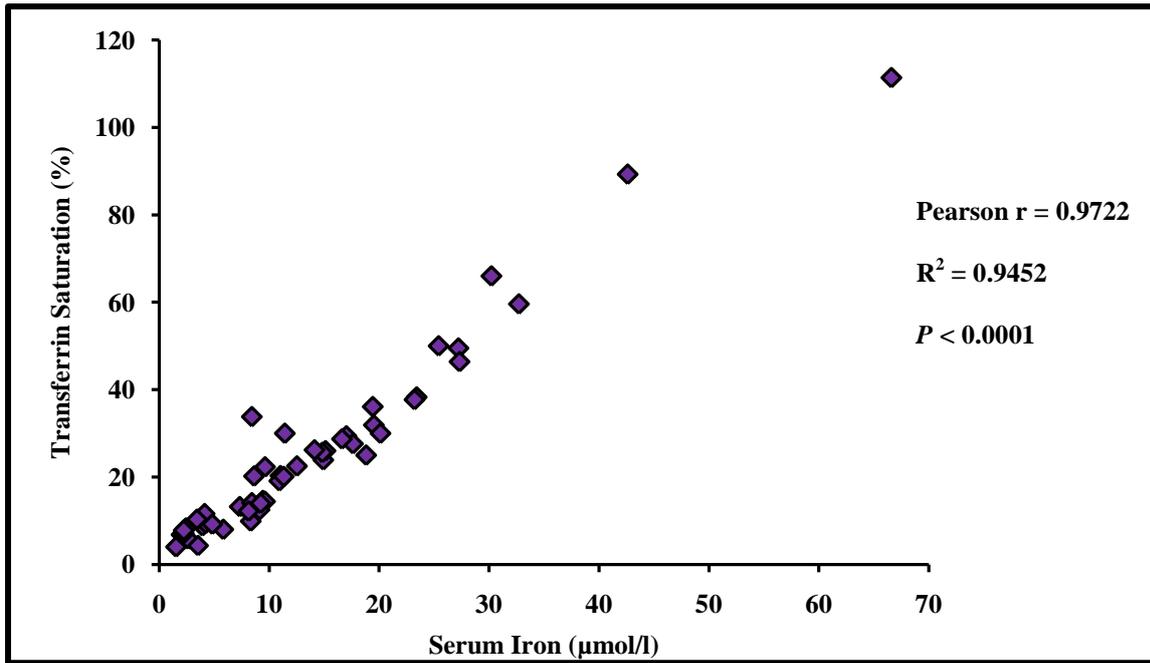
the mean and median for each transformed parameter is similar. Serum transferrin does not show the same pattern.

### 3.4.3.2 Pearson *r* Correlation

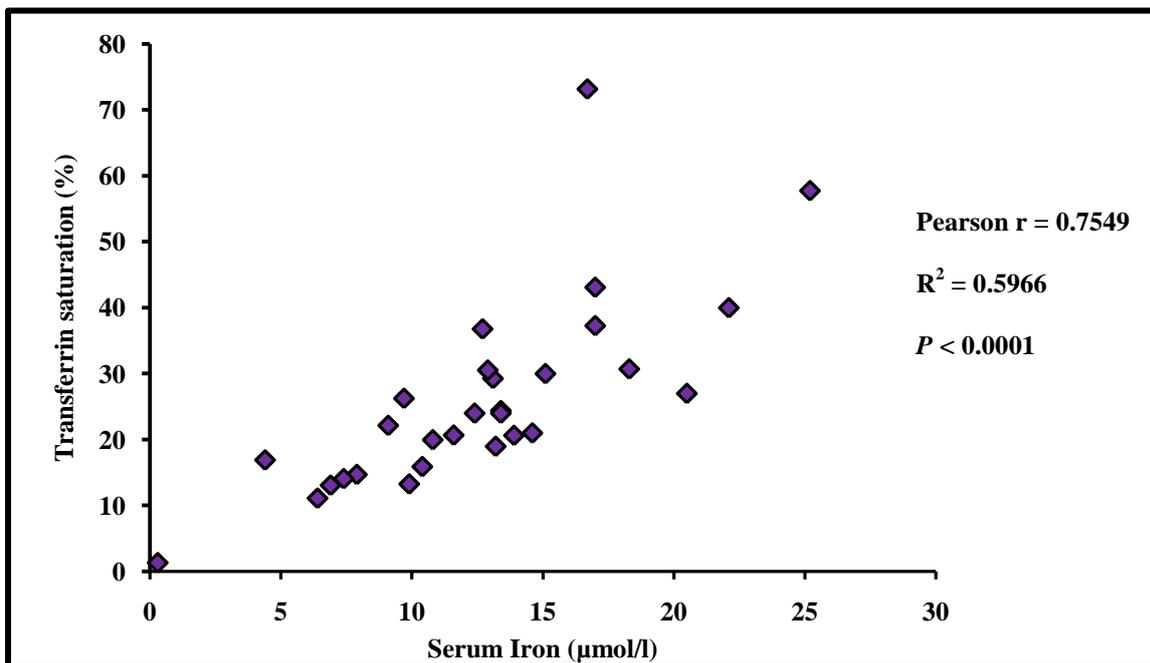
The Pearson *r* correlation was performed using GraphPad Prism™ (version 5.0.1) (GraphPad Software, San Diego, CA) ([www.graphpad.com](http://www.graphpad.com)) to determine whether any correlations exist between iron parameters measured, including serum ferritin, serum iron, serum transferrin and transferrin saturation. These correlations were performed between iron parameters within the entire cohort as well as within the patient and control groups separately. Only analyses demonstrating statistically significant correlations are shown. Figure 3.1 shows a strong positive correlation between serum iron and transferrin saturation levels for the entire cohort. This relationship is mimicked in both the patient and control groups, shown in Figures 3.2 and 3.3 respectively, although the relationship is weaker in the control group.



**Figure 3.1 Graph depicting the strong positive correlation between serum iron and transferrin saturation levels for the entire cohort**

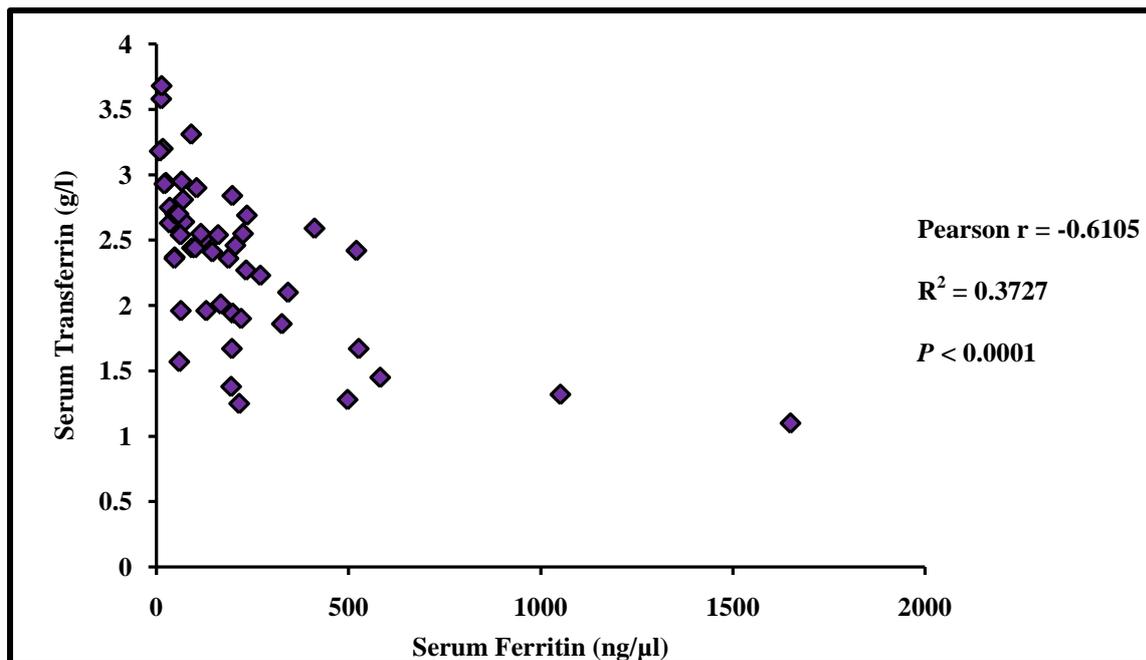


**Figure 3.2** Graph depicting the strong positive correlation between serum iron and transferrin saturation levels in the patient group



**Figure 3.3** Graph depicting the strong positive correlation between serum iron and transferrin saturation levels in the control group

Further analyses showed a weak negative correlation between serum transferrin and serum ferritin levels only in the patient group (Figure 3.4).



**Figure 3.4 Graph depicting the negative correlation between serum ferritin and serum transferrin in the patient group**

No other combinations of iron parameter comparisons yielded statistically significant results.

### ***3.4.3.3 Normality Distribution and Data Transformation***

The untransformed patient group iron parameter data for serum ferritin, serum iron and transferrin saturation were non-normally distributed and the natural logarithm function was used to transform the data. Following data transformation it was found that the patient group's iron parameters were normally distributed. The data for serum transferrin was found to be normally distributed for both the patient and control groups. The remaining control group iron parameter data (for serum ferritin, serum iron and transferrin saturation) were also transformed but were found to be non-normally distributed regardless of transformation. This resulted in a non-parametric test being used to compare the data of the patient group to that of the control group for these iron parameters.

In order to determine if a variant had any effect on iron parameters, the data was split into two groups for each variant (this was only performed for variants that had iron parameter data available). The first group was the “variant absent” group (composed of only wild type individuals) while the second group was the “variant present” group (composed of

individuals both heterozygous and homozygous for the variant). In this manner the groups were compared to one another for each variant to determine whether an overall effect could be seen for each iron parameter. For every redistribution of the data for each new variant being investigated, it was determined whether the data were normally distributed and depending on the result either a parametric (normally distributed data) or a non-parametric (non-normally distributed data) were used to compare the two groups

#### 3.4.3.4 T-tests (*Parametric and Non-parametric*)

Figures 3.5-3.8 represent the results obtained when comparing the iron parameters of the patient to the control group as well as the results of the tests determining the influence of a particular variant on each of the iron parameters. All genotyped variants with iron parameter data available were tested. Only graphs for variants with statistically significant results have been shown (graphs for the other variants tested can be seen in Addendum C)

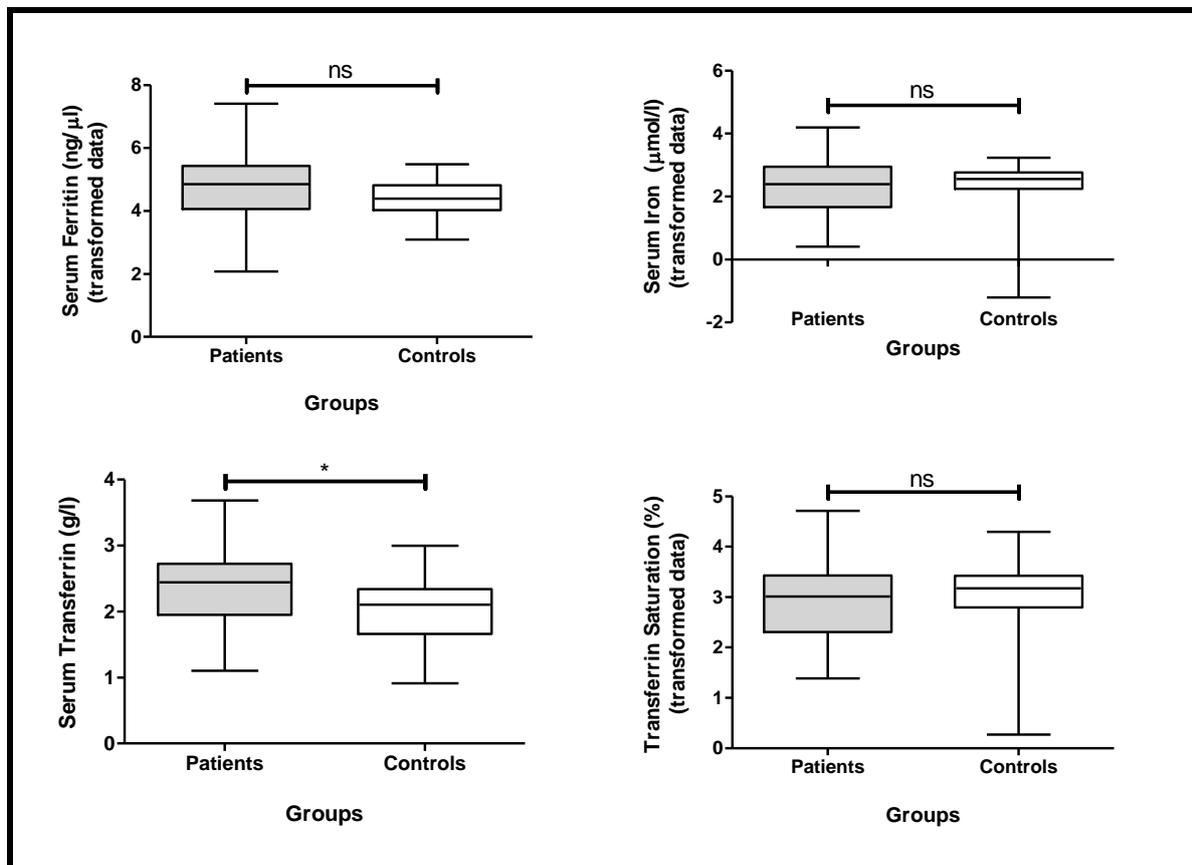
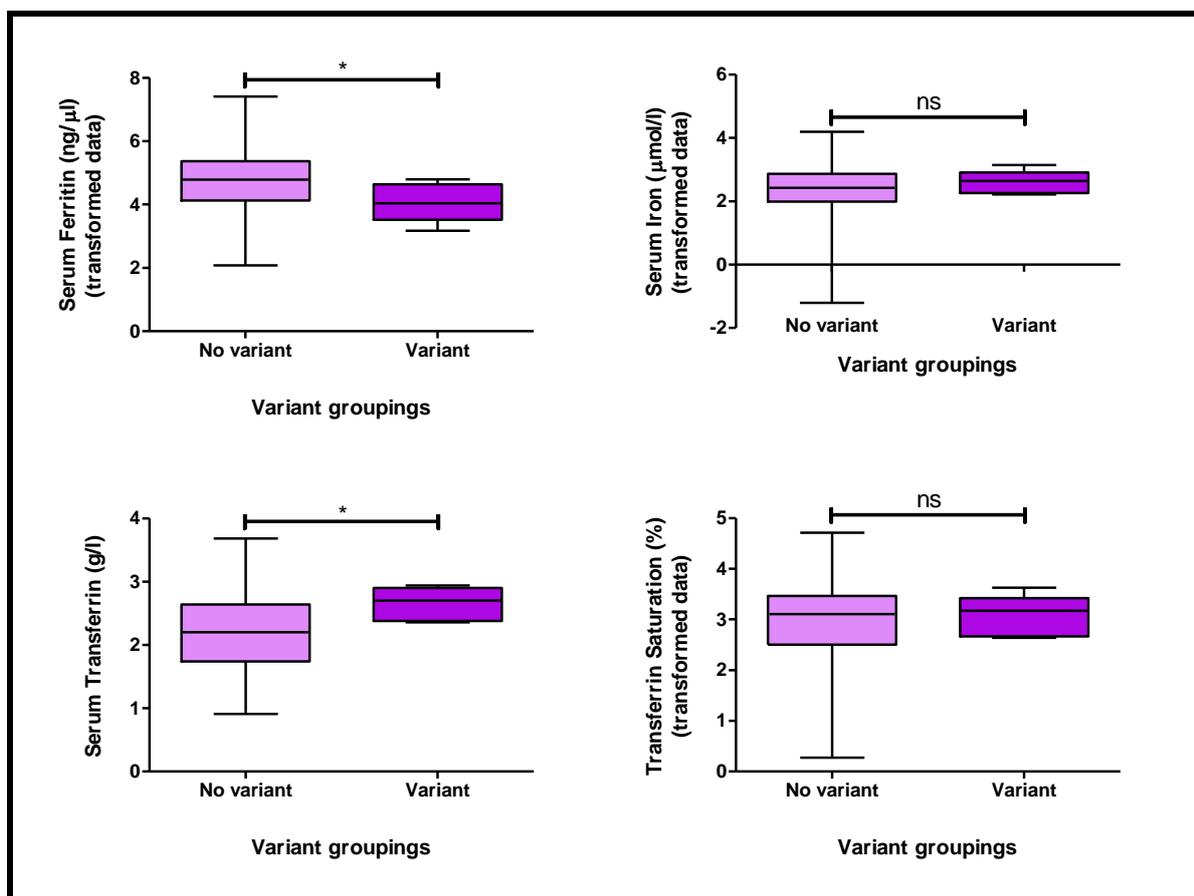


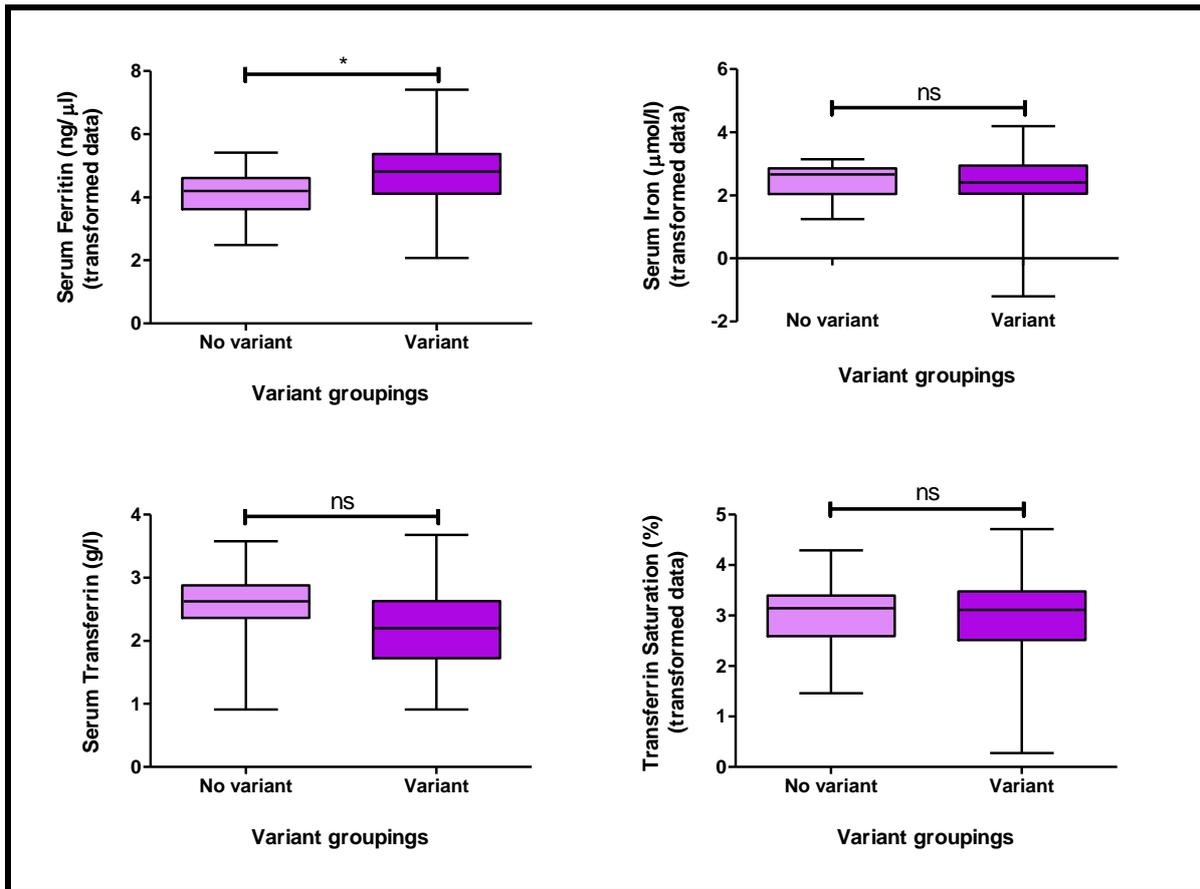
Figure 3.5 Comparison of iron parameters between the patient and control groups

It was determined that only serum transferrin was statistically significantly different ( $P < 0.05$ ) between the patient and control groups (Figure 3.5). Although the remaining iron parameters are not significantly different, differences can still be observed in the mean values between the patient and control groups.



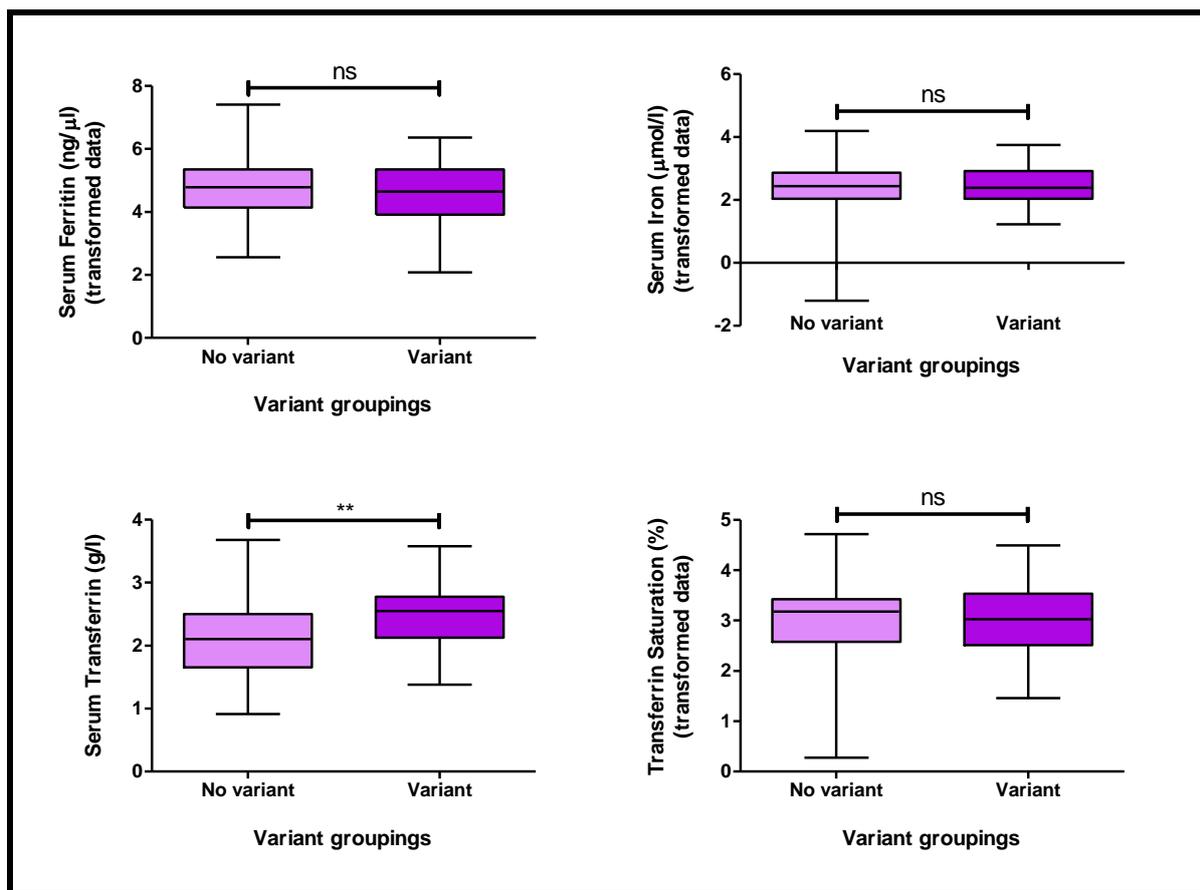
**Figure 3.6 Comparison of iron parameters to determine the effect of the variant -1470 C/T in the promoter region of *SLC40A1***

The variant -1470 C/T situated in the promoter region of *SLC40A1* showed statistically significant differences ( $P < 0.05$ ) in both serum ferritin and serum transferrin when comparing the variant absent to the variant present group (Figure 3.6). Although visible differences could be observed in both serum iron and transferrin saturation when comparing the two groups, these differences demonstrated no statistical significance.



**Figure 3.7 Comparison of iron parameters to determine the effect of the variant -1355 G/C in the promoter region of *SLC40A1***

The variant -1355 G/C showed no statistically significant differences in serum iron, serum transferrin or transferrin saturation when comparing the variant-containing to the non-variant-containing groups but did demonstrate a statistically significant difference when comparing serum ferritin for the two groups ( $P < 0.05$ ) [Figure 3.7].



**Figure 3.8 Comparison of iron parameters to determine the effect of the variant -1098 G/A in the promoter region of *SLC40A1***

The last variant to be investigated for differences in iron parameters between individuals either wild type for the variant or carrying the variant (heterozygotes as well as homozygotes) was -1098 G/A located in the promoter region of the *SLC40A1* gene (Figure 3.8). While serum ferritin, serum iron and transferrin saturation showed differences between the two groups, these differences exhibited no statistical significance. Serum transferrin however demonstrated strong statistical significance ( $P < 0.01$ ) when comparing the variant absent and variant present groups.

Bonferroni correction was performed where a  $P$ -value of less than 0.0017 was required to remain statistically significant. All previously statistically significant results do not stand up to this correction but it is important to note that the Bonferroni correction is one of the most stringent tests to deter the finding of false positives and thus the results should not be dismissed outright (Hochberg & Benjamini, 1990).

### 3.5 DISCUSSION

An assortment of techniques was employed in this study in order to determine whether any variation exists in the promoter region of 5 genes – *CYBRD1*, *HAMP*, *HFE*, *LTF* and *SLC40A1* – implicated in iron metabolism. Initially, a small subset of samples was subjected to semi-automated bi-directional sequencing analysis and this yielded a total of 40 variants identified in the promoter region of the genes under investigation. Ten novel variants were identified in the current study and 30 variants proved to be previously identified (dbSNP database).

The remaining samples of the cohort were genotyped for these variants using either HEX-SSCP analysis, RFLP analysis where possible, or, in cases where the variants were too closely spaced for the other analysis techniques, semi-automated bi-directional sequencing analysis was used. Once genotyping was complete, the data was analysed using Fisher's exact test or the chi-squared test (as described in section 2.6.2) in TFPGA, which tested for departure from HWE (Table 3.1). The *LTF* promoter region variants proved difficult to genotype and were subsequently omitted from further analyses. It should be noted that only homozygous individuals were observed for the variant -457 T/C in both the patient and control groups, which may indicate a different ancestral allele for this population group.

The program TFPGA was also utilised to calculate the genotype and allele frequencies and was used to perform the Exact test, which is used to determine whether the patient and controls groups have statistically significant differences in allele frequencies (Table 3.2). The Exact test performs an algorithm with a set number of replications. The replications of the test reduce the number of false positives obtained through familywise error rate and so it is unnecessary to use a correction method such as the Bonferroni correction (Moran, 2003; Nakagawa, 2004). TFPGA only performs the Exact test when comparing allelic population differentiation; for this reason the program Genepop was used to examine genotypic population differentiation. The default settings for both programs were used (Table 3.5).

**Table 3.5 Default settings for the Exact test used in TFPGA and Genepop**

	TFPGA	Genepop
Number of dememorization steps	1 000	10 000
Number of batches	10	100
Number of permutation/ iterations	2 000	5 000

---

Following mutation detection, *in silico* analyses were performed in order to ascertain whether any putative TFBSs were abolished or created due to the presence of a variant (Table 3.3). The predicted implications of these TFBSs and their transcription factors were investigated. Finally, the iron parameter data were investigated to determine whether any of the variants had an effect on the parameter levels. This was ascertained by performing unpaired *t*-tests to compare groupings of data. The data were grouped for each variant based on whether the variant was present or absent, i.e., there was a ‘variant present’ group and a ‘variant absent’ group. The ‘variant present’ group consisted of individuals that are either heterozygous or homozygous for the variant, while the ‘variant absent’ group consisted of only wild type individuals. This test was used to determine whether a statistically significant difference occurred between the two groups based on the presence of each variant (Figures 3.5 to 3.8).

Of the 31 loci tested, 18 showed departure from HWE. Genotyping errors and small sample size may inflate the occurrence of type 1 errors when assessing HWE but other factors may also play a role (Li & Li, 2008). When selecting individuals for a genetic study from a specific group in the population (in this case from a TB patient group), one is no longer sampling randomly and an over-representation occurs of these individuals when compared to the whole population. This non-random sampling may be one of the reasons for departure from HWE. Other documented reasons for departure from the model may include: migration, mutation, selection, the presence of a rare allele, undetected deleted alleles in heterozygotes, inbreeding and admixture of the population (Schaid *et al.*, 2006). It is possible that one or more of these factors, such as selection or the presence of rare alleles, are influencing the study cohort, which could also account for the deviation of so many loci from HWE.

The deviation of so many loci from HWE, while troubling, is most likely due to the small sample size of the study population and so future studies should correct for this. However, it is important to note that linkage disequilibrium (LD) may cause deviation from HWE as it affects the random assortment of alleles, thereby influencing the allele frequencies of the linked loci (Feder *et al.*, 1996; Nielsen *et al.*, 1999). Feder *et al.* (1996) go on to explain that it may even be possible to identify disease-associated loci by their departure from HWE, thus it should not be surprising to find that some of the variants identified in this study are found to be in Hardy-Weinberg disequilibrium (HWD) within the patients groups. Feder *et al.* (1996) noted that alleles at these loci appear to have an excess of homozygosity, which is

---

mimicked in some of the current study's data, which could partly explain the large number of loci that deviate from HWE.

Goudet *et al.* (1996a) suggest the use of an Exact test for genotypic population differentiation as it does not assume HWE for the loci tested; the test assumes random sampling of genotypes rather than alleles. For this reason Genepop was used since it performs this test, thus more emphasis is placed on the genotype differences between the patient and control group than on the allele differences (as the TFPGA equivalent test assumes HWE for the allele).

The Pearson  $r$  correlation was performed to determine if any of the iron parameters shared a relationship or showed reliance on one another. This was performed for the entire cohort's iron parameter data as well as for the patient and control groups separately, to determine if disease status affected these relationships. A relationship was discovered to exist between serum iron and transferrin saturation and this was visible in the entire cohort as well as within the patient and control groups (Figures 3.1 to 3.3). Since iron has the potential to be toxic when unbound, binding of it to transferrin is encouraged to the point where virtually no unbound iron is found in the bloodstream (Halliwell & Gutteridge, 1984; Ong *et al.*, 2006).

The relationship between serum iron and transferrin saturation is such that when the level of iron in the bloodstream increases the potential for instability and free radical formation increases, which leads to more iron becoming transferrin-bound and therefore more transferrin will become saturated. It was observed that the relationship is much stronger in the patient group than in the control group. The Pearson  $r$  value in the patient group was 0.9722 (i.e., 97% dependence between the two parameters) in comparison with the control group that had a Pearson  $r$  value of 0.7549 (only 75% dependence). This may be explained by the body's innate ability to compete for iron with invading pathogens (Ong *et al.*, 2006). The infection of *M. tuberculosis* could have caused the system to react by binding as much iron as possible in order to minimise iron availability for the pathogen.

Interestingly, a relationship exists between serum ferritin and serum transferrin in the patient group alone (Figure 3.4). Although the relationship is not as strong as that of serum iron and transferrin saturation, it is still considered statistically significant as it has a Pearson  $r$  value of more than 0.6 (Pearson  $r = 0.6105$ ). This could be an adaptive response by the body to

---

reduce iron availability for the pathogen by producing excess ferritin and transferrin in response to infection and inflammation, thereby decreasing the amount of unbound iron (Tran *et al.*, 1997).

Unpaired *t*-tests were performed to determine whether there is a statistically significant difference in the iron parameter levels when comparing the patient group to the control group. Although differences were observed between the groups for each parameter, only serum transferrin proved to be statistically significantly different ( $P < 0.05$ ). The patient group had a higher mean value for this iron parameter than the control group. These results and the results of the Pearson's *r* correlation analysis may indicate an adaptive response of the host to pathogenic infection in an attempt to limit iron availability (Ong *et al.*, 2006). Therefore it is not surprising that the patient group has higher levels of serum transferrin, probably in response to the original infection.

The *HAMP*, *HFE* and *SLC40A1* promoter region variants were investigated for their effect on the iron parameters serum ferritin, serum iron, serum transferrin, and transferrin saturation. No statistically significant results were found for either the *HAMP* or the *HFE* promoter variants, suggesting that these variants do not influence the levels of these iron parameters. The *HAMP* promoter variant -323 C/T was not tested in this manner as only one individual was found to be heterozygous while the remaining individuals in the cohort were wild type. Three of the *SLC40A1* promoter variants yielded statistically significant results when determining the influence of a variant on iron parameters (Figures 3.6-3.8).

The variant -1470 C/T was found to be statistically significantly different ( $P < 0.05$ ) when comparing the variant absent group to the variant present group for the iron parameters serum ferritin and serum transferrin. The median value of serum ferritin in the variant absent group (Figure 3.6) was higher than that of the variant present group. Conversely, serum transferrin was, on average, higher in the variant present group than that of the variant absent group. This would indicate that the variant might be involved with the decreased accumulation of serum ferritin and the increased accumulation of serum transferrin. This variant occurs in the promoter region of the *SLC40A1* gene and as such may influence the export of iron from cells. From the iron parameter data it can be expected that this variant could cause the increased expression of *SLC40A1* which in turn would lead to increased iron accumulation in the blood stream and thus increased transferrin accumulation and decreased ferritin

---

accumulation. Future studies should investigate the effect of this variant on gene expression in order to corroborate this. Interestingly, this variant seems to be related to the increased accumulation of transferrin and decreased accumulation of ferritin which would suggest that transferrin is being used as the primary mechanism to bind iron. This could provide the ideal environment for *M. tuberculosis* proliferation as the bacterium more readily outcompetes transferrin for iron than it does ferritin (Gobin & Horwitz, 1996).

The variants -1355 G/C and -1098 G/A both, on average, had higher levels of serum ferritin and serum transferrin respectively in the variant present group (Figures 3.7 and 3.8). Both variants therefore play some role in the accumulation of these proteins. As described for the variant -1470 C/T, the increased accumulation of transferrin may create an advantageous environment for *M. tuberculosis* proliferation. However, the increased accumulation of ferritin related to these variants may be attributed to an adaptive response to bind iron to ferritin in order to decrease iron availability to the pathogen.

The exact mechanism by which these three *SLC40A1* promoter variants (-1470 C/T; -1355 G/C, -1098 G/A) affect iron parameter levels is unclear and requires further elucidation. Functional studies in combination with staining different bodily tissues for iron (liver, pancreas, macrophage etc.) may aid in the understanding of the influence these variants have on iron parameter levels.

Although variants were identified in the promoter region of *HAMP*, *LTF* and *SLC40A1*, they were excluded from further analyses either due to the difficulty in genotyping or to the non-significant results obtained from genotype and allele frequency comparisons between the patient and control groups. The following sections will deal with the results of the *in silico* analyses.

### **3.4.1 *CYBRDI***

Mutation detection analysis of the promoter region of the *CYBRDI* gene revealed the presence of four novel (-849 C/G -492 A/G, -454 C/T and -397 A/C) and 16 previously described variants [-1849 T/G, -1844 C/G, -1813 C/T, -1749 del(T)<sub>6</sub>G, -1540 G/A, -1477 G/A, -1459 T/C, -1346 T/C, -1272 T/C, -1011 T/A, -627 T/C, -624 G/A, -449 C/G, -399 T/G, -238 A/G and -166 C/G]. The control group used for comparison for this promoter region was derived from a previous study and, as such, no iron parameter data was available. For this

---

reason no determinations could be made pertaining to the influence of any of the *CYBRDI* promoter region variants on iron parameters.

Nine of the 20 identified variants were shown to differ statistically significantly when comparing the genotype frequencies of the patient and control groups, and these variants were subjected to *in silico* analyses to determine if any TFBSs were created or disrupted by their presence (Table 3.3). The variants -1849 T/G and -1346 T/C were not predicted to create or abolish any TFBSs due to their presence or absence.

The variant -1813 C/T was predicted to create a p53 and a LEF-1 TFBS but to abolish a SP1 and a RXR-alpha TFBS. The transcription factor p53 has been shown to cause increased expression of genes to which it is linked, specifically during periods of infection (Weizer-Stern *et al.*, 2007). If this factor caused the increased expression of *CYBRDI* it could lead to the enhanced absorption of iron, increasing the body's iron content, and could therefore potentially amplify an individual's susceptibility to TB infection. The lymphoid enhancer-binding factor 1 (LEF-1) transcription factor has been connected to lymphocyte differentiation and, as yet, has not been seen to be expressed in any tissue other than lymphoid tissue (Travis *et al.*, 1991). For this reason it is unlikely that this transcription factor is associated with the *CYBRDI* gene, which is predominantly expressed in the gut.

Specificity protein 1 (SP1) is a zinc finger transcription factor that binds to GC-rich regions of several promoters (Kaczynski *et al.*, 2003). This transcription factor has been implicated in many processes such as cell growth, apoptosis, differentiation and immune responses. This factor can both repress and stimulate expression of target genes in response to physiological and pathological stimuli (Dhar *et al.*, 2006). Its exact functioning when related to *CYBRDI* warrants further investigation because the presence of other variants in this gene is either predicted to create or abolish SP1 sites.

The retinoid X-receptor alpha (RXR-alpha) transcription factor has been shown to form heterodimers with HNF-4alpha which exert their function by binding to DNA (Li *et al.*, 2000). RXR-alpha is part of the same steroid/thyroid nuclear receptor superfamily as HNF-4alpha and, as such, is potentially involved in the same gene expression (Garcia *et al.*, 1993). As HNF-4alpha has been implicated in *TF* expression it is possible that it, together with

---

RXR-alpha, is involved in the expression of other iron metabolism genes such as *CYBRDI* (Zakin, 1992).

The *CYBRDI* promoter variant -1749 del (T)<sub>6</sub>G caused the elimination of the transcription factors HNF-3 and Hb. The hepatocyte nuclear factor (HNF) transcription factors have been shown to be involved in liver development (Li *et al.*, 2000). Zakin (1992) showed that HNF-3 $\alpha$  (a subgroup of HNF-3) acts as an enhancer for *TF* mRNA transcription in a liver cell line and this could indicate some function in *CYBRDI*. In fact, Collins and Hu (2007) showed that HNF-3 was among a group of transcription factors found to be enriched in a large-scale microarray-based study of intestinal gene expression during iron deprivation. Hunchback (Hb) is a transcription factor predominantly derived from the species *Drosophila melanogaster* and thus is not likely to play a role here (Berman *et al.*, 2002).

The presence of the promoter variant -1477 G/A predicted the creation of the TFBSs HNF-4alpha and C/EBPbeta and elimination of the sites E1, HNF-4alpha, and MEF-2A. HNF-4alpha is part of the HNF transcription factor family responsible for the regulation of liver development (Li *et al.*, 2000). Zakin (1992) showed that HNF-4alpha acts as a transcription factor for the expression of *TF*. Binding of this factor to the promoter region of *TF* induced increased expression of this gene. From this it could be hypothesised that HNF-4alpha is involved in the expression of other iron genes such as *CYBRDI*. It is unlikely that this transcription factor played a role here, however, as it was shown to be both gained and lost due to the presence of the variant. The CCAAT/enhancer binding protein (C/EBP) transcription factor has been shown to interact with HNF-4alpha specifically in the promoter region of *HAMP* (Courselaud *et al.*, 2002). Courselaud and colleagues (2002) explain that C/EBPalpha is a potent enhancer of *HAMP* expression while C/EBPbeta is a weak enhancer. The creation of a C/EBPbeta site might lead to a weak increase in the expression of *CYBRDI*.

Very little information could be found regarding the E1 TFBS and any link it may have to any form of iron gene expression and so it is unlikely that this factor is associated with *CYBRDI* expression. The monocyte-specific enhancer factor 2A (MEF-2A) transcription factor is a class of skeletal-specific factors which can indirectly either inhibit or increase the expression of certain skeletal genes (McGee & Hargreaves, 2004). Since this is a tissue-specific transcription factor it is unlikely that it acts on the *CYBRDI* gene as the necessary tissue environment is not present to direct the use of this factor.

---

The introduction of the variant -1459 T/C is predicted to produce the TFBSs Pax-5, p53, and HNF-1A and to abolish the sites for c-Myc, SP1, HNF-1A, and PR-B. Pax-5 is part of a family of highly conserved (in vertebrates) transcription factors that have a basic helix-loop-helix structure and have the ability to bind E-box motifs in the promoter region of the *transferrin (TF)* gene within Sertoli cells, which subsequently increases *TF* expression (Wang *et al.*, 2010; Chaudhary *et al.*, 1997). With the exception of the research published by Chaudhary and colleagues (1997), little evidence exists to link Pax-5 to the regulation of genes involved in iron metabolism. As described for variant -1813 C/T, the introduction of a p53 TFBS could lead to the increased expression of *CYBRD1*.

The c-Myc transcription factor contains basic helix-loop-helix and leucine zipper domains as well as a transactivation domain (Wu *et al.*, 1999). This transcription factor functions by forming a heterodimer with a related protein, MAX, which has a similar structure. Wu and associates (1999) showed that c-Myc has the ability to both repress and stimulate the expression of *FTH* (the heavy subunit of ferritin) and *IRP2* respectively, and so its precise influence on *CYBRD1*, if any, should be determined through functional analysis.

The progesterone receptor B (PR-B) transcription factor is a hormone receptor that is part of the steroid superfamily. The PRs are generally involved in reproductive functions but have been shown to be expressed in other tissues such as the brain, liver, mammary glands, and spermatozoa (Lösel & Wehling, 2003; Gadkar-Sable *et al.*, 2005). Hardy *et al.* (2008) demonstrated PR's ability to antagonise the inflammatory response pathway, which is known to cause the accumulation of iron. One could then suggest that if PR were truly acting in the promoter region of *CYBRD1* then its abolishment could lead to accumulated iron.

HNF-1A is another of the HNF family of transcription factors, the alpha and beta subtypes of which have been shown to influence the expression of genes localised exclusively in the small intestine within absorptive enterocytes (Wu *et al.*, 1994). HNFs are generally thought to primarily influence the expression of genes in the liver but the evidence presented by Wu and colleagues (1994) indicates the possibility of HNF-1A to influence genes involved in dietary absorption. However, the results showed that HNF-1A was both lost and gained, so the overall effect of this variant in terms of HNF-1A is negligible.

---

The variant -1272 T/C in the *CYBRDI* promoter region was predicted to create a NF-1 TFBS and disrupt a PR and a c-Myb site. The nuclear factor 1 (NF-1) transcription factor is known for its involvement in the ras signal transduction pathway and has been associated with human cognitive development (Weeber & Sweatt, 2002). Little proof exists of a connection between NF-1 and iron metabolism and if this transcription factor does truly interact with *CYBRDI* then this would merit further investigation into the method and reasons for interaction. The c-Myb transcription factor is employed in the hematopoietic system and gastrointestinal tract (Ramsay *et al.*, 2003). This transcription factor has also been shown to play a critical role in the development of lymphocytes (Xiao *et al.*, 2007). Whether this transcription factor influences the expression of *CYBRDI* remains to be investigated and this is required before the actual importance of this site being abolished due to the presence of this variant can be established.

The *CYBRDI* promoter variant -238 A/G was predicted to create a NF-1 and a OTX2 TFBS and it was predicted to abolish the GR-beta, CTF, C/EBPalpha, C/EBPbeta and XBP-1 TFBSs. The orthodenticle homolog 2 (OTX2) transcription factor is a member of the bicoid subfamily of proteins containing a homeodomain that plays a role in brain development (Lamonerie *et al.*, 1996; Simeone, 1998). It is unclear if a connection exists between this transcription factor and *CYBRDI* expression. The CCAAT-binding transcription factor (CTF) has been shown to be a transcriptional activator and influences the  $\beta$ -globin gene *in vitro*, among others (Jones *et al.*, 1985). Whether a connection exists between CTF and *CYBRDI* expression is unclear and merits further investigation.

Glucocorticoids are known to be integral in the basal and stress-related homeostasis of all higher organisms (Chrousos & Kino, 2005). They have been shown to influence approximately 20% of the human genome and their effects are demonstrated in almost every organ and tissue. The GR-alpha group of glucocorticoids have been described as complex, multifunctional domain proteins that operate in a ligand-dependent fashion when acting as transcription factors. Glucocorticoids have been implicated in the inhibition of inflammation and as such would inhibit the accumulation of iron within cells such as macrophages (Hentze *et al.*, 2004; Rhen & Cidlowski, 2005). The GRs have been shown to influence the expression of several iron genes but have not yet been connected to the *CYBRDI* gene. Despite this the possibility of such an interaction occurring is high since GRs are known to influence so much of the genome.

---

Unlike GR-alpha, GR-beta does not bind to GC-rich regions and despite being thought to be transcriptionally inactive it has been demonstrated to have the ability to influence GC-mediated transcription (Oakley *et al.*, 1996; Haarman *et al.*, 2004). More specifically GR-beta can inhibit signal transduction by binding and subsequently inactivating GR-alpha. Since GR-beta is transcriptionally inactive in the absence of GR-alpha, it can be postulated that this transcription factor will either have no effect on the expression of *CYBRDI*, or it will inhibit any GR-alpha present, which will cause the increased absorption of iron from the gut and possibly increase iron accumulation in macrophages (Hentze *et al.*, 2004; Rhen & Cidlowski, 2005).

The transcription factor X-box binding protein 1 (XBP-1) consists of a basic leucine zipper (bZIP) domain and has been implicated in the regulation of MHC class II genes by binding to a promoter element referred to as an X-box (Yoshida *et al.*, 2001). Some iron genes, such as *HFE*, bear structural similarities to the MHC-class proteins and so one could postulate that XBP-1 may interact with it but it would be a far leap to connect XBP-1 to *CYBRDI* expression.

The *CYBRDI* promoter variant -166 C/G was predicted to produce SP1, ETF, and E2F-1 TFBSs but to disrupt Krox-20, CACCC-bi, NF-1, and NFI/CTF sites. The epidermal growth factor receptor (EGFR) has been recognised as an oncogene, which has been shown to be over-expressed in certain types of cancers (Kageyama *et al.*, 1988). Furthermore, this gene seems to lack both a 'TATA' box and a 'CAAT' box but contains multiple GC boxes, which are recognised by the SP1 transcription factor. Evidence suggests that another transcription factor regulates the EGFR gene and this transcription factor was isolated and purified by Kageyama and colleagues (1988), who termed the factor EGFR-specific transcription factor (ETF). A study by Bem *et al.* (2009) on mechanical ventilation and inflammation in the lungs of pneumovirus infected mice found an enrichment of ETF, among other transcription factors. These results demonstrate that downstream proinflammatory pathways may be transcriptionally activated due to the interaction between mechanical ventilation and pathogens. This would indicate that ETF is involved in the proinflammatory response, which would indicate that its role in this pathway could lead to iron accumulation in the cells. Thus it could be hypothesised that the introduction of an ETF TFBS may increase an individual's susceptibility to TB infection since it could aid in the activation of an inflammatory response.

---

This variant was also predicted to create an E2F-1 TFBS. Although there appears to be no direct link between this transcription factor and iron metabolism, the factor has been identified as a key player in regulation of the cell cycle progression from the G1 phase to the S phase (Jamshidi-Parsian *et al.*, 2005). It has also been identified for its role in apoptosis together with the transcription factor p53 (Wu & Levine, 1994). E2F-1 directly affects the expression of the transcription factor c-Myc, which is involved in iron homeostasis, so it is possible that the presence of this transcription factor may indirectly affect genes involved in iron metabolism (Wu & Levine, 2005).

The transcription factor Krox-20 (also known as early growth factor 2 or *egr2*) is a zinc finger protein that is rapidly activated in fibroblasts by growth factors (Chavrier *et al.*, 1989; Bhat *et al.*, 1992). Research has shown that Krox-20 has decreased expression in *TF* knockout mice indicating that the transcription factor interacts with this gene (Levy *et al.*, 1999). This would signify a role for Krox-20 in the regulation of iron metabolism. Of specific relevance to the current study was the discovery of the enrichment for Krox in a microarray study on intestinal gene expression during iron deprivation by Collins and Hu (2007). This would link Krox transcription factors to *CYBRDI* expression but the exact mechanism by which this occurs requires further investigation.

The CACCC-binding factor (CACCC-bi) is part of a broad range of zinc finger transcription factors that bind 'CACCC-boxes' in DNA, which leads to either activation or suppression of gene expression (van Vliet *et al.*, 2000). Since these proteins share a similar binding sequence (CACCC) the program ALIBABA2 could not distinguish which CACCC-bi factor this was. This would need to be determined in order to assess the effect this factor has on *CYBRDI* expression, if any. The nuclear factor-I/CCAAT-binding transcription factor (NFI/CTF), which comes from a group of proteins that recognise and bind to the cellular DNA sequence GCCAAT, has been implicated in eukaryotic transcription (Uramoto *et al.*, 2003). These transcription factors are known to play a role in the regulation of genes that are ubiquitously expressed as well as those that are influenced by hormones, developmental stages, and nutrition. As described for the variant -1272 T/C, these NF-1 family factors have the ability to both repress and stimulate transcription of genes and, as such, will require further investigation in order to determine their effect on *CYBRDI* expression (Gronostajski, 2000).

---

### 3.4.2 *HFE*

Through mutation detection analysis 4 variants were identified in the promoter region of *HFE*. Two novel SNPs (-561 A/G and -331 A/C) and 2 previously described SNPs were identified (-1168 A/G and -467 G/C). The -331 A/C variant proved difficult to genotype as no RE recognition sites were present and analysis on HEX-SSCP PAA gels proved impossible. It was then decided to exclude this variant from further analyses.

Of the 3 remaining variants only -561 A/G yielded statistically significant results when comparing the patient and control groups for deviations in genotype and allele frequencies. The putative TFBS prediction program PROMO was the only program to predict changes to occur due to the presence of this variant. It predicted the creation of 4 TFBSs and their potential involvement in the regulation of the *HFE* gene was investigated.

The first TFBS predicted to be created by the presence of the variant -561 A/G is transcription factor II-I (TFII-I). TFII-I is a ubiquitously expressed multifunctional transcription factor that is believed to link signal transduction processes to transcription (Roy, 2001). This transcription factor is activated in reaction to extracellular signals. It has also been proposed that TFII-I can function as both a basal transcription factor and an activator. This would imply that TFII-I facilitates communication between the core promoter region and upstream activator sites. As a transcription factor it is known to upregulate expression of a gene in response to external stimuli. It is possible that the increased expression of *HAMP* could lead to increased macrophage iron concentrations, which would dissuade systemic pathogenic infection but would be ideal for mycobacterial infection of the macrophage (Sow *et al.*, 2007). This would indicate the potential for this variant to increase an individual's susceptibility to TB.

Paired box 5 (Pax-5) was the next TFBS predicted to be created in the presence of this variant. As described in section 3.4.1, its role in the regulation of *TF* may suggest that it is capable of regulating other genes involved in iron metabolism, such as *HAMP*, and, if so, may also implicate this variant in TB susceptibility since Pax-5 could potentially cause increased expression of hepcidin.

The TFBS for p53 was also identified as a putative binding site created by the presence of the -561 A/G variant. This transcription factor has been directly linked to the regulation of

*HAMP*, particularly during times of inflammation and infection (Weizer-Stern *et al.*, 2007). As described for TFII-I and Pax-5, the increased expression of *HAMP* could possibly cause increased susceptibility to TB. The last TFBS identified by PROMO to have been created by the presence of the -561 A/G variant is E2F-1. This transcription factor may affect iron gene regulation as described for *CYBRDI* promoter variant -166 C/G.

## **Chapter 4**

### Concluding Remarks and Future Prospects

---

## 4.1 CONCLUDING REMARKS AND FUTURE PROSPECTS

The aim of this study was to elucidate the role of five genes involved in iron metabolism – *CYBRD1*, *HAMP*, *HFE*, *LTF* and *SLC40A1* – in relation to *M. tuberculosis* susceptibility. The first objective identified in order to achieve this aim was to identify a suitable cohort consisting of TB patients and population-matched controls. A group comprising 49 TB patients was compared to a control group made up of a total of 51 healthy control individuals split into two subgroups for specific analyses. A limitation of this study was the small sample size of the cohort, but this was a pilot study and future work will include a larger cohort. The standard deviation for age between the two groups was significantly different and future work should aim to collect samples from individuals that are truly population-matched (for age and sex) or corrections for age and sex may be performed to minimise their influence on the data. The DNA samples from the control group were particularly difficult to amplify due to minimal quantity and age of the samples. Future researchers could corroborate the findings from this study and validate them through the use of a larger cohort.

The second objective identified by this study was to identify and compare iron parameters (including serum ferritin, serum iron, serum transferrin and transferrin saturation) between the groups. It was determined that serum iron and transferrin saturation share a strong positive relationship and that this relationship is particularly notable in the patient group. This could suggest an influence on iron parameters by disease status and adaptive response to infection. This relationship could be explained as the body's response to competitively bind iron to limit its availability for invading pathogens (Ong *et al.*, 2006). Interestingly, a relationship between serum ferritin and serum transferrin was exclusively observed in the patient group. This could be due to an adaptive response by the body in an attempt to limit free iron availability by producing excess ferritin (Tran *et al.*, 1997). This relationship shows a decrease in transferrin as ferritin increases. This could also be explained as an adaptive response by the body to preferentially bind iron in ferritin rather than transferrin as *M. tuberculosis* can compete with TF and LTF for the binding of iron (Gobin & Horwitz, 1996). Therefore, it is beneficial to the system if most iron is bound in ferritin which the bacterium cannot access as easily.

The third and fourth objectives recognised were to genotype all individuals for any identified novel and/or previously discovered variants in the promoter region of the 5 genes under

---

investigation, and to determine allele and genotype frequencies for comparison respectively. One of the limitations presented here is that the *LTF* gene could not be genotyped due to difficulties in amplifying the fragments and differentiating the different genotypes on HEX-SSCP gels. In future this gene should be investigated as it is likely to play a role in *M. tuberculosis* susceptibility indicated by its direct interaction with the pathogen (Olakanmi *et al.*, 2007). It is suggested that the promoter region of this gene be genotyped by semi-automated bi-directional sequencing analysis to limit redundancies and genotyping difficulty.

The genotyping of some other fragments did involve HEX-SSCP analysis which is a system with only 70-100% sensitivity which can be influenced by number a of factors such as temperature, DNA fragment size, gel composition, and the fluorescent stain used (Fujita & Silver, 1994; Bonner & Ballard, 1999). The variation for these variables was minimised where possible. Funding was limited for this project and it is advised that future research should utilise more sensitive genotyping methods such as RFLP or semi-automated bi-directional sequencing analysis should more funding be available. Genotyping efficiency could also be increased as not all samples were genotyped for each variant (as indicated by the *n* values in Table 3.2). Many of the variants were found to deviate from HWE but this can most likely be attributed to the small sample size or to a true disease association for some of the variants (Schaid & Jacobsen, 1999).

The genotype and allele frequencies were determined for the genotyped variants as shown in Table 3.2. The variants that were found to be statistically significant between the patient and control groups were : **CYBRDI**: -1849 T/G, -1813 C/T, -1749 del (T)<sub>6</sub>G, -1477 G/A, -1459 T/C, -1346 T/C, -1272 T/C, -238 A/G and -166 C/G; **HFE**: -561 A/G. These variants were used to fulfil the fifth identified objective which was to perform *in silico* analyses to determine their effect on gene expression.

The majority of the TFBSs predicted to be abolished or created in the presence or absence of the *CYBRDI* promoter region variants could lead to the increased expression of the gene. If *CYBRDI* expression is implicated in *M. tuberculosis* susceptibility then this result would be expected. The increase would lead to excess iron absorption and this, in turn, would cause the total iron levels within the body to rise. This could potentially increase an individual's risk for TB infection as the pathogen thrives in an iron rich host (Schaible & Kaufmann, 2004). This is coupled with the results from the *HFE in silico* analyses which also show a general

---

trend of TFBSs that could lead to the increase of gene expression. The increased activity of *HFE* would lead to the decreased ability of TF to transport iron into cells but would not affect LTF which mainly transports iron into macrophages and other white blood cells (Sánchez *et al.*, 1992). This could potentially lead to the increased iron availability within macrophages which is advantageous for *M. tuberculosis* infection.

The results of the *in silico* analyses are supported by gene expression studies performed in our laboratory (unpublished data). Increased expression of both *CYBRDI* and *HFE* in the same TB cohort are currently observed. From these results it can be determined that more research is required as to the exact mechanism by which these genes, their protein products, and the variants identified play a role in *M. tuberculosis* susceptibility. It should be noted that haplotype analyses may be performed in future to determine the compound effect of variants on disease susceptibility.

The results of this study suggest that a link may exist between variants in the promoter region of iron metabolism genes and *M. tuberculosis* susceptibility. This emphasises the need for further investigation into the complex relationship between iron metabolism and pathogenic infection. This research can possibly contribute to our understanding of the interactions that occur and the genetic factors at play when distinguishing individuals in terms of their susceptibility to this pathogen. Research such as this could assist in the discovery of alternative treatment methods – such as iron chelation therapy suggested by Cronje *et al.* in 2005 – which could circumnavigate the need for drug treatments which have adverse drug reactions and create difficulty in terms of adherence and therefore the proliferation of multiple drug-resistant mycobacteria.

## **Chapter 5**

### **REFERENCES**

---

## 5. REFERENCES

### 5.1 GENERAL REFERENCES

- Allen, K. J., Gurrin, L. C., Constantine, C. C., Osborne, N. J., Delatycki, M. B., Nicoll, A. J., McLaren, C. E., Bahlo, M., Niselle, A. E., Vulpe, C. D., Anderson, G. J., Southey, M. C., Giles, G. G., English, D. R., Hopper, J. L., Olynyk, J. K., Powell, L. W. and Gertig, D. M. 2008. Iron-Overload-Related Disease in *HFE* Hemochromatosis. *New England Journal of Medicine* 358: 221-230.
- Bagu, E. T. and Santos, M. M. 2011. Friend of GATA suppresses the GATA-induced transcription of hepcidin in hepatocytes through a GATA-regulatory element in the *HAMP* promoter. *Journal of Molecular Endocrinology* 47: 299-313.
- Baker, E. N. and Baker, H. M. 2005. Molecular structure, binding properties and dynamics of lactoferrin. *Cellular and Molecular Life Sciences* 62: 2531-2539.
- Baker, M. A., Wilson, D., Wallengren, K., Sandgren, A., Iartchouk, O., Broodie, N., Goonesekera, S. D., Sabeti, P. C. and Murray, M. B. 2012. Polymorphisms in the gene that encodes the iron transport protein ferroportin 1 influence susceptibility to tuberculosis. *The Journal of Infectious Diseases* 205(7): 1043-1047.
- Bellamy, R., Beyers, N., McAdam, K. P. W. J., Ruwende, C., Gie, R., Samaai, P., Bester, D., Meyer, M., Corrah, T., Collin, M., Camidge, D. R., Wilkinson, D., Hoal-van Helden, E., Whittle, H. C., Amos, W., van Helden, P. and Hill, A. V. S. 2000. Genetic susceptibility to tuberculosis in Africans: A genome-wide scan. *Proceedings of the National Academy of Sciences* 97(14): 8005-8009.
- Bem, R. A., van Woensel, J. B. M., Bos, A. P., Koski, A., Farnand, A. W., Domachowske, J. B., Rosenberg, H. F., Martin, T. R. and Matute-Bello, G. 2009. Mechanical ventilation enhances lung inflammation and caspase activity in a model of mouse pneumovirus infection. *The American Journal of Physiology – Lung Cellular and Molecular Physiology* 296: L46-L56.
- Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M. and Eisen, M. B. 2002. Exploiting transcription factor binding site clustering to

- 
- identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proceedings of the National Academy of Sciences* 99(2): 757-762.
- Bhat, R. T., Worley, P. F., Cole, A. J. and Baraban, J. M. 1992. Activation of the zinc finger encoding gene *krox-20* in adult rat brain: comparison with *zif268*. *Molecular Brain Research* 13: 263-266.
- Bonner, M. R. and Ballard, L. W. 1999. Considerations in adding mutation detection services to a sequencing core facility. *Journal of Biomolecular Techniques* 10: 177-186.
- Bossone, S. A., Asselin, C., Patel, A. J. and Marcu, K. B. 1992. MAZ, a zinc-finger protein, binds to *c-MYC* and *C2* gene sequences regulating transcriptional initiation and termination. *Proceedings of the National Academy of Sciences* 89: 7452-7456.
- Bjorstad, A. 2009. Functional Dualism of Antimicrobial Host Defence Peptides. Doctoral Thesis. Department of Rheumatology and Inflammation Research, Institute of Medicine, University of Gothenburg, Göteborg, Sweden.
- Camirero, J. A., Sotgiu, G., Zumla, A. and Migliori, G. B. 2010. Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis. *The Lancet* 10: 621-629.
- Camus, J. C., Pryor, M. J., Medigue, C. and Cole, S. T. 2002. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 148: 2967-2973.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A. M., Taylor, M. S., Engström, P. G., Frith, M. C., Forrest, A. R. R., Alkema, W. B., Tan, S. L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S. M., Wells, C. A., Orlando, V., Wahlestedt, C., Liu, E. T., Harbers, M., Kawai, J., Bajic, V. B., Hume, D. A. and Hayashizaki, Y. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics* 38(6): 626-635.
- Causse, M., Ruiz, P., Gutierrez-Aroca, J. B. and Casal, M. 2011. Comparison of Two Molecular Methods for Rapid Diagnosis of Extrapulmonary Tuberculosis. *Journal of Clinical Microbiology* 49(8): 3065-3067.

- 
- Chaudhary, J., Cupp, A. S. and Skinner, M. K. 1997. Role of basic helix-loop-helix transcription factors in Sertoli cell differentiation: identification of an E-box response element in the transferrin promoter. *Endocrinology* 138(2): 667-675.
- Chavrier, P., Janssen-Timmen, U., Mattéi, M-G., Zerial, M., Bravo, R. and Charnay, P. 1989. Structure, chromosome location, and expression of the mouse zinc finger gene *Krox-20*: multiple gene products and coregulation with the proto-oncogene *c-fos*. *Molecular and Cellular Biology* 9(2): 787-797.
- Chrousos, G. P. and Kino, T. 2005. Intracellular glucocorticoid signaling: a formerly simple system turns stochastic. *Science Signaling* 2005(304): pe48.
- Coffer, P. J. and Burgering, M. T. 2004. Forkhead-box transcription factors and their role in the immune system. *Nature Reviews Immunology* 4: 889-899.
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., Tekaiia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T. and Gentles, S. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393:537-544.
- Collins, J. F. and Hu, Z. 2007. Promoter analysis of intestinal genes induced during iron deprivation reveals enrichment of conserved SP1-like binding sites. *BioMed Central Genomics* 8: 420.
- Collins, H. L. 2008. Withholding iron as a cellular defense mechanism – friend or foe? *European Journal of Immunology* 38: 1803-1806.
- Comstock, G. W. 1978. Tuberculosis in twins: a re-analysis of the Proffit survey. *The American Review of Respiratory Disease* 117(4): 621-624.
- Cowie, D. 2013. Iron and TB pathogenesis. Department of Genetics. Stellenbosch University, Stellenbosch, Western Cape.
- Cronjé, L., Edmondson, N., Eisenach, K. D. and Bornman, L. 2005. Iron and iron chelating agents modulate *Mycobacterium tuberculosis* growth and monocyte-macrophage viability and effector functions. *FEMS Immunology and Medical Microbiology* 45: 103-112.

- 
- De Domenico, I., Ward, D. M., Langelier, C., Vaughn, M. B., Nemeth, E., Sundquist, W. I., Ganz, T., Musci, G. and Kaplan, J. The Molecular Mechanism of Hepcidin-mediated Ferroportin Down-Regulation. *Molecular Biology of the Cell* 18: 2569-2578.
- Dhar, S. K., Xu, Y., Chen, Y. and St.Clair, D. K. 2004. Specificity protein 1-dependent p53-mediated suppression of human manganese superoxide dismutase gene expression. *The Journal of Biological Chemistry* 281(31): 21698-21709.
- Dillon, D. C., Alderson, M. R., Day, C. H., Bement, T., Campos-Neto, A., Skeiky, Y. A. W., Vedvick, T., Badaro, R., Reed, S. G. and Houghton, R. 2000. Molecular and Immunological Characterization of *Mycobacterium tuberculosis* CFP-10, an Immunodiagnostic Antigen Missing in *Mycobacterium bovis* BCG. *Journal of Clinical Microbiology* 39(9):3285-3290.
- Donovan, A., Brownlie, A., Zhou, Y., Shepard, J., Pratt, S. J., Moynihan, J., Paw, B. H., Grejer, A., Barut, B., Zapata, Z., Law, T. C., Brugnara, C., Lux, S. E., Pinkus, G. S., Pinkus, J. L., Kingsley, P. D., Palis, J., Fleming, M. D., Andrews, N. C. and Zon, L. I. 2000. Positional cloning of zebrafish ferroportin1 identifies a conserved vertebrate iron exporter. *Nature* 403: 776-781.
- Dreher, D. and Junod, A. F. 1996. Role of oxygen free radicals in cancer development. *European Journal of Cancer* 32(1): 30-38.
- Dye, C., Scheele, S., Dolin, R., Pathania, G. and Raviglione, M. 1999. Global burden of tuberculosis. Estimated incidence, prevalence, and mortality by country. *The Journal of the American Medical Association* 282: 677-686.
- Ellison, R. T., Giehl, T. J. and LaForce, F. M. 1988. Damage of the outer membrane of enteric gram-negative bacteria by lactoferrin and transferrin. *Infection and Immunity* 56(11): 2774-2781.
- Fairbairn, I. P. 2004. Macrophage apoptosis in host immunity to mycobacterial infections. *Biochemical Society Transactions* 32(3): 496-498.
- Fares, A. 2012. Racial differences in susceptibility to infection by *Mycobacterium tuberculosis*. *Annals of Tropical Medicine and Public Health* 5(4): 307.
- Feder, J. N., Gnirke, A., Thomas, W., Tsuchihashi, Z., Ruddy, D. A., Basava, A., Dormishian, F., Domingo Jr, R., Ellis, M. C., Fullan, A., Hinton, L. M., Jones, N. L.,

- 
- Kimmel, B. E., Kronmal, G. S., Lauer, P., Lee, V. K., Loeb, D. B., Mapa, F. A., McClelland, E., Meyer, N. C., Mintier, G. A., Moeller, N., Moore, T., Morikang, E., Prass, C. E., Quintana, L., Starnes, S. M., Schatzman, R. C., Brunke, K. J., Drayna, D. T., Risch, N. J., Bacon, B. R. and Wolff, R. K. 1996. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nature Genetics* 13: 399-408.
- Ferreira, R., Ohneda, K., Yamamoto, M. and Philipsen, S. 2005. GATA1 function, a paradigm for transcription factors in hematopoiesis. *Molecular and Cellular Biology* 25(4): 1215-1227.
- Finkelstein, R. A., Sciortino, C. V. and McIntosh, M. A. 1983. Role of iron in microbe-host interactions. *Reviews of Infectious Diseases* 5: S759-S777.
- Fujita, K. and Silver, J. 1994. Single-strand conformation polymorphism. *Genome Research* 4: S137-S140.
- Gadkar-Sable, S., Shah, C., Rosario, G., Sachdeva, G. and Puri, C. 2005. Progesterone receptors: various forms and function in reproductive tissues. *Frontiers in Bioscience* 10: 2118-2130.
- Ganz, T. 2003. Heparin, a key mediator of iron metabolism and mediator of anemia of inflammation. *Blood* 102(3): 783-788.
- Ganz, T. 2007. Molecular control of iron transport. *The Journal of the American Society of Nephrology* 18(2): 394-400.
- Garcia, A. D., Ostapchuk, P. and Hearing, P. 1993. Functional interaction of nuclear factors EF-C, HNF-4, and RXR alpha with hepatitis B virus enhancer I. *Journal of Virology* 67(7): 3940-3950.
- Geissler, C. and Singh, M. 2011. Iron, Meat and Health. *Nutrients* 3: 283-316.
- Gobin, J. and Horwitz, M. A. 1996. Exochelins of *Mycobacterium tuberculosis* remove iron from human iron-binding proteins and donate iron to mycobactins in the *M. tuberculosis* cell wall. *The Journal of Experimental Medicine* 183: 1527-1532.
- Gordeuk, V. R., McLaren, C. E., MacPhil, A. P., Deichsel, G. and Bothwell, T. H. 1996. Associations of iron overload in Africa with hepatocellular carcinoma and tuberculosis: Strachan's 1929 Thesis Revisited. *Blood* 87(8): 3470-3476.

- 
- Goudet, J., Raymond, M., de Meeüs, T. and Rousset, F. 1996a. Testing differentiation in diploid populations. *Genetics* 144: 1933-1940.
- Gregersen, T. 1978. Rapid method for distinction of gram-negative from gram-positive bacteria. *Applied Microbiology and Biotechnology* 5(2): 123-127.
- Gronostajski, R. M. 2000. Roles of NFI/CTF gene family in transcription and development. *Gene* 249: 31-45.
- Haarman, E. G., Kaspers, G. J. L., Pieters, R., Rottier, M. M. A. and Veerman, A. J. P. Glucocorticoid receptor alpha, beta and gamma expression vs *in vitro* glucocorticoid resistance in childhood leukemia. *Leukemia* 18: 530-537.
- Haile, D. J. 2000. Assignment of Slc11a3 to mouse chromosome 1 band 1B and SLC11A3 to human chromosome 2q21 by *in situ* hybridization. *Cytogenetics and Cell Genetics* 88: 328-329.
- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41:95-98.
- Halliwell, B. and Gutteridge, J. M. C. 1984. Oxygen toxicity, oxygen radicals, transition metals and disease. *Biochemical Journal* 219: 1-14.
- Hentze, M. W. and Kuhnt, L. C. 1996. Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. *Proceedings of the National Academy of Sciences* 93: 8175-8182.
- Hentze, M. W., Muckenthaler, M. U. and Andrews, N. C. 2004. Balancing acts: molecular control of mammalian iron metabolism. *Cell* 117: 285-297.
- Herzog, H. 1998. History of Tuberculosis. *Respiration* 65:5-15.
- Hochberg, Y. and Benjamini, Y. 1990. More powerful procedures for multiple significance testing. *Statistics in Medicine* 9: 811-818.
- Horiuchi, Y., Higuchi, T., Tatsumi, K., Takakura, K., Fujii, S. and Konishi, I. 2009. Lactoferrin is associated with a decrease in oocyte depletion in mice receiving cyclophosphamide. *Fertility and Sterility* 91(5): 2069-2078.
- Hsu, T., Trojanowska, M. and Watson, D. K. 2004. Ets proteins in biological control and cancer. *Journal of Cellular Biochemistry* 91: 896-903.

- 
- Hubbard, T. J. P., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Herrero, J., Holland, R., Howe, K., Howe, K., Johnson, N., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Melsopp, C., Megy, K., Meidl, P., Overduin, B., Parker, A., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Severin, J., Slater, G., Smedley, D., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wood, M., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Flicek, P., Kasprzyk, A., Proctor, G., Searle, S., Smith, J., Ureta-Vidal A. and Birney, E. 2007. Ensembl. *Nucleic Acids Research* 35(Database Issue):D610-D617.
- Huebner, R. E., Schein, M. F. and Bass, J. B. 1993. The tuberculin skin test. *Clinical Infectious Diseases* 17(6): 968-975.
- Jamshidi-Parsian, A., Dong, Y., Zheng, X., Zhou, H. S., Zacharias, W. and McMasters, K. M. 2005. Gene expression profiling of E2F-1-induced apoptosis. *Gene* 344: 67-77.
- Jones, K. A., Yamamoto, K. R. and Tjian, R. 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. *Cell* 42: 559-572.
- Juven-Gershon, T. and Kadonaga, J. T. 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology* 339: 225-229.
- Kaczynski, J., Cook, T. and Urrutia, R. 2003. Sp1- and Krüppel-like transcription factors. *Genome Biology* 4(2): 206.1-203.8.
- Kageyama, R., Merlino, G. T. and Pastan, I. 1988. A transcription factor active on the epidermal growth factor receptor gene. *Proceedings of the National Academy of Sciences* 85: 5016-5020.
- Karim, F. D., Urness, L. D., Thummel, C. S., Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C., Maki, R. A., Gunther, C. V., Nye, J. A. and Graves, B. J. 1990. The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. *Genes and Development* 4: 1451-1453.
- Katoh, M. and Katoh, M. 2004. Human FOX gene family (Review). *International Journal of Oncology* 25(5): 1495-1500.

- 
- Kawamata, T., Tooyama, I., Yamada, T., Walker, D. G. and McGeer, P. L. 1993. Lactotransferrin immunocytochemistry in Alzheimer and normal human brain. *The American Journal of Pathology* 142(2): 1574-1585.
- Kotze, M. J., Theart, L., Callis, M., Peeters, A. V., Thiart, R. and Langenhoven, E. 1995. Nonradioactive multiplex PCR screening strategy for the simultaneous detection of multiple low-density receptor gene mutations. *PCR Methods and Applications* 4:352-356.
- Kumar, V., Abbas, A. K., Fausto, N. and Mitchell, R. N. 2007. *Robbins Basic Pathology* (8<sup>th</sup> edition). Saunders Elsevier. 516-522.
- Lamonerie, T., Tremblay, J. J., Lactot, C., Therrien, M., Gauthier, Y. and Drouin, J. 1996. Ptx-1, a *bicoid*-related homeo box transcription factor involved in transcription of the pro-opiomelanocortin gene. *Genes and Development* 10: 1284-1295.
- Lawn, S. D. and Zumla, A. 2011. Tuberculosis. *The Lancet* 378: 57-72.
- Lebron, J. A., Bennett, M. J., Vaughn, D. E., Chirino, A. J., Snow, P. M., Mintier, G. A., Feder, J. N. and Bjorkman, P. J. 1998. Crystal structure of the haemochromatosis protein HFE and characterization of its interaction with transferrin receptor. *Cell* 93(1): 111-123.
- Levy, J. E., Jin, O., Fujiwara, Y., Kuo, F. and Andrews, N. C. 1999. Transferrin receptor is necessary for development of erythrocytes and the nervous system. *Nature Genetics* 21: 396-399.
- Li, J., Ning, G. and Duncan, S. A. 2000. Mammalian hepatocyte differentiation requires the transcription factor HNF-4 $\alpha$ . *Genes and Development* 14: 464-474.
- Li, M. and Li, C. 2008. Assessing departure from Hardy-Weinberg equilibrium in the presence of disease association. *Genetic Epidemiology* 32: 589-599.
- Lipsitch, M. and Sousa, A. O. 2002. Historical Intensity of Natural Selection for Resistance to Tuberculosis. *Genetics* 161: 1599-1607.
- Liu, X. B., Yang, F. and Haile, D. J. 2005. Functional consequences of ferroportin 1 mutations. *Blood Cells, Molecules, and Diseases* 35: 33-46.
- Lösel, R. and Wehling, M. 2003. Nongenomic actions of steroid hormones. *Nature Reviews Molecular Cell Biology* 4: 46-56.

- 
- Lounis, N., Truffot-Pernot, C., Grosset, J., Gordeuk, V. R. and Boelaert, J. R. 2001. Iron and *Mycobacterium tuberculosis* infection. *Journal of Clinical Virology* 20: 123-126.
- Madison, B. M. 2001. Application of stains in clinical microbiology. *Biotechnic and Histochemistry* 76(3): 119-125.
- Martin, H. M., Hancock, J. T., Salisbury, V. and Harrison, R. 2004 Role of Xanthine Oxidoreductase as an Antimicrobial Agent. *Infection and Immunity* 72(9): 4933-4939.
- McGee, S.L. and Hargreaves, M. 2004. Exercise and myocyte enhancer factor 2 regulation in human skeletal muscle. *Diabetes* 53: 1208-1214.
- McKie, A. T., Barrow, D., Latunda-Dada, G. O., Rolfs, A., Sager, G., Mudaly, E., Mudaly, M., Richardson, C., Barlow, D., Bomford, A., Peters, T. J., Raja, K. B., Shirali, S., Hediger, M. A., Farzaneh, F. and Simpson, R. J. 2001. An Iron-Regulated Ferric Reductase Associated with the Absorption of Dietary Iron. *Science* 291(5509): 1755-1759.
- Metcalfe, J. Z., Everett, C. K., Steingart, K. R., Cattamanchi, A., Huang, L., Hopewell, P. C. and Pai, M. 2011. Interferon- $\gamma$  Release Assays for Active Pulmonary Tuberculosis Diagnosis in Adults in Low- and Middle-Income Countries: Systematic Review and Meta-analysis. *The Journal of Infectious Diseases* 204: S1120 – S1129.
- Miller, M. P. 1997. Tools for population genetic analysis (TFPGA) 1.3. A windows program for the analysis of allozyme and molecular population genetic data. *Crop Science* 40(6):1521-1528.
- Miller, S. A., Dykes, D. D. and Polesky, H. F. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* 16(3): 1215.
- Miret, S., Simpson, R. J. and McKie, A. T. 2003. Physiology and molecular biology of dietary iron absorption. *Annual Review of Nutrition* 23: 283-301.
- Moran, M.D. 2003. Arguments for rejecting the sequential Bonferroni in ecological studies. *OIKOS* 100(2): 403-405.
- Moskwa, P., Lorentzen, D., Excoffon, K. J. D. A., Zabner, J., McCray, P. B., Nauseef, W. M., Dupuy, C. and Banfi, B. 2007. A novel host defence system of airways is defective in cystic fibrosis. *American Journal of Respiratory and Critical Care Medicine* 175(2): 174-183.

- 
- Nakagawa, S. 2004. A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioural Ecology* 15(6): 1044-1045.
- Nandar, W. and Connor, J. R. 2011. HFE Gene Variants Affect Iron in the Brain. *The Journal of Nutrition* 141(4): 729S-739S.
- Nicas, M., Nazaroff, W. W. and Hubbard, A. 2005. Toward understanding the risk of secondary airborne infection: emission of respirable pathogens. *Journal of Occupational and Environmental Hygiene* 2(3): 143-154.
- Nielsen, D.A., Ehm, M. G. and Weir, B. S. 1999. Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *American Journal of Human Genetics* 63: 1531-1540.
- Oakhill, J. S., Marritt, S. J., Garcia Gareta, E., Cammack, R. and McKie, A. T. 2008. Functional characterization of human duodenal cytochrome *b* (Cybrd1): Redox properties in relation to iron and ascorbate metabolism. *Biochimica et Biophysica Acta* 1777: 260-268.
- Oakley, R. H., Sar, M. and Cidlowski, J. A. 1996. The human glucocorticoid receptor  $\beta$  isoform: expression, biochemical properties, and putative function. *The Journal of Biological Chemistry* 271(16): 9550-9559.
- Ohashi, Y., Ishida, R., Kojima, T., Goto, E., Matsumoto, Y., Watanabe, K., Ishida, N., Nakata, K., Takeuchi, T. and Tsubota, K. 2003. Abnormal protein profiles in tears with dry eye syndrome. *American Journal of Ophthalmology* 136(2): 291-299.
- Olakanmi, O., Schlesinger, L. S. and Britigan, B. E. 2007. Hereditary hemochromatosis results in decreased iron acquisition and growth by *Mycobacterium tuberculosis* within human macrophages. *Journal of Leukocyte Biology* 81(1): 195-204.
- Ong, S. T., Ho, J. Z. S. and Ding, J. L. 2006. Iron-withholding strategy in innate immunity. *Immunobiology* 211: 295-314.
- Orphanides, G., Lagrange, T. and Reinberg, D. 1996. The general transcription factors of RNA polymerase II. *Genes and Development* 10: 2657-2683.
- Orsi, N. 2004. The antimicrobial activity of lactoferrin: Current status and perspectives. *BioMetals* 17: 189-196.

- 
- Pamer, E. and Cresswell, P. 1998. Mechanisms of MHC class I-restricted antigen processing. *Annual Review of Immunology* 16: 323-358.
- Papanikolaou, G. and Pantopoulos, K. 2005. Iron Metabolism and Toxicity. *Toxicology and Applied Pharmacology* 202: 199-211.
- Papanikolaou, G., Tzilianos, M., Christakis, J. I., Bogdanos, D., Tsimirika, K., MacFarlane, J., Goldberg, Y. P., Sakellaropoulos, N., Ganz, T. and Nemeth, E. 2005. Hepcidin in iron overload disorders. *Blood* 105: 4103-4105.
- Park, C. H., Valore, E. V., Waring, A. J. and Ganz, T. 2001. Hepcidin, a Urinary Antimicrobial Peptide Synthesized in the Liver. *The Journal of Biological Chemistry* 276(11): 7806-7810.
- Perez-Osorio, A. C., Boyle, D. S., Ingham, Z. K., Ostash, A., Gautom, R. K., Colombel, C., Houze, Y. and Leader, B. T. 2012. Rapid Identification of Mycobacteria and Drug-Resistant *Mycobacterium tuberculosis* by Use of a Single Multiplex PCR and DNA Sequencing. *Journal of Clinical Microbiology* 50(2): 326-336.
- Pietrangelo, A. 2004. The ferroportin disease. *Blood Cells, Molecules, and Diseases* 32: 131-138.
- Ponka, P. 1997. Tissue-specific regulation of iron metabolism and heme synthesis: distinct control mechanisms in erythroid cells. *Blood* 89(1): 1-25.
- Ponka, P., Beaumont, C. and Richardson, D. R. 1998. Function and regulation of transferrin and ferritin. *Seminars in Hematology* 35: 35-54.
- Prism, G. 2007. Version 5.01. *GraphPad Software Inc.: San Diego, CA, USA*.
- Quinlan, G. J., Evans, T. W. and Gutteridge, J. M. C. 2002. Iron and the redox status of the lungs. *Free Radical Biology and Medicine* 33: 1306-1313.
- Raffin, S. B., Woo, C. H., Roost, K. T., Price, D. C. and Schmid, R. 1974. Intestinal absorption of hemoglobin iron-heme cleavage by mucosal heme oxygenase. *The Journal of Clinical Investigation* 54: 1344-1352.
- Ramsay, R. G., Barton, A. L. and Gonda, T. J. Targeting c-Myb expression in human disease. *Expert Opinion on Therapeutic Targets* 7(2): 235-248.
- Ratledge, C. 2004. Iron, mycobacteria and tuberculosis. *Tuberculosis* 84: 110-130.

- 
- Richardson, D. R. and Ponka, P. 1997. The molecular mechanisms of the metabolism and transport of iron in normal and neoplastic cells. *Biochimica et Biophysica Acta* 1331: 1-40.
- Rothel, J. S. and Andersen, P. 2005. Diagnosis of latent *Mycobacterium tuberculosis* infection: is the demise of the Mantoux test imminent? *Expert Review of Anti-Infective Therapy* 3(6): 981-993.
- Rousset, F. 2008. GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8(1): 103-106.
- Roy, A. L. 2001. Biochemistry and biology of the inducible multifunctional transcription factor TFII-I. *Gene* 274: 1-13.
- Sánchez, L., Calvo, M. and Brock, J. H. 1992. Biological role of lactoferrin. *Archive of Disease in Childhood* 67: 657-661.
- Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y. and Hume, D. A. 2007. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature Reviews Genetics* 8: 424-436.
- Santos, P. C. J. L., Cancado, R. D., Pereira, A. C., Schettert, I. T., Soares, R. A. G., Pagliusi, R. A., Hirata, R. D. C., Teixeira, A. C., Figueiredo, M, S., Chiattonne, C. S., Krieger, J. E. and Guerra-Shinohara, E. M. 2011. Hereditary hemochromatosis: Mutations in genes involved in iron homeostasis in Brazilian patients. *Blood Cells, Molecules, and Diseases* 46: 302-307.
- Saunders, B. M. and Britton, W. J. 2007. Life and death in the granuloma: immunopathology of tuberculosis. *Immunology and Cell Biology* 82: 103-111.
- Schaible, U. E., Collins, H. L., Priem, F. and Kaufmann, S. H. E. 2002. Correction of the iron overload defect in  $\beta$ -2-microglobulin knockout mice by lactoferrin abolishes their increased susceptibility to tuberculosis. *Journal of Experimental Medicine* 196(11):1507-1513.
- Schaible, U. E. and Kaufmann, H. E. 2004. Iron and microbial infection. *Nature Reviews* 2: 946-953.
- Schiad, D. J. and Jacobsen, S. J. 1999. Biased tests of association: comparisons of allele frequencies when departing from Hardy-Weinberg proportions. *American Journal of Epidemiology* 149(8): 706-711.

- 
- Schaid, D. J., Batzler, A. J., Jenkins, G. D. and Hildebrandt, M. A. T. 2006. Exact tests of Hardy-Weinberg equilibrium and homogeneity of disequilibrium across strata. *The American Journal of Human Genetics* 79: 1071-1080.
- Sester, M., Sotgiu, G., Lange, C., Giehl, C., Girardi, E., Migliori, G. B., Bossink, A., Dheda, K., Diel, R., Dominguez, J., Lipman, M., Nemeth, J., Ravn, P., Winkler, S., Huitric, E., Sandgren, A. and Manissero, D. 2011. Interferon- $\gamma$  release assays for the diagnosis of active tuberculosis: a systematic review and meta-analysis. *European Respiratory Journal* 37(1): 100-111.
- Sharrocks, A. D. 2001. The ETS-domain transcription factor family. *Molecular Cell Biology* 2: 827-837.
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. and Sirotkin, K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research* 29(1):308-311.
- Simeone, A. 1998. *Otx1* and *Otx2* in the development and evolution of the mammalian brain. *The EMBO Journal* 17(23): 6790-6798.
- Singh, P. K., Parsek, M. R., Greenberf, E. P. and Welsh, M. J. 2002. A component of innate immunity prevents bacterial biofilm development. *Nature* 417: 552-555.
- Smith, N. H., Hewinson, R. G., Kremer, K., Brosch, R. and Gordon, S. V. 2009. Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. *Nature Reviews* 7: 537-544.
- Sow, F. B., Florence, W. C., Satoskar, A. R., Schlesinger, L. S., Zwilling, B. S. and Lafuse, W. P. 2007. Expression and localization of hepcidin in macrophages: a role in host defence against tuberculosis. *Journal of Leukocyte Biology* 82: 934-945.
- Stewart, G. R., Robertson, B. D. and Young, D. B. 2003. Tuberculosis: A problem with persistence. *Nature Reviews Microbiology* 1: 97-105.
- Strickland, N. J. 2013. *In silico* and functional analyses of the iron metabolism pathway. Department of Genetics. Stellenbosch University, Stellenbosch, Western Cape.
- Teng, C. T. 2002. Lactoferrin gene expression and regulation: an overview. *Biochemistry and Cell Biology* 80(1): 7-16.

- 
- Teo, S. S. S. and Shingadia, D. V. 2006. Does BCG have a role in tuberculosis control and prevention in the United Kingdom? *Archives of Disease in Childhood* 91(6): 529-531.
- Tran, T. N., Eubanks, S. K., Schaffer, K. J., Zhou, C. Y. J. and Linder, M. C. 1997. Secretion of ferritin by rat hepatoma cells and its regulations by inflammatory cytokines and iron. *Blood* 90: 4979-4989.
- Travis, A., Amsterdam, A., Belanger, C. and Grosschedl, R. 1991. LEF-1, a gene encoding a lymphoid-specific with protein, an HMG domain, regulates T-cell receptor  $\alpha$  enhancer function. *Genes and Development* 5: 880-894.
- Uramoto, H., Izumi, H., Nagatani, G., Ohmori, H., Nagasue, N., Ise, T., Yoshida, T., Yasumoto, K. and Kohno, K. 2003. Physical interaction of tumour suppressor p53/p73 with CCAAT-binding transcription factor 2 (CTF2) and differential expression of human high-mobility group 1 (HMG1) gene expression. *The Biochemical Journal* 371: 301-310.
- Van Vliet, J., Turner, J. and Crossley, M. 2000. Human Krüppel-like factor 8: a CACCC-box binding protein that associates with CtBP and represses transcription. *Nucleic Acids Research* 28(9): 1955-1962.
- Vaupel, P., Kallinowski, F. And Okunieff, P. 1989. Blood flow, oxygen and nutrient supply, and metabolic microenvironment of human tumors: a review. *Cancer research* 49: 6449-6465.
- Verelst, W. and Asard, H. 2003. A phylogenetic study of cytochrome b561 proteins. *Genome Biology* 4 (6): R38.
- Vergne, I., Chua, J., Lee, H-H., Lucas, M., Belisle, J. and Deretic, V. 2005. Mechanism of phagolysosome biogenesis block by viable *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences* 102(11): 4033-4038.
- Vincze, T., Posfai, J. and Roberts R. J. 2003. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Research* 31: 3688-3691.
- Wallace, D. F., Harris, J. M. and Subramaniam, V. N. 2010. Functional analysis and theoretical modelling of ferroportin reveals clustering of mutations according to phenotype. *American Journal of Physiology: Cell Physiology* 298(1): C75-C84.
- Wagner, D., Maser, J., Lai, B., Cai, Z., Barry III, C. E., Honer zu Bentrup, K., Russell, D. G. and Bermudez, L. E. 2005. Elemental Analysis of *Mycobacterium avium*-, *Mycobacterium*

- 
- tuberculosis*-, and *Mycobacterium smegmatis*-Containing Phagosomes Indicates Pathogen-Induced Microenvironments within the Host Cell's Endosomal System. *The Journal of Immunology* 174: 1491-1500.
- Wang, W., Zhong, J. and Wang, Y-Q. 2010. Comparative genomic analysis reveals evolutionary conservation of the *Pax* gene family. *Genes Genetics Systems* 85: 193-206.
- Ward, P. P., Zhou, X. and Conneely, O. M. 1996. Cooperative Interactions between the Amino- and Carboxyl-terminal Lobes Contribute to the Unique Iron-binding Stability of Lactoferrin. *The Journal of Biological Chemistry* 271(22): 12790-12794.
- Ward, R. J., Crichton, R. R., Taylor, D. L., Corte, L. D., Srai, S. K and Dexter, D. T. 2011. Iron and the immune system. *Journal of Neural Transmission* 118: 315-328.
- Weeber, E. J. and Sweatt, J. D. 2002. Molecular neurobiology of human cognition. *Neuron* 33: 845-848.
- Weizer-Stern, O., Adamsky, K., Margalit, O., Ashur-Fabian, O., Givol, D., Amariglio, N. and Rechavi, G. 2007. Hepcidin, a key regulator of iron metabolism, is transcriptionally activated by p53. *British Journal of Haematology* 138: 253-262.
- Wheeler, T. T., Hodgkinson, A. J., Prosser, C. G. and David, S. R. 2007. Immune Components of Colostrum and Milk – A Historical Perspective. *Journal of Mammary Gland Biology and Neoplasia* 12: 237-247.
- Wilson, S. M., Al-Suwaidi, Z., McNerney, R., Porter, J., and Drobniewski, F. 1997. Evaluation of a new rapid bacteriophage-based method for the drug susceptibility testing of *Mycobacterium tuberculosis*. *Nature medicine* 3(4): 465-468.
- Wolfe, N. D., Dunavan, C. P. and Diamond, J. 2007. Origins of major human infectious diseases. *Nature* 447: 279-283.
- World Health Organization. 2004. Assessing the iron status of populations. 2<sup>nd</sup> Edition. Geneva. WHO Press.
- World Health Organization. 2011. Global tuberculosis control: WHO report 2011. Geneva. WHO Press.
- World Health Organization. 2012. Global tuberculosis report 2012. Geneva. WHO Press.

- 
- Wilson, S. M., Al-Suwaidi, Z., McNerney, R., Porter, J. and Drobniewski, F. 1997. Evaluation of a new rapid bacteriophage-based method for the drug susceptibility testing of *Mycobacterium tuberculosis*. *Nature Medicine* 3(4): 465-468.
- Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., Krull, M., Matys, V., Michael, H., Ohnauer, R., Pruss, M., Schacherer, F., Thiele, S. and Urbach, S. 2001. The TRANSFAC system on gene expression regulation. *Nucleic Acids Research* 29: 281-283.
- Wu, G. D., Chen, L., Forslund, K. and Traber, P. G. 1994. Hepatocyte nuclear factor-1 $\alpha$  (HNF-1 $\alpha$ ) and HNF-1 $\beta$  regulate transcription via two elements in an intestine-specific promoter. *The Journal of Biological Chemistry* 269(25): 17080-17085.
- Wu, K-J., Polack, A. and Dalla-Favera, R. 1999. Coordinated regulation of iron-controlling genes, h-ferritin and IRP2, by c-Myc. *Science* 283(5402): 676-679.
- Wu, X. and Levine, A. J. 1994. p53 and E2F-1 cooperate to mediate apoptosis. *Proceedings of the National Academy of Sciences* 91: 3602-3606.
- Wurster, A. L., Tanaka, T. and Grusby, M. J. 2000. The biology of Stat4 and Stat6. *Oncogene* 19: 2577-2584.
- Xiao, C., Calado, D. P., Galler, G., Thai, T-H., Patterson, H. C., Wang, J., Rajewsky, N., Bender, T. P. and Rajewsky, K. 2007. MiR150 controls B cell differentiation by targeting the transcription factor c-Myb. *Cell* 131: 146-159.
- Yoshida, H., Matsui, T., Yamamoto, A., Okada, T. and Mori, K. 2001. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* 107: 881-891.
- Zaahl, M. G., Merryweather-Clarke, A. T., Kotze, M. J., van der Merwe, S., Warnich, L. and Robson, K. J. H. 2004. Analysis of genes implicated in iron regulation in individuals presenting with primary iron overload. *Human Genetics* 115: 409-417.
- Zakin, M. M. 1992. Regulation of transferrin gene expression. *The FASEB Journal* 6: 3253-3258.
- Zhang, A. S., Davies, P. S., Carlson, H. L. and Enns, C. A. 2003. Mechanisms of HFE-induced regulation of iron homeostasis: insights from the W81A HFE mutation. *Proceedings of the National Academy of Sciences* 100: 9500-9505.

## 5.2 ELECTRONIC REFERENCES

ALIBABA2: <http://www.gene-regulation.com/pub/programs/alibaba2>

dbSNP: <http://www.ncbi.nlm.nih.gov/projects/SNP/>

Ensembl: <http://www.ensembl.org>

GenBank: <http://www.ncbi.nlm.nih.gov/genbank>

GENEPOP: [kimura.univ-montp2.fr/~rousset/Genepop.htm](http://kimura.univ-montp2.fr/~rousset/Genepop.htm)

Graphpad Prism: [www.graphpad.com](http://www.graphpad.com)

HapMap (Online): <http://www.hapmap.org>

PROMO: [algen.lsi.upc.es/cgi-bin/promo\\_v3/promo/promoinit.cgi?dirDB=TF\\_8.3](http://algen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3)

TFBlast: [www.gene-regulation.com/cgi-bin/pub/programs/tfbblast/tfbblast.cgi](http://www.gene-regulation.com/cgi-bin/pub/programs/tfbblast/tfbblast.cgi)

TFPGA: [http://www.marksgeneticssoftware.net/vti\\_bin/shtml.exe/tfpga.htm](http://www.marksgeneticssoftware.net/vti_bin/shtml.exe/tfpga.htm)

TRANSFAC®7: <http://www.gene-regulation.com/pub/databases.html#transfac>

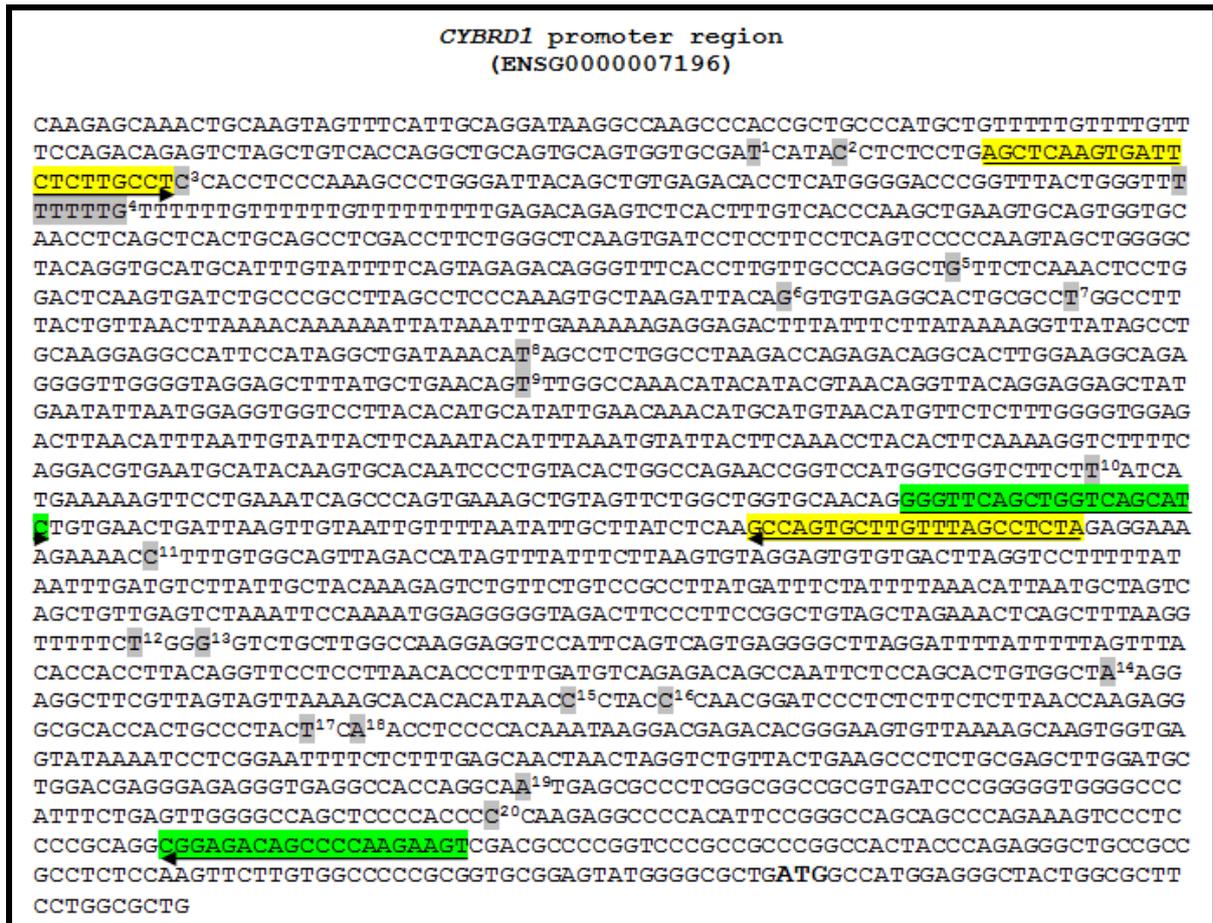
# **Addendum A**

---

## ADDENDUM A

### A.1 PROMOTER REGION SEQUENCES OF *CYBRD1*, *HAMP*, *HFE*, *LTF* AND *SLC40A1*

The sequences of each promoter region were sourced from the Ensembl ([www.ensembl.org](http://www.ensembl.org)) website. The region investigated for each gene is provided in Figures A.1 to A.5. Each figure contains the sequence investigated as well as the primers used (highlighted in various colours described in the key and in Tables 2.1 to 2.3), the primers' orientation, the Ensembl reference number (below the sequence title), the start site of transcription (the 'ATG' site is emphasised in bold), and the variants identified (underlined, assigned a number in superscript, and identified in the key to the figure). Arrows below the primers indicate their orientation – arrows pointing left indicate forward primers while arrows pointing right indicate reverse primers.



**Figure A.1** The promoter region investigated upstream of the *CYBRD1* gene

Key to Figure A.1:

Primers

Variants identified in the current study

**BP1 Sequence**

<sup>1</sup> -1849 T/G    <sup>2</sup> -1844 C/G    <sup>3</sup> -1813 C/G    <sup>4</sup> -1749 del (T)<sub>6</sub>G

**BP2 Sequence**

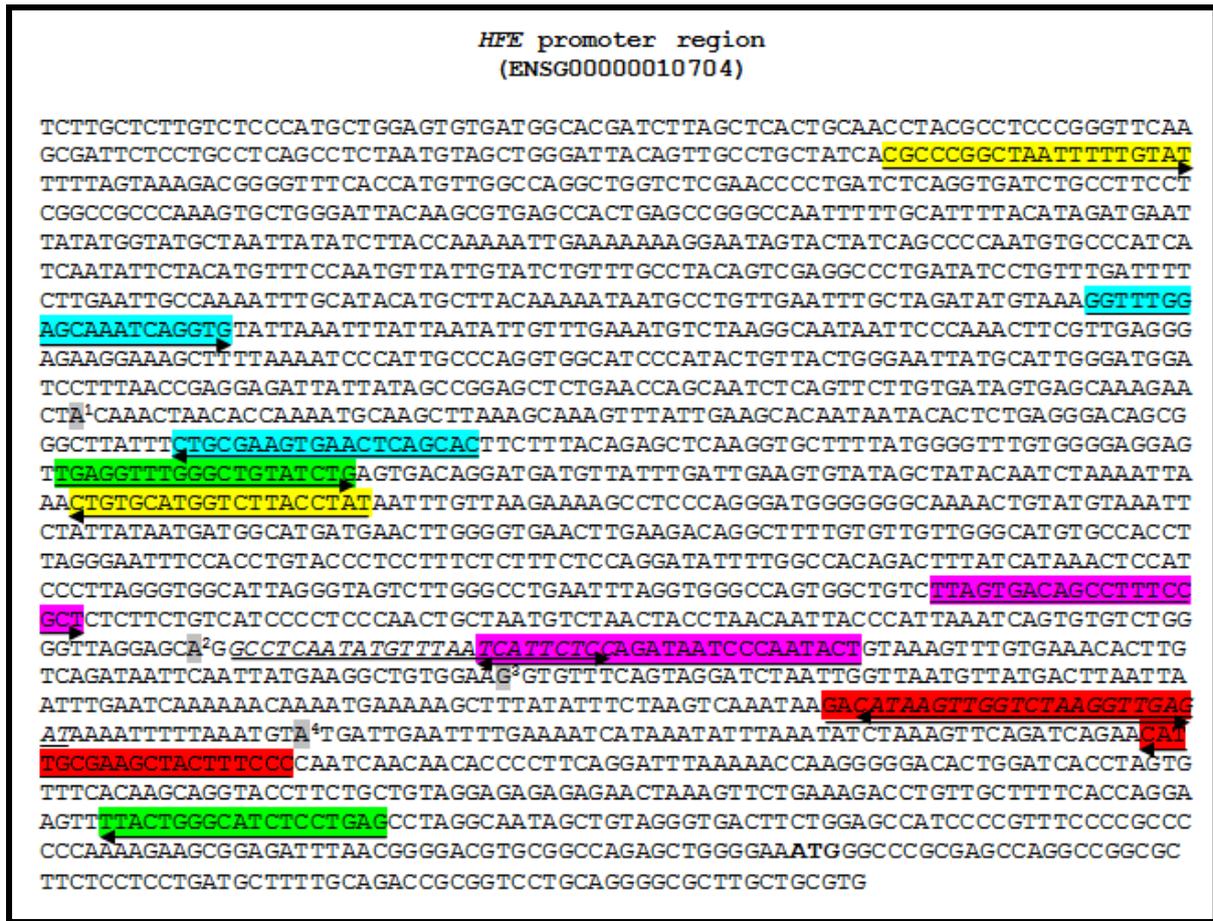
<sup>5</sup> -1540 G/A    <sup>6</sup> -1477 G/A    <sup>7</sup> -1459 T/C    <sup>8</sup> -1346 T/C

<sup>9</sup> -1272 T/C    <sup>10</sup> -1011 T/A    <sup>11</sup> -849 C/G    <sup>12</sup> -627 T/C

<sup>13</sup> -624 G/A    <sup>14</sup> -492 A/G    <sup>15</sup> -454 C/T    <sup>16</sup> -449 C/G

<sup>17</sup> -399 T/G    <sup>18</sup> -397 A/C    <sup>19</sup> -238 A/G    <sup>20</sup> -166 C/G





**Figure A.3 The promoter region investigated upstream of the *HFE* gene**

Key to Figure A.3:

Primers

Variants identified in the current study

HP1 Sequence <sup>1</sup>-1168 A/G <sup>2</sup>-561 A/G <sup>3</sup>-467 G/C <sup>4</sup>-331 A/C

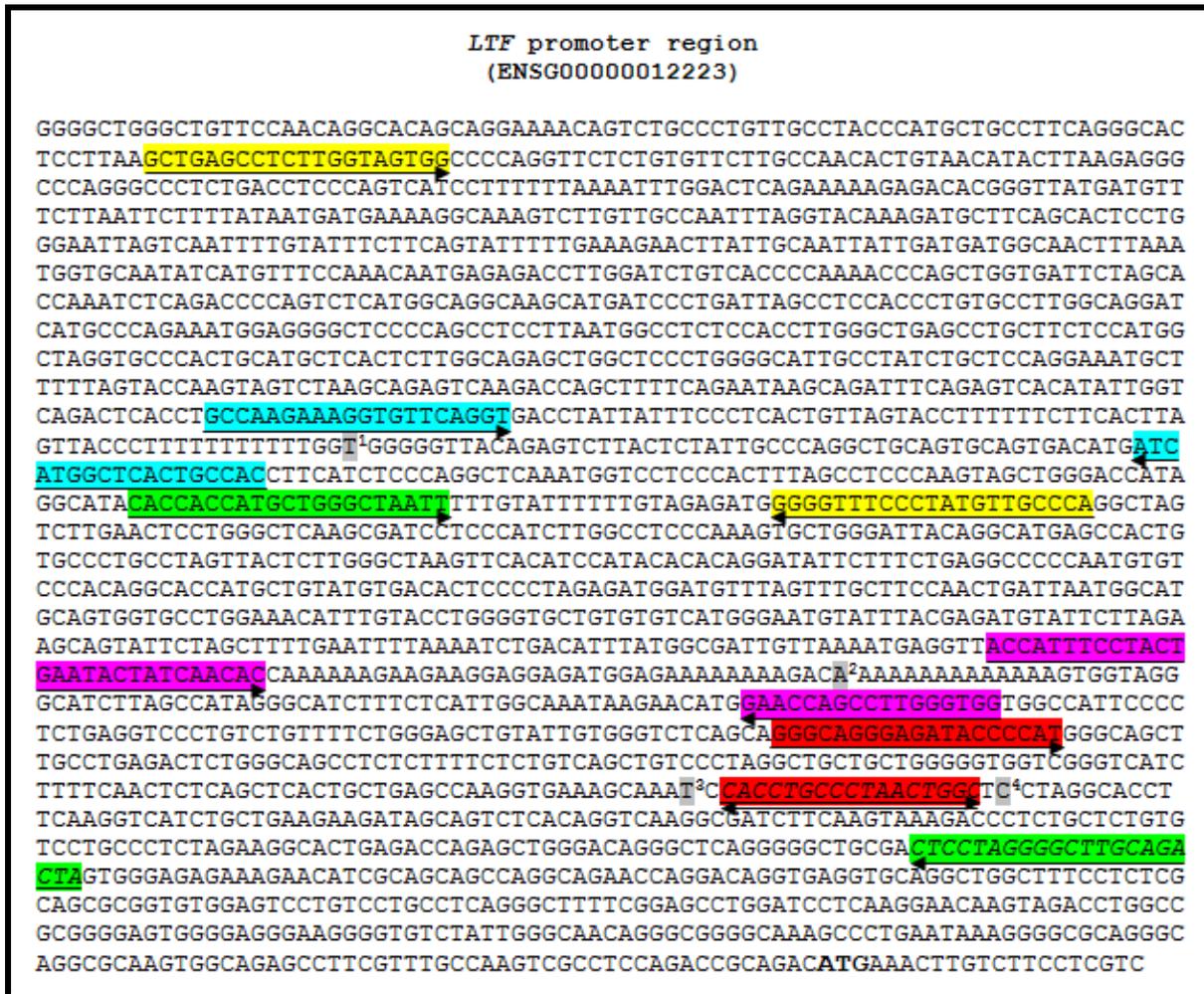
HP2 Sequence

HP1 (-1168 A/G)

HP2 (-561 A/G)

HP3 (-467 G/C)

HP4 (-331 A/C)



**Figure A.4** The promoter region investigated upstream of the *LTF* gene

Key to Figure A.4:

Primers

Variants identified in the current study

**LP1 Sequence**

<sup>1</sup>-1377 T/G    <sup>2</sup>-746 del AAA/ins AA    <sup>3</sup>-457 T/C    <sup>4</sup>-437 C/G

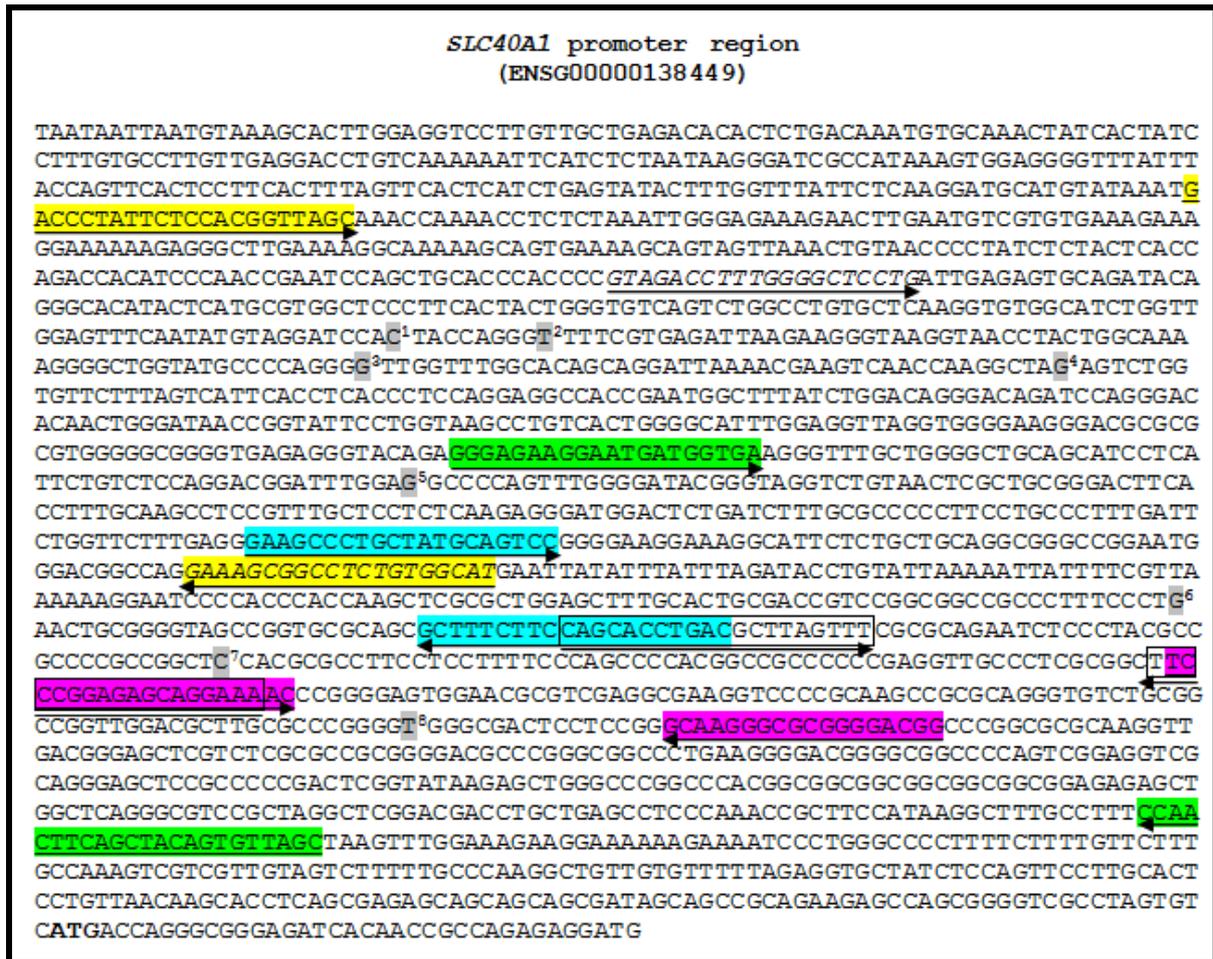
**LP2 Sequence**

**LP1 (-1377 T/G)**

**LP2 (-746 del AAA/ins AA)**

**LP3 (-457 T/C)**

**LP4 (-437 C/G)**



**Figure A.5 The promoter region investigated upstream of the *SLC40A1* gene**

Key to Figure A.5:

Primers

Variants identified in the current study

**SP1 Sequence**

<sup>1</sup>-1470 C/T    <sup>2</sup>-1461 T/C    <sup>3</sup>-1399 G/A    <sup>4</sup>-1355 G/C

**SP2 Sequence**

<sup>5</sup>-1098 G/A    <sup>6</sup>-750 G/A    <sup>7</sup>-662 C/T    <sup>8</sup>-501 T/C

*SP3 Sequence (5 variants)*

**SP1 (-750 G/A)**

**SP2 (-662 C/T)**

**SP3 (-501 T/C)**

## **Addendum B**

---

## ADDENDUM B

### B.1 VARIANT IDENTIFICATION

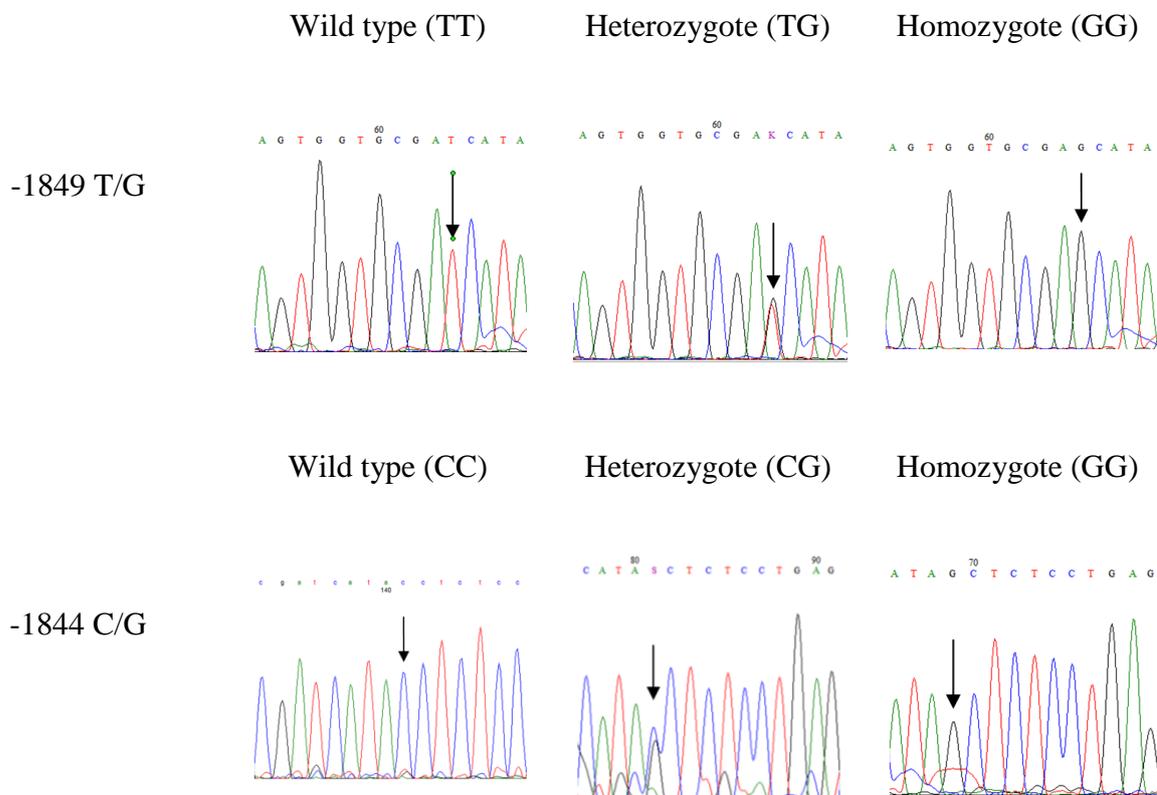
#### B.1.1 Chromatograms

Chromatograms were obtained from the Central Analytical Facility (CAF) from sequence reactions on PCR amplified samples using primers as indicated in Tables 2.1 to 2.3. All variants were identified in the initial screened group and chromatograms are provided. The titles of the variants that were excluded from further analyses are highlight in red. Horizontal arrows represent the point at which the base pair change occurs for the assigned variant.

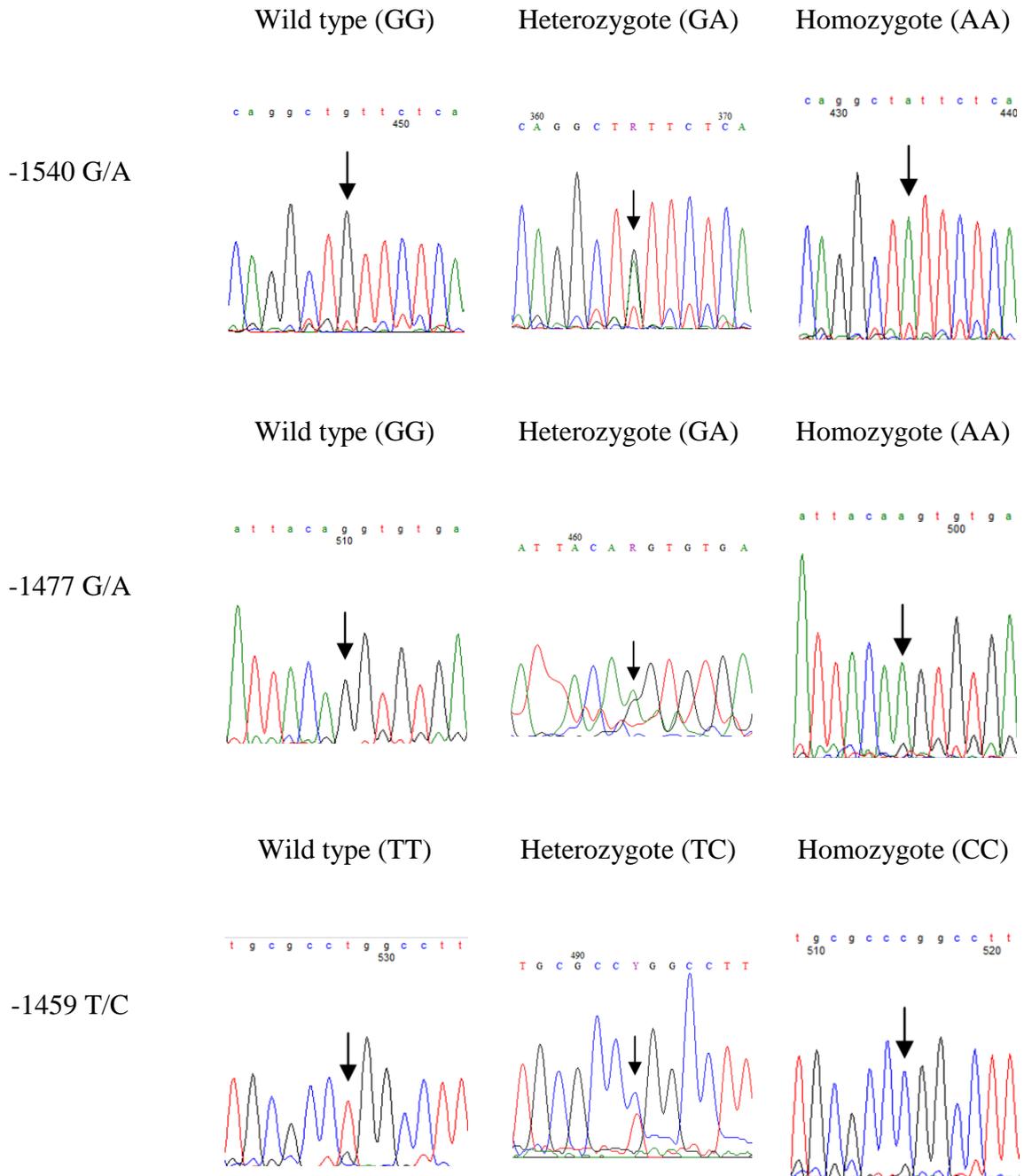
---

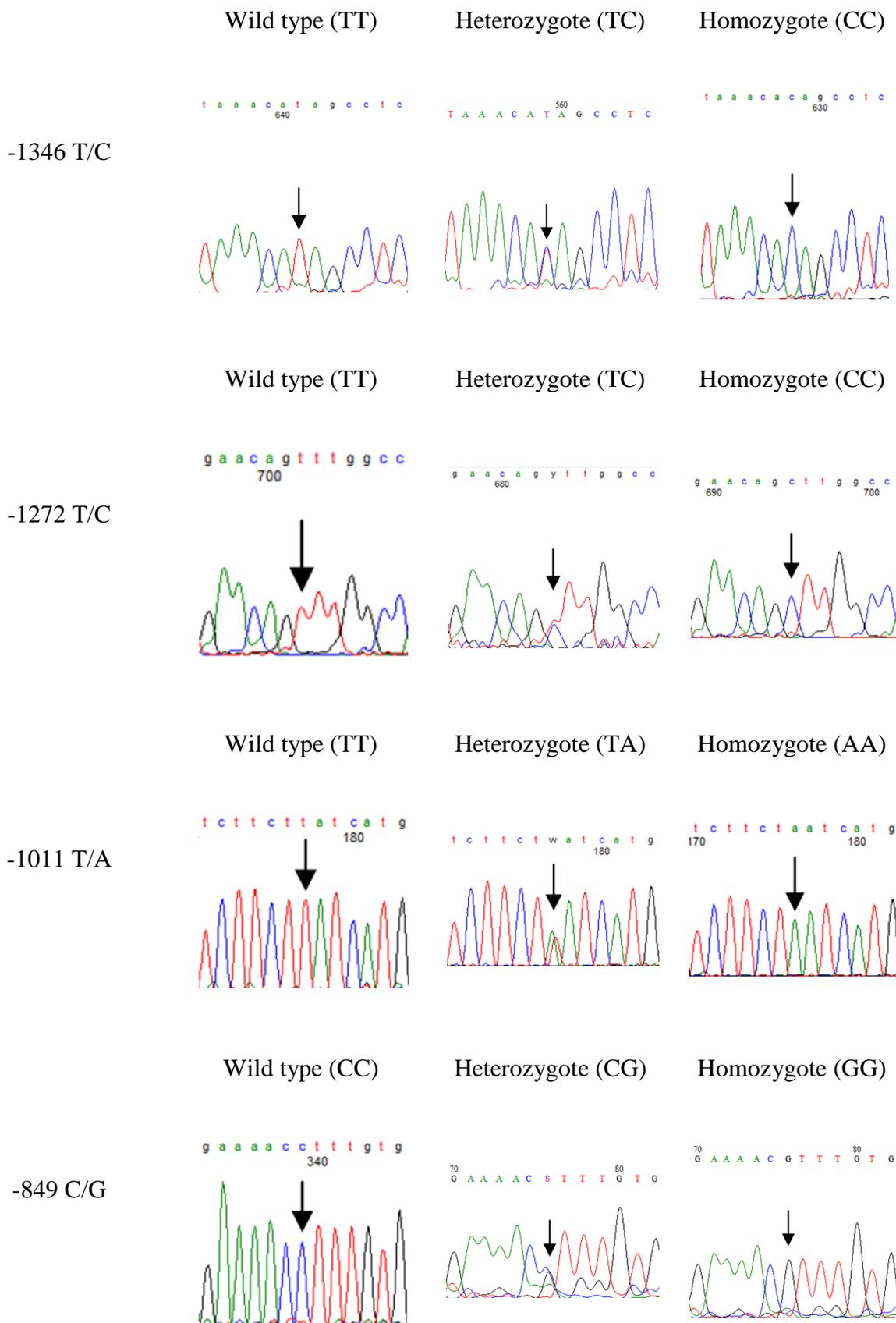
#### *CYBRD1* variants

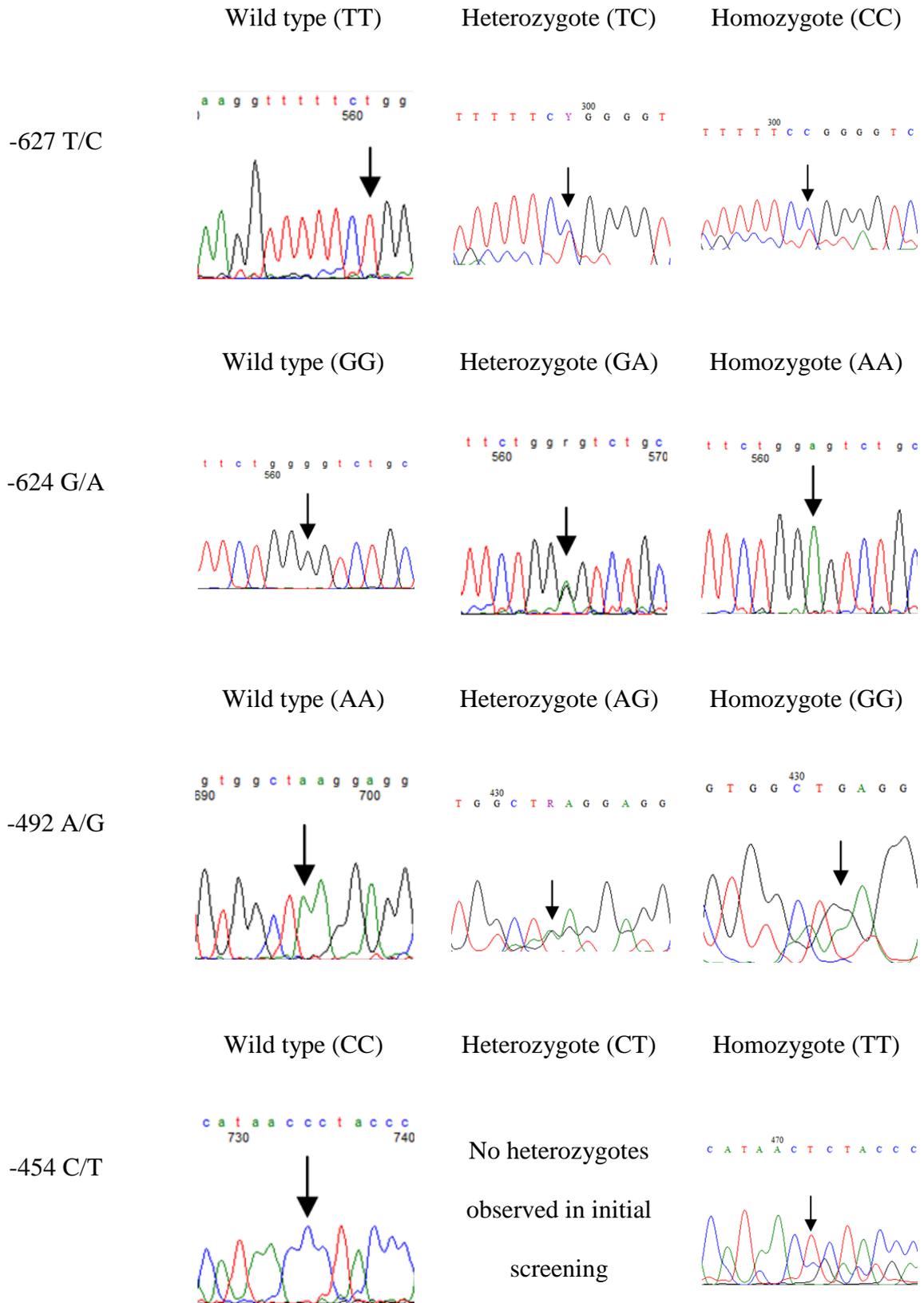
---

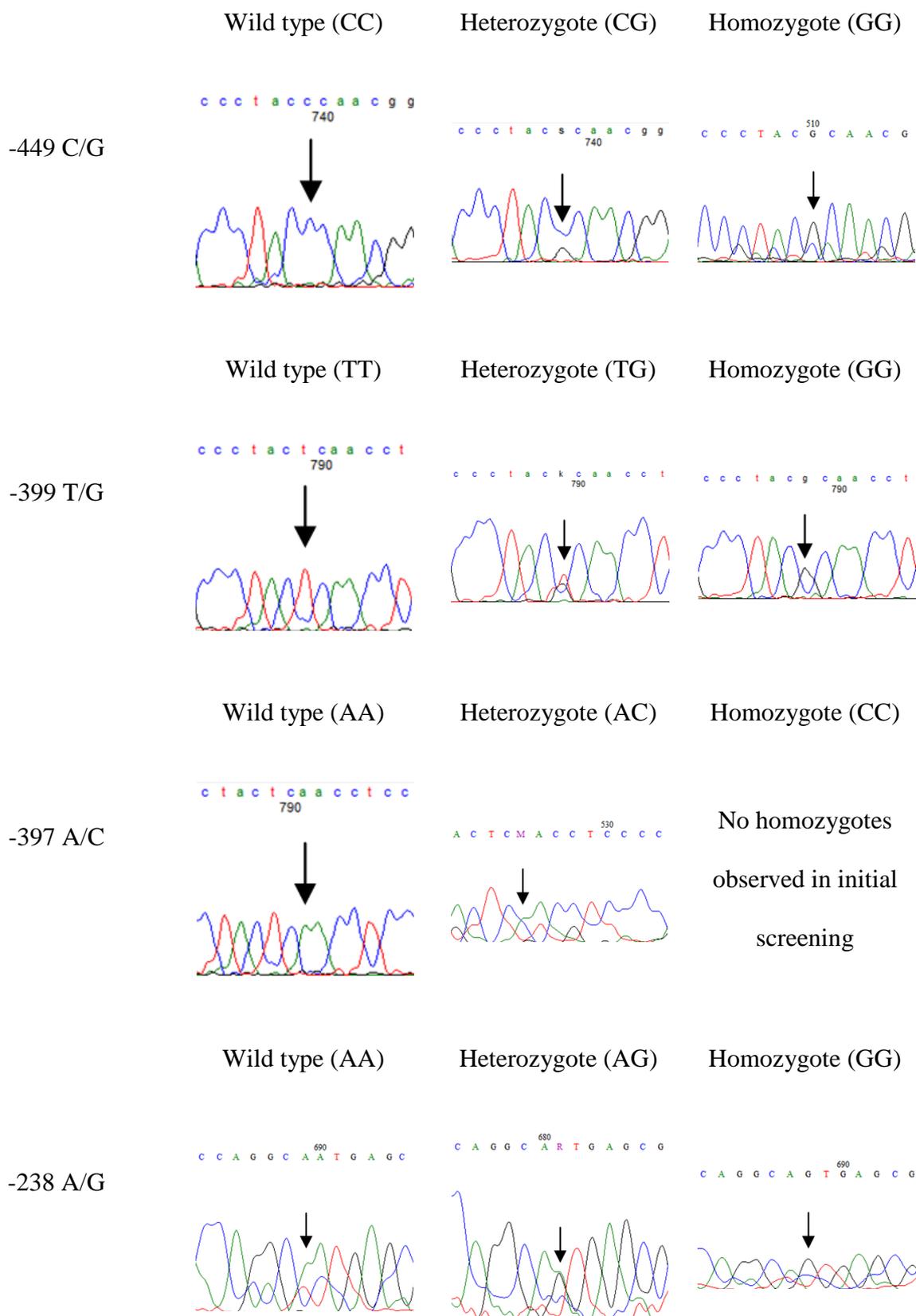


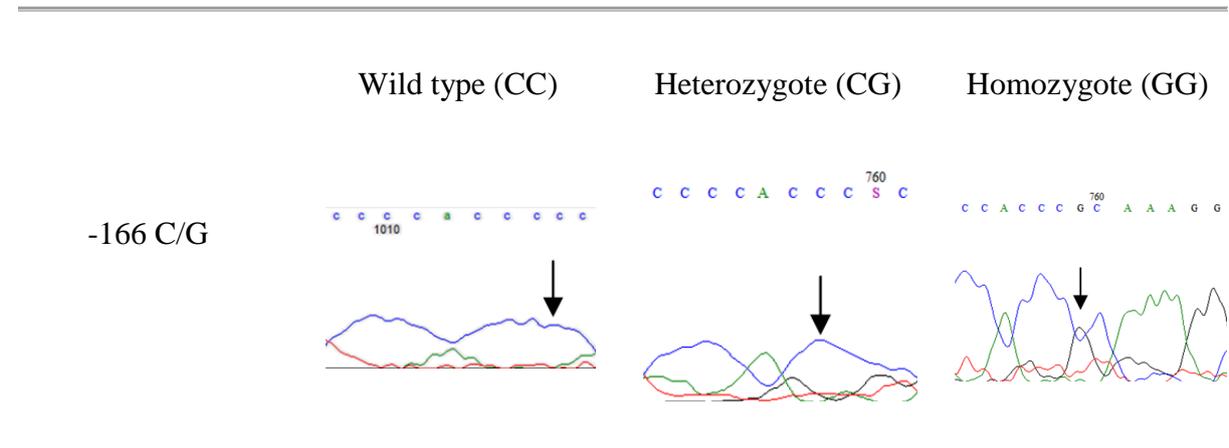




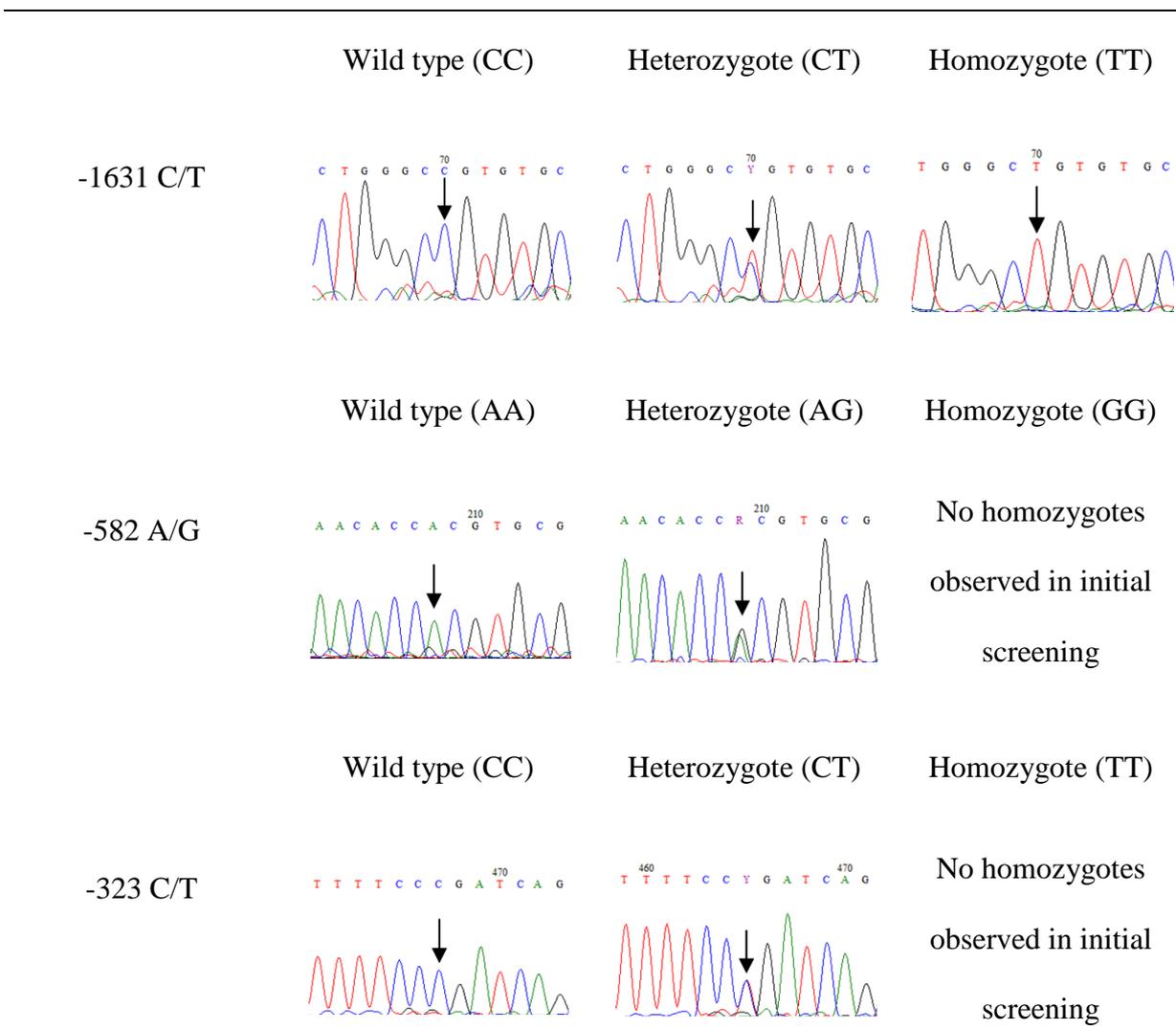


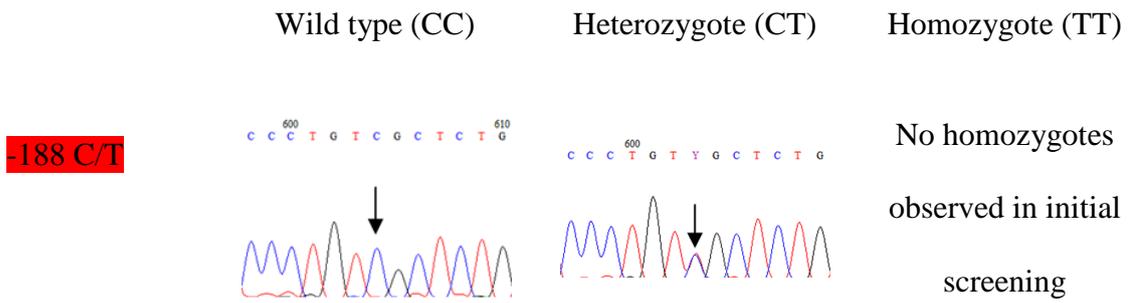




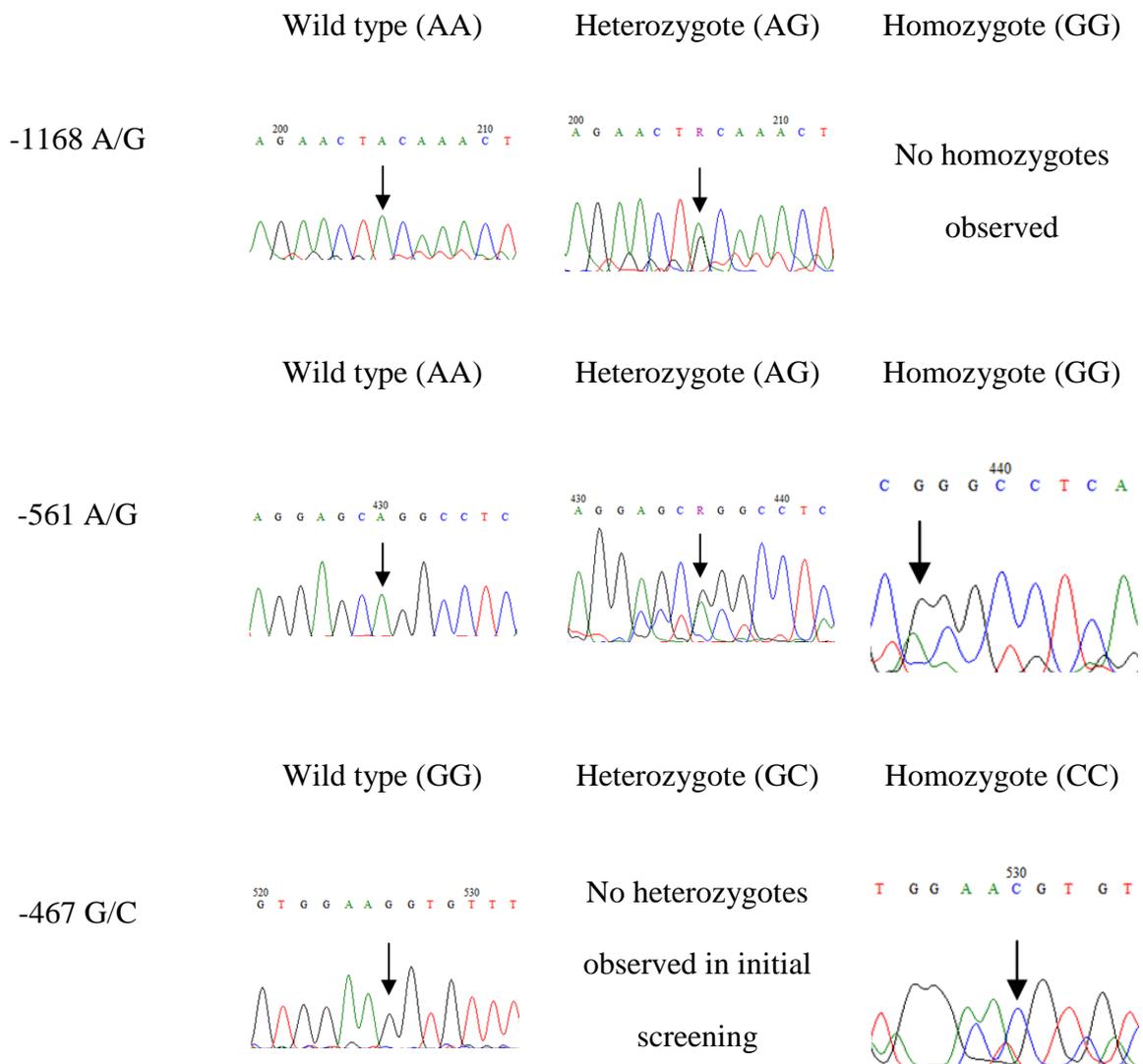


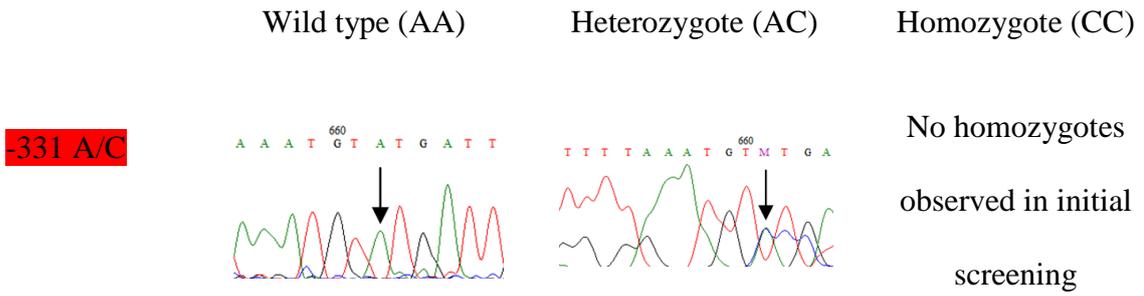
**HAMP variants**



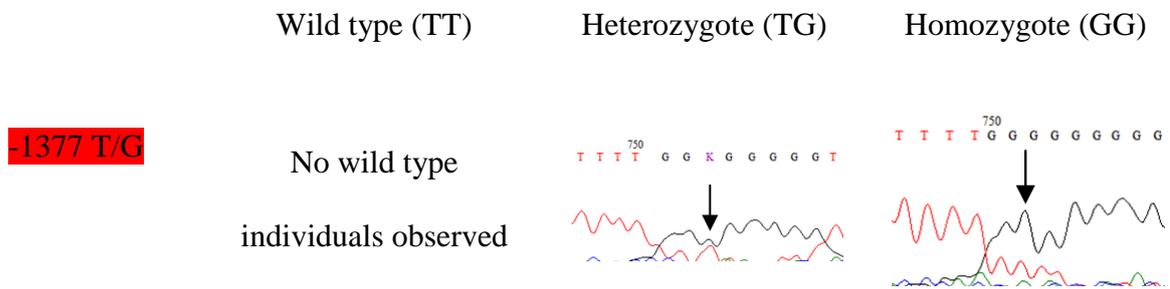


***HFE* variants**

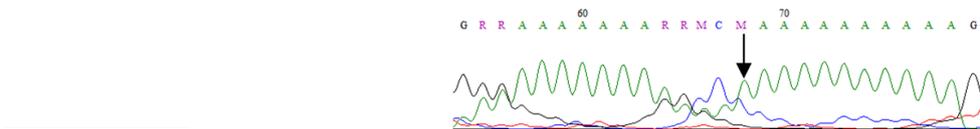




***LTF* variants**

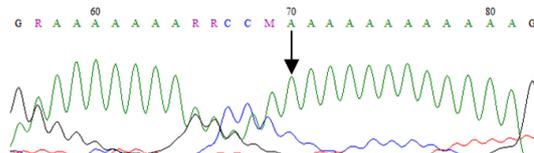


(WT/-AAA) 11 and 14 adenines in sequence



**insAA**

(WT/[-AAA/insAA]) 13 and 14 adenines in sequence

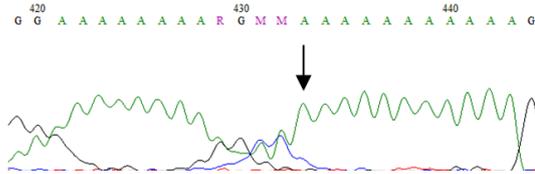


(-AAA/[-AAA/insAA]) 11 and 13 adenines in sequence

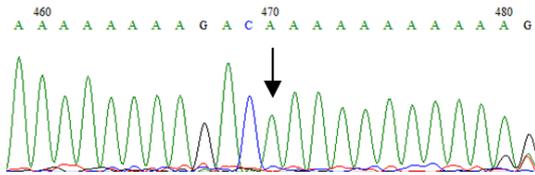
-746 delAAA/

insAA

(continued)



(-AAA/-AAA) 11 adenines in sequence



Wild type (TT)

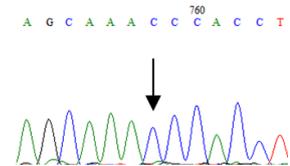
Heterozygote (TC)

Homozygote (CC)

-457 T/C

No wild types  
observed

No heterozygotes  
observed

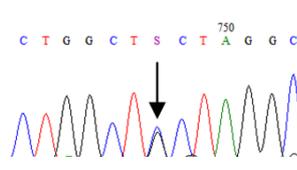
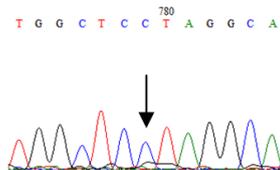


Wild type (CC)

Heterozygote (CG)

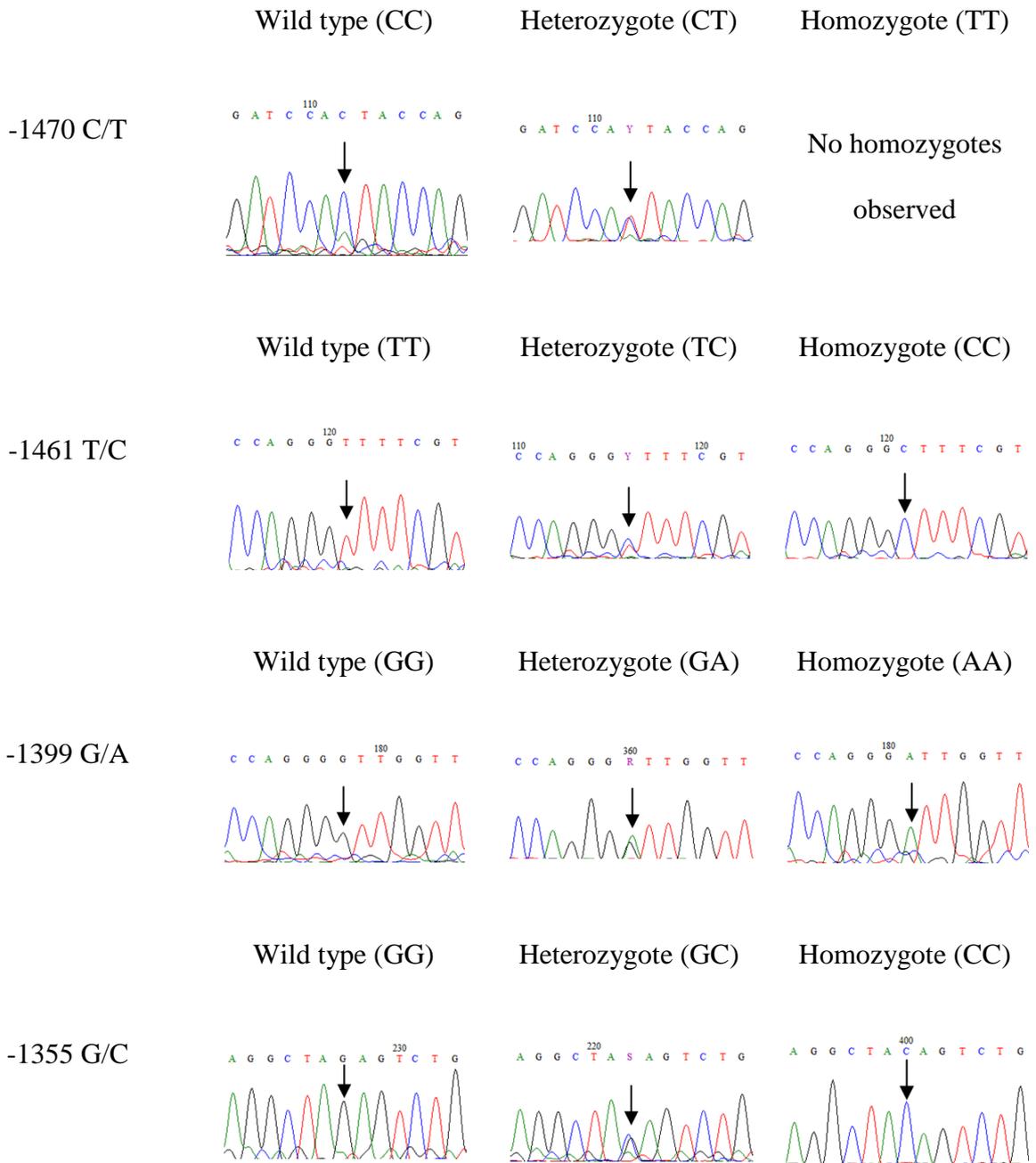
Homozygote (GG)

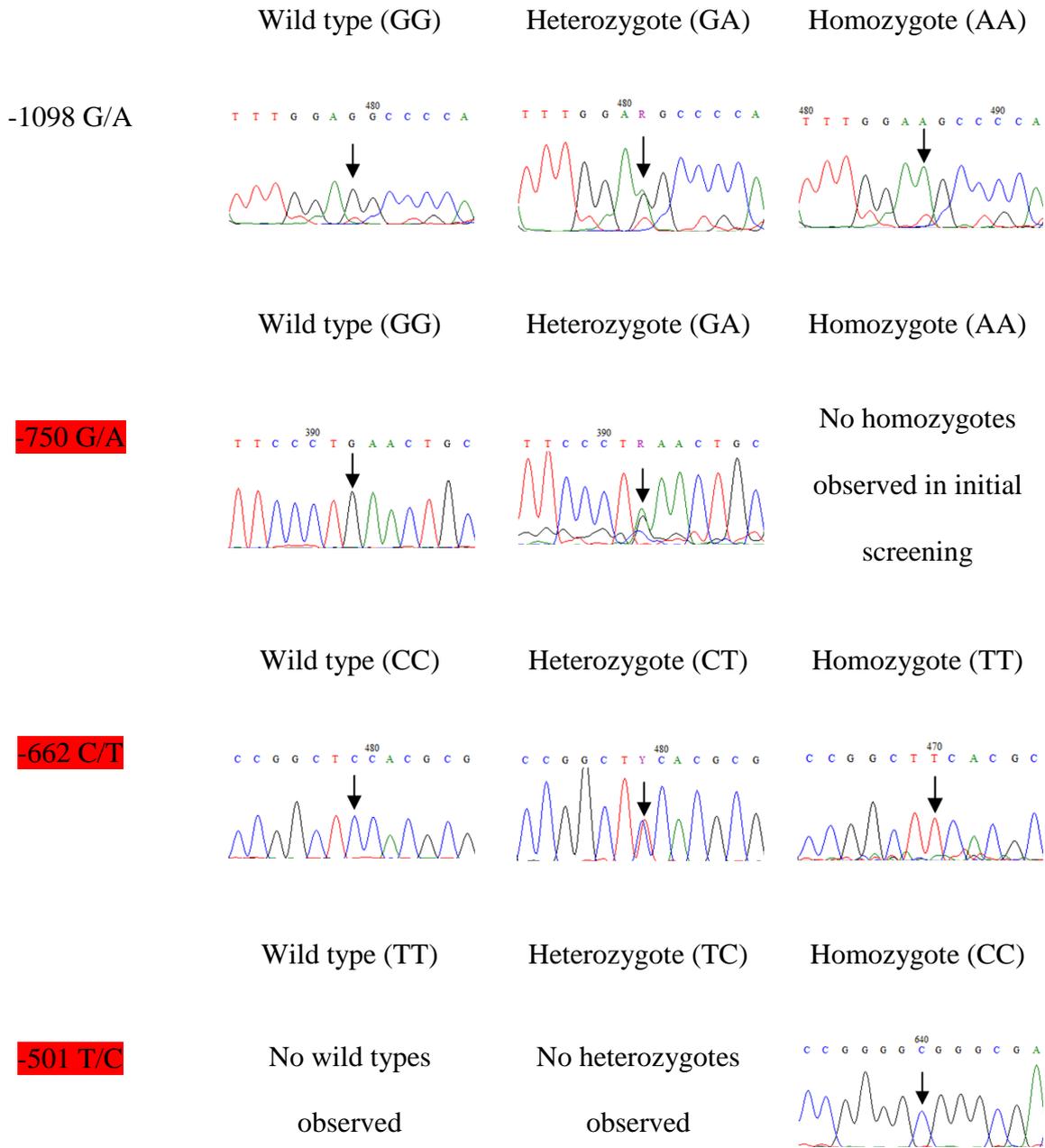
-437 C/G



No homozygotes  
observed

***SLC40A1* variants**



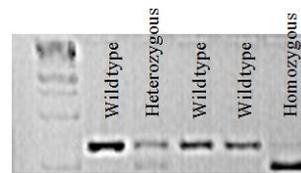
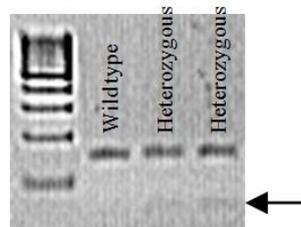
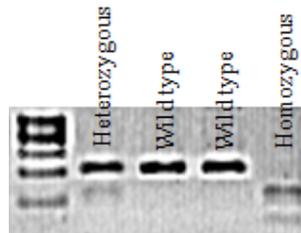


---

**B.1.2 RFLP Gel Images**

PCR product digestion was performed as described in section 2.3.2 and electrophoresed (section 2.2.3). Examples of results obtained are provided below. The horizontal arrow in the gel image of variant -561 A/G indicates the lower band which makes the individual a heterozygote (this band is light thus the arrow is added as an aid).

---

***HAMP*****-582 A/G*****HFE*****-561 A/G****-467 G/C**

---

**B.1.3 HEX-SSCP Gel Images**

The images here illustrate variants detected by HEX-SSCP analysis. Horizontal arrows indicate differences in banding patterns allowing one to distinguish between genotypes.

---

***HAMP***

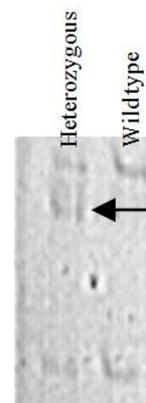
---



-323 C/T



-188 C/T



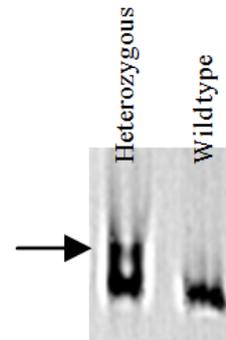
---

---

*HFE*

---

-1168 A/G

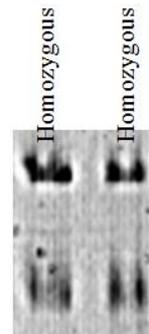


---

*LTF*

---

-457 T/C

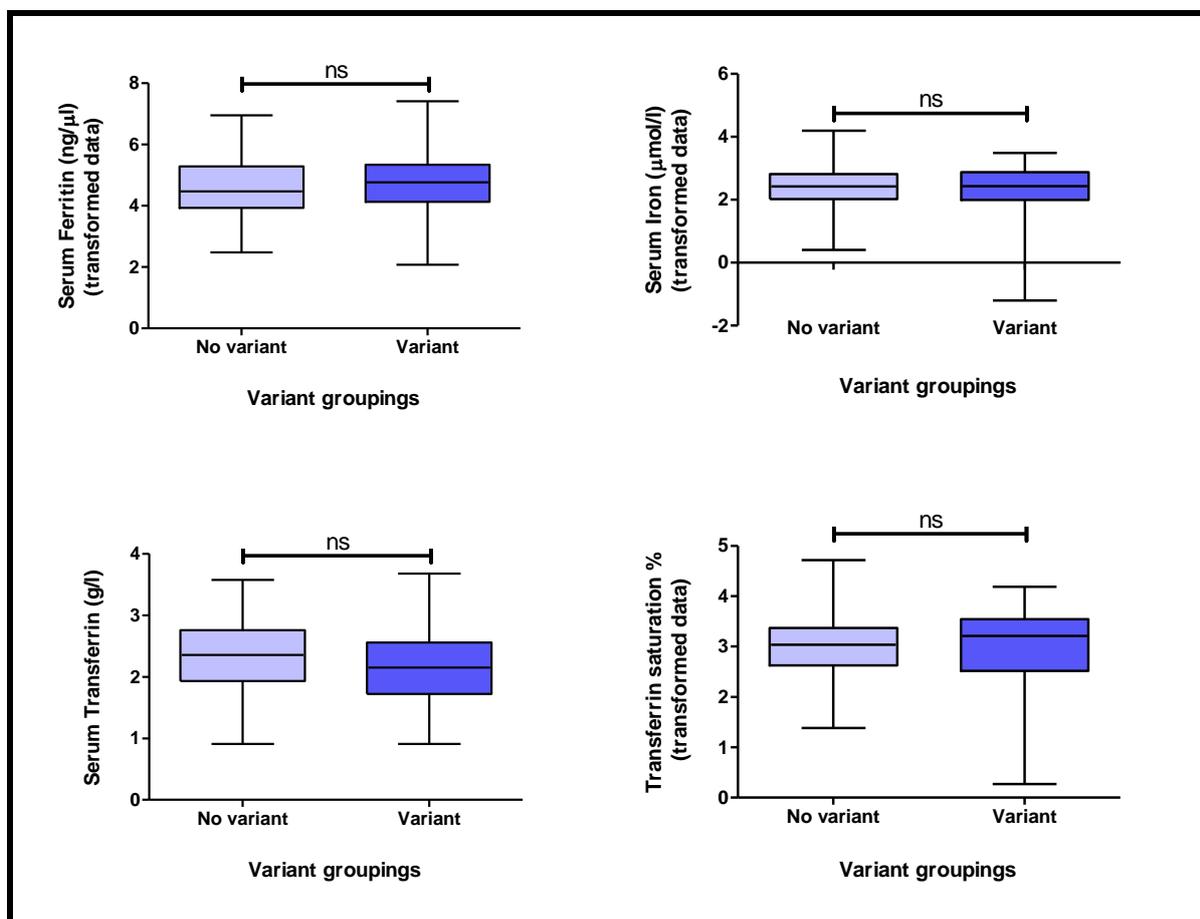


## **Addendum C**

## ADDENDUM C

### C.1 T-TESTS FOR IRON PARAMETER LEVEL COMPARISONS

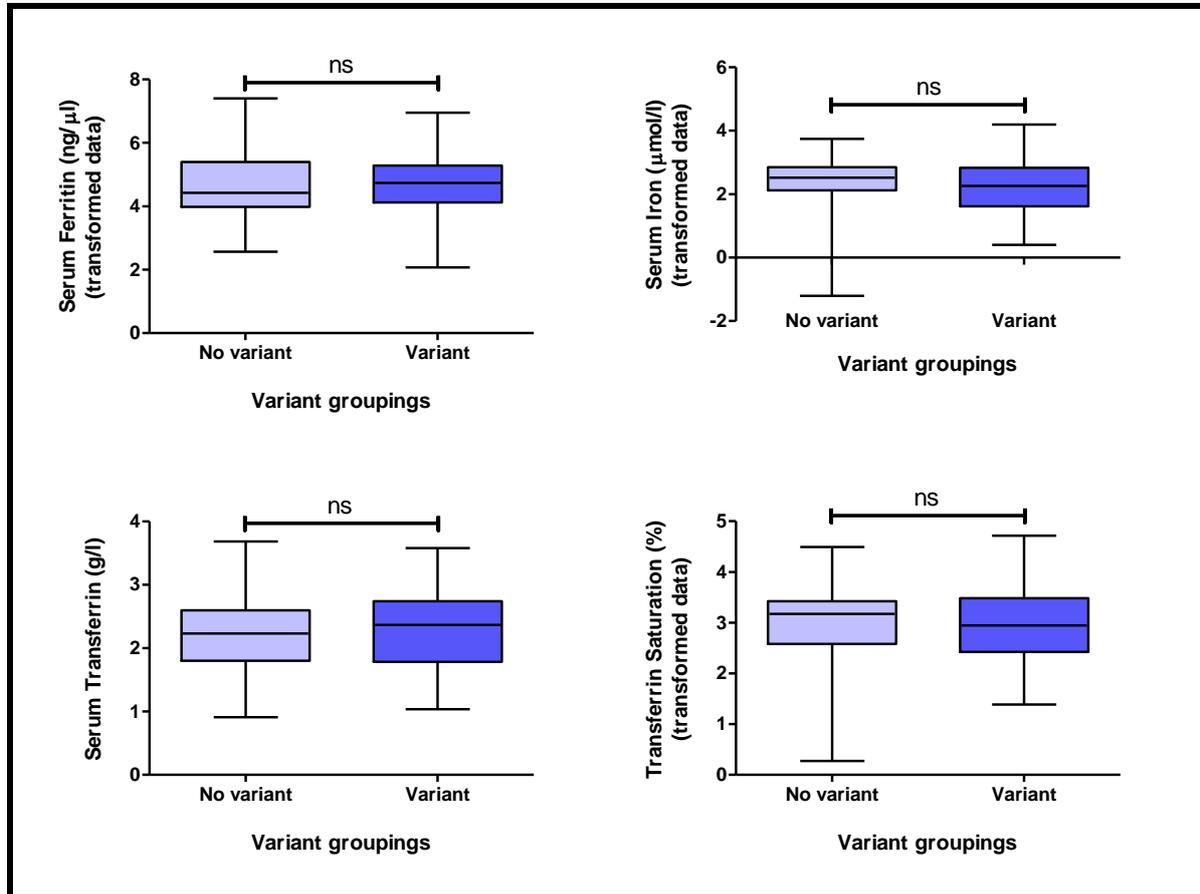
Unpaired *t*-tests were performed to ascertain whether a statistically significant difference could be observed between iron parameter levels in WT individuals, and those with the variant. Figures C.1 to C.7 represent those results which showed differences due to the presence of the variant but did not yield statistically significant results



**Figure C.1 Comparison of iron parameters to determine the effect of the variant -1631 C/T in the promoter region of *HAMP***

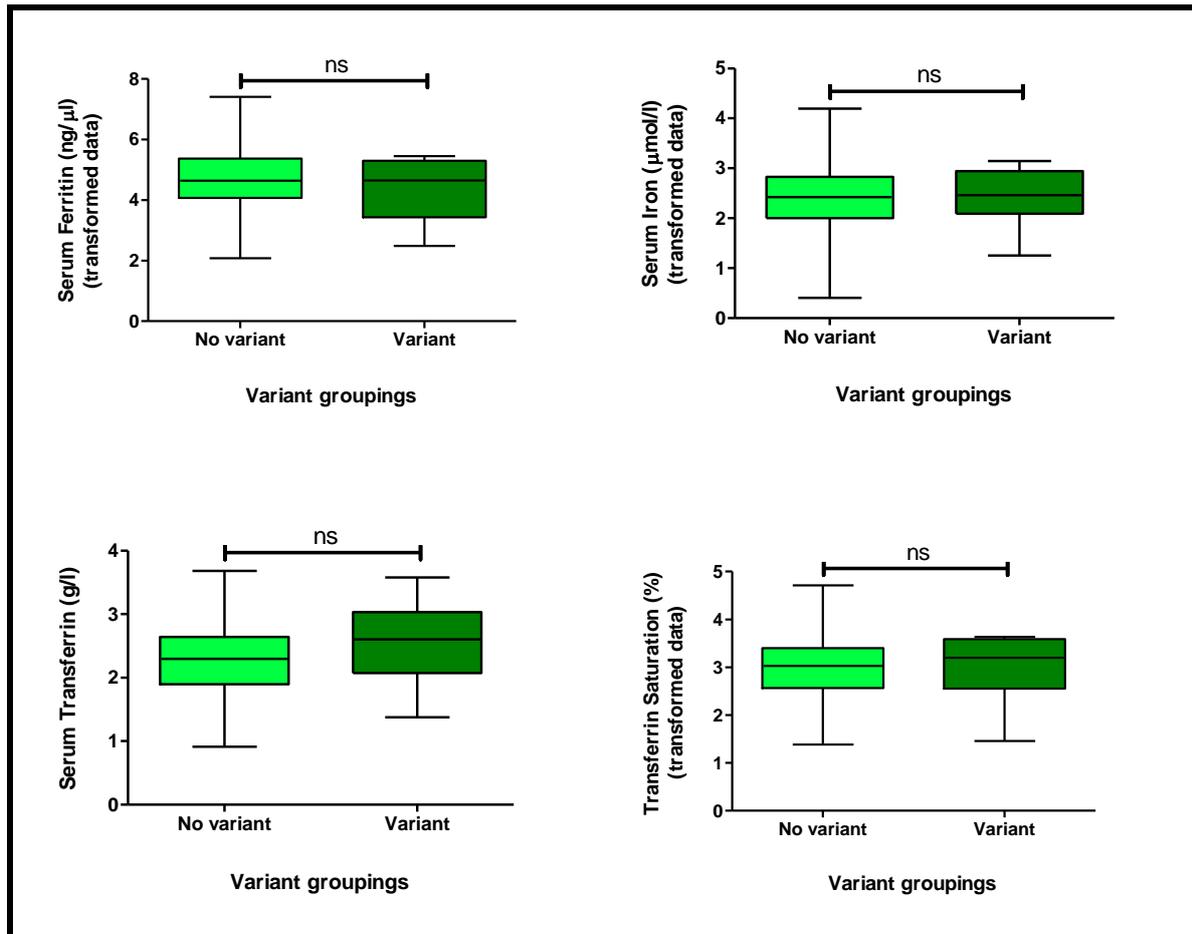
Differences in iron parameters were observed for the -1631 C/T variant located in the promoter region of the *HAMP* gene when comparing the 'no variant' group to the 'variant'

group but none of these differences proved to be statistically significant (ns – as depicted in Figure C.1).



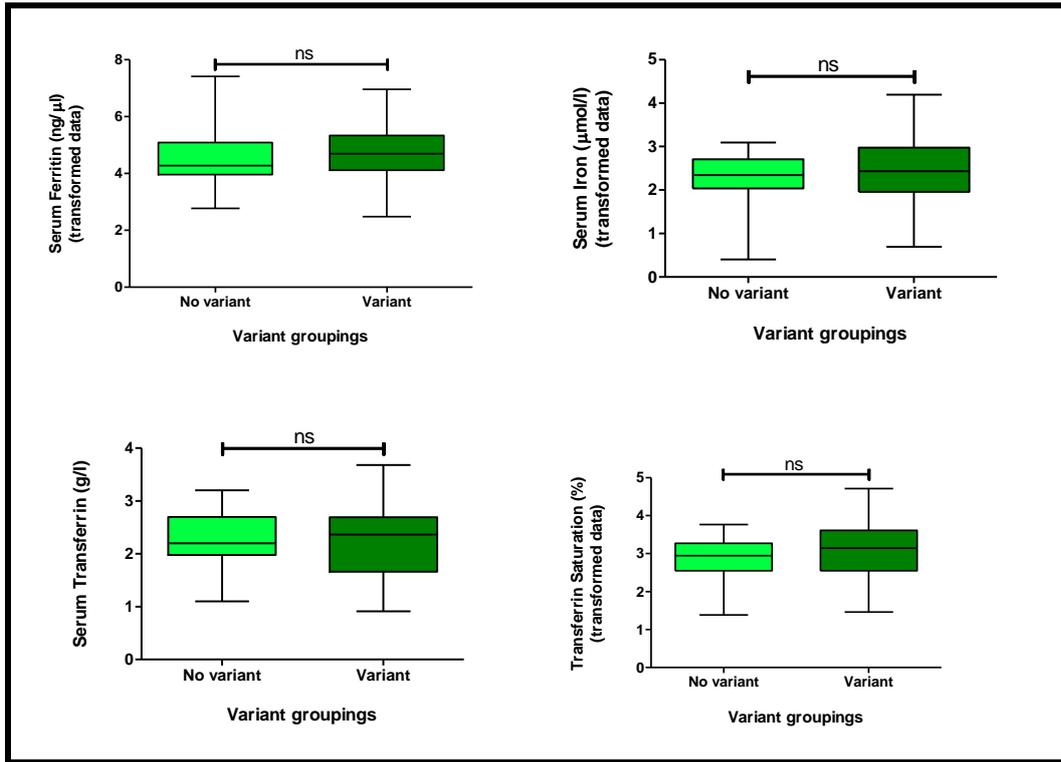
**Figure C.2 Comparison of iron parameters to determine the effect of the variant -582 A/G in the promoter region of *HAMP***

The variant -582 A/G located in the promoter region of *HAMP* showed visible differences in iron parameters between the variant-containing and non-variant-containing groups but these results proved to be non-significant (Figure C.2). The same results were found for the variant -1168 A/G located in the promoter region of the *HFE* gene (Figure C.3).

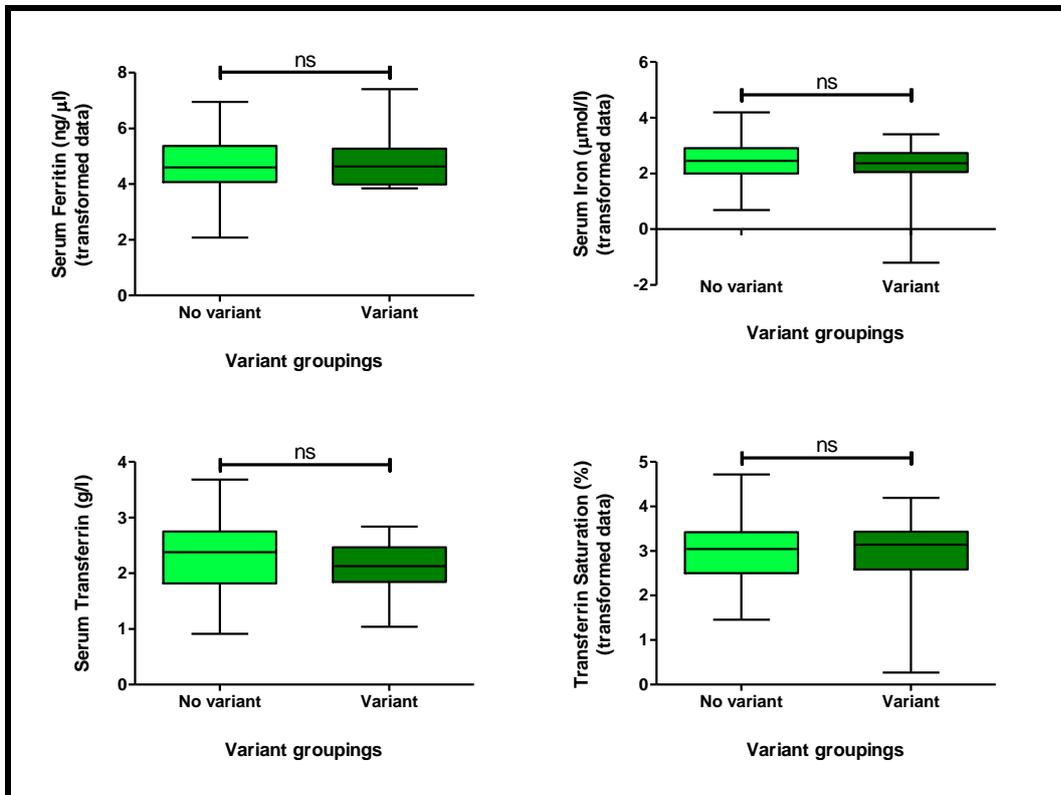


**Figure C.3 Comparison of iron parameters to determine the effect of the variant -1168 A/G in the promoter region of *HFE***

These results were imitated for the two remaining variants investigated in this manner: variants -561 A/G and -467 G/C both located in the promoter region of the *HFE* gene (Figures C.4 and C.5).

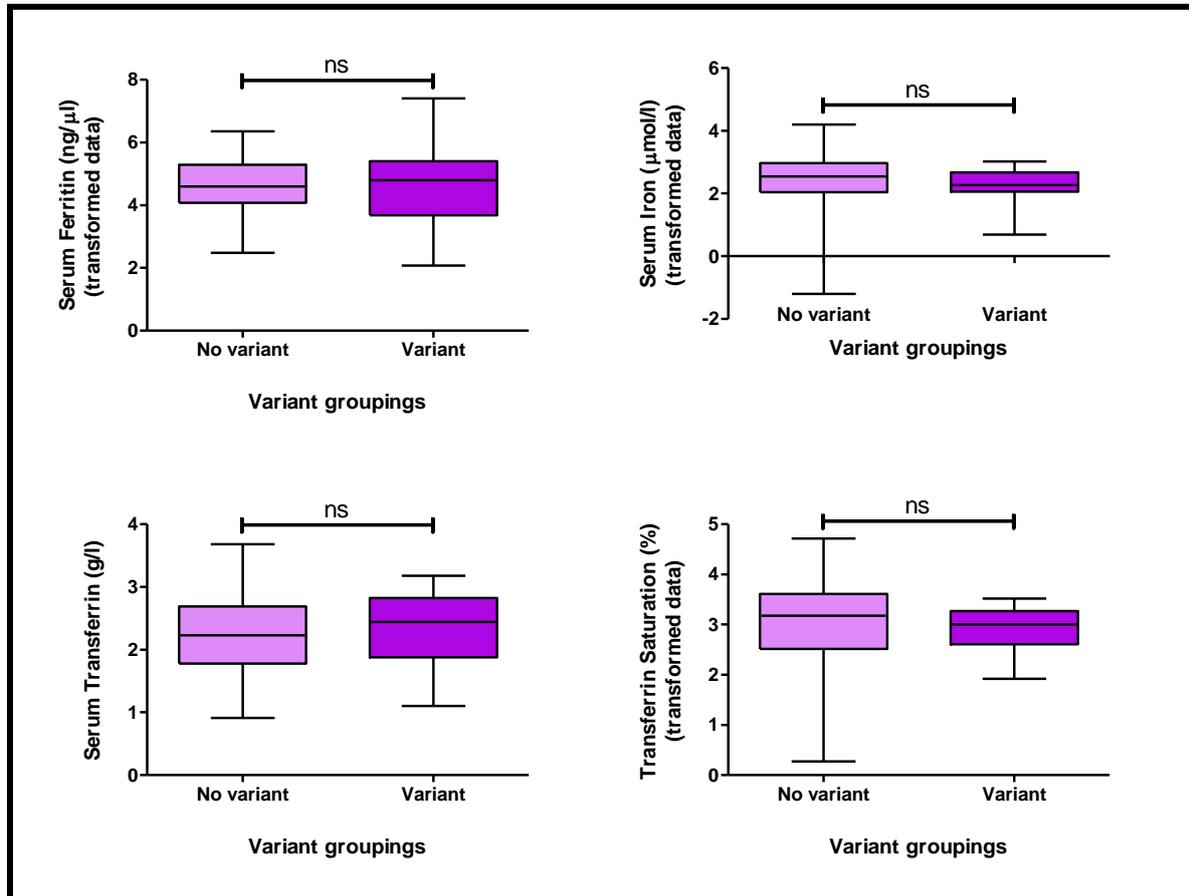


**Figure C.4 Comparison of iron parameters to determine the effect of the variant -561 A/G in the promoter region of *HFE***

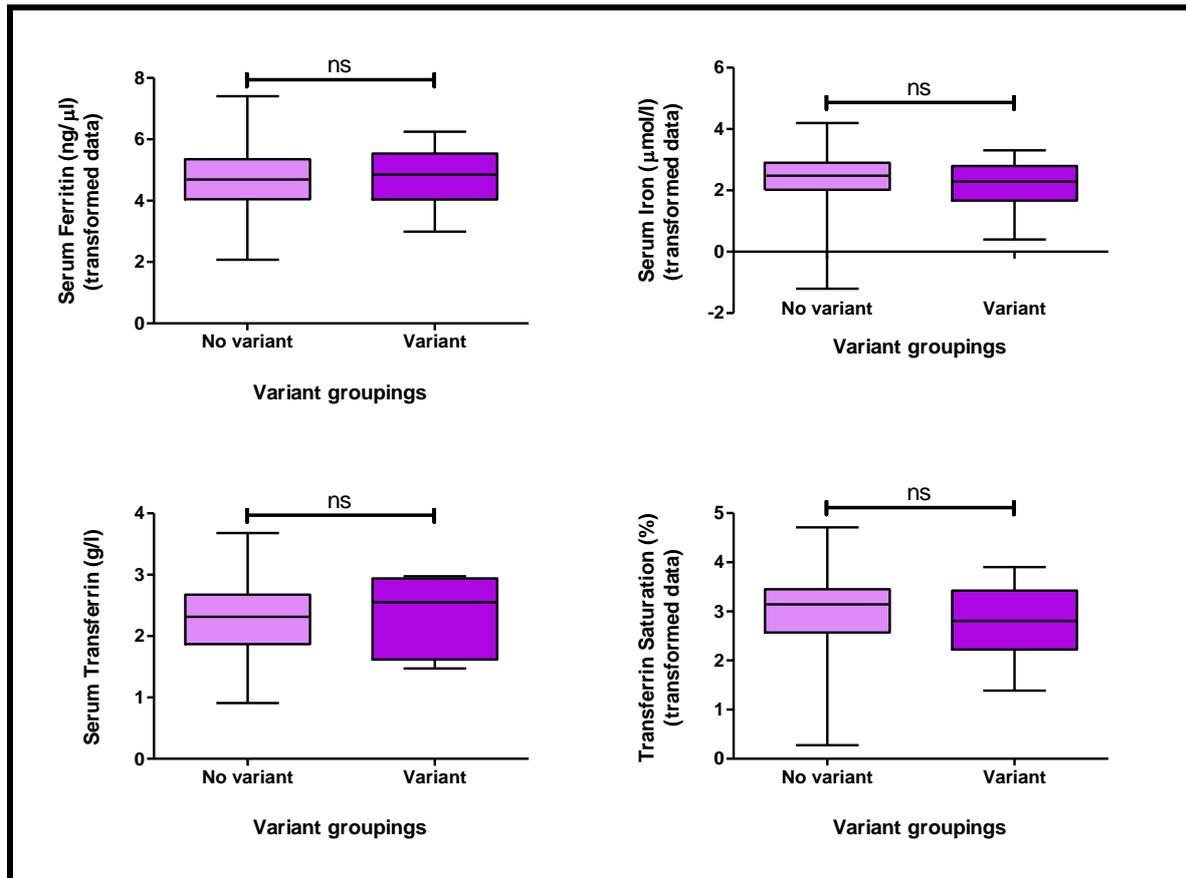


**Figure C.5 Comparison of iron parameters to determine the effect of the variant -467 G/C in the promoter region of *HFE***

No statistically significant differences were seen for any of the four iron parameters when comparing the variant absent group to the variant present group for either the variant -1461 T/C or the variant -1399 G/A (Figures C.6 and C.7 respectively).



**Figure C.6 Comparison of iron parameters to determine the effect of the variant -1461 T/C in the promoter region of *SLC40A1***



**Figure C.7** Comparison of iron parameters to determine the effect of the variant -1399 G/A in the promoter region of *SLC40A1*