Molecular Genetic Analysis of *Ceruloplasmin* in Oesophageal Cancer

by

Natalie Strickland

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science (MSc) in Genetics at Stellenbosch University

Supervisor: Dr MG Zaahl

Co-supervisor: Dr T Matsha

March 2009

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 owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature: .....

Date: .....

SUMMARY

Oesophageal cancer (OC) is characterised by the development of malignant tumours in the epithelial cells lining the oesophagus. It demonstrates marked ethnic variation, with squamous cell carcinoma (SCC) being more prevalent in the Black population and adenocarcinoma (ADC) occurring more often in Caucasians. The aetiology of this complex disease has been attributed to a variety of factors, including an excess of iron (resulting in increased tumourigenesis), oesophageal injury and inflammation.

The present study attempted to determine the mutation spectrum of the regulatory and coding regions of the ceruloplasmin (*CP*) gene, involved in iron metabolism, in the Black South African OC population. The patient cohort was comprised of 96 (48 male and 48 female) unrelated individuals presenting with SCC of the oesophagus. The control group consisted of 88 unrelated, healthy population-matched control individuals. The techniques employed for mutation detection in this study included polymerase chain reaction (PCR) amplification, heteroduplex single-strand conformation polymorphism (HEX-SSCP) analysis, restriction fragment length polymorphism (RFLP) analysis followed by bidirectional semi-automated DNA sequencing analysis to verify the variants identified.

Mutation detection of *CP* resulted in the identification of fourteen previously described (5'UTR-567C→G, 5'UTR-563T→C, 5'UTR-439C→T, 5'UTR-364delT, 5'UTR-354T→C, 5'UTR-350C→T, 5'UTR-282A→G, V223, Y425, R367C, D544E, IVS4-14C→T, IVS7+9T→C and IVS15-12T→C) and four novel (5'UTR-308G→A, T83, V246A and G633) variants. Statistical analysis revealed that two of the novel variants were significantly associated with OC in this study; the promoter variant 5'UTR-308G→A (P=0.012) and the exonic variant G633 (P=0.0003). It is possible that these variants may contribute to OC susceptibility in the Black South African population.

OC symptoms generally present late in the development of the disease, and as a result treatment after diagnosis is highly ineffective. Early detection of symptoms and subsequent treatment is therefore the most effective manner of disease intervention. In high incidence areas, such as the Transkei region of South Africa, the implementation of a screening programme would be the ideal way to achieve this goal. The information that can be gathered from the identification of potential modifier genes for OC can lead to improvements in early detection, which in turn may lead to advancements in the treatment and counselling to individuals with OC. To our knowledge, this is the first study concerning *CP* and its effects on iron dysregulation in the Black South African population with oesophageal cancer.

OPSOMMING

Oesofageale kanker word gekenmerk deur die ontwikkeling van kwaardaardige gewasse in die epiteelweefsel van die oesofageale voering. Hierdie siekte demonstreer opvallende etniese variasie, met plaveisel selkarsinoom meer algemeen in die Swart populasie en adenokarsinoom meer algemeen in die Kaukasiese populasie. Die ontwikkeling van hierdie komplekse siekte word aan ‘n aantal faktore toegeskryf, insluitend ‘n oormaat yster (wat lei tot ‘n vermeerdering van gewasse) en oesofageale besering en -ontsteking.

Die doel van die hierdie studie was om die mutasie spektrum van die regulatoriese- en koderingsarea van die ceruloplasmin (*CP*) geen, betrokke in yster metabolisme, in die Swart Suid Afrikaanse oesofageale kanker populasie te bepaal. Die pasiënt groep het bestaan uit 96 (48 manlik en 48 vroulik) onverwante individue met plaveisel selkarsinoom van die oesofagus. Die kontrole groep het uit 88 nie-geaffekteerde onverwante, populasie spesifieke individue bestaan. Die tegnieke aangewend vir mutasie deteksie in hierdie studie sluit in polimerase kettingsreaksie amplifikasie, heterodupleks enkelstring konformasie polimorfisme analise en restriksie fragment lengte polimorfisme analise, gevolg deur tweerigting semi-geoutomatiseerde DNS volgorde-bepalingsanalise om die geïdentifiseerde variante te bevestig.

Mutasie deteksie van *CP* het tot die identifikasie van veertien reeds beskryfde (5’UTR-567C→G, 5’UTR-563T→C, 5’UTR-439C→T, 5’UTR-364delT, 5’UTR-354T→C, 5’UTR-350C→T, 5’UTR-282A→G, V223, Y425, R367C, D544E, IVS4-14C→T, IVS7+9T→C en IVS15-12T→C) en vier nuwe (5’UTR-308G→A, T83, V246A en G633) variante gelei. Statistiese analise het getoon dat twee van die nuwe variante betekenisvol geassosieerd was met oesofageale kanker in hierdie studie; die promotor variant 5’UTR-308G→A (*P*=0.012) en

die eksoniese variant G633 (*P*=0.0003). Dit is moontlik dat hierdie variante mag bydra tot oesofageale kanker vatbaarheid in die Swart Suid Afrikaanse populasie.

Oesofageale kanker simptome vertoon gewoonlik op ‘n latere stadium in die ontwikkelingsproses van die siekte, en as ‘n gevolg is behandeling na diagnose hoogs oneffektief. Vroegtydige identifikasie van die simptome en daaropvolgende behandeling is die mees effektiewe manier vir ingryping. In hoë voorkoms streke, soos die Transkei gebied van Suid Afrika, sal die implementasie van ‘n siftingsprogram die ideale manier wees om hierdie doel te bereik. Die inligting wat dan versamel word, insluitend identifisering van modifiserende gene vir oesofageale kanker, kan lei tot ‘n verbetering in vroegtydige deteksie van die siekte. In effek kan dit dan lei tot beter behandeling en berading vir individue met oesofageale kanker. So ver ons kennis strek, is hierdie die eerste studie wat *CP* en sy effek op yster disregulasie in die Swart Suid-Afrikaanse populasie met oesofageale kanker behels.

Dedicated to my family

‘A person who never made a mistake never tried anything new.’ – Albert Einstein

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to the following institutions and individuals without whom this study would not have been possible:

The National Research Foundation (Thuthuka) and the University of Stellenbosch for providing financial support.

The University of Stellenbosch and the Department of Genetics for providing the infrastructure and facilities utilised in the completion of this study.

My supervisor, Dr MG Zaahl, for affording me the opportunity to be part of her research group, interesting discussions, support, enthusiasm for the project and for critical reading of this manuscript.

My co-supervisor, Dr T Matsha, for providing DNA samples, performing the biochemical analysis and for critical reading of this manuscript.

Veronique Human, for DNA extractions, and for providing a positive working environment, emotional support and for never being too busy to help or answer questions.

My housemates, Marika Bosman and Daleen Badenhorst, for putting up with me through the long hours spent writing this thesis and for providing many nights of laughter and fun.

My fellow researchers in Lab 242, Nathan McGregor, Jessica Vervalle, Marika Bosman, Alisa Postma, Jomien de Jager for providing constant amusement, a wonderful working environment and invaluable advice.

My father and mother for providing financial assistance, emotional support and encouragement throughout the writing of this thesis and my years at the University of Stellenbosch. Thank you for making my dreams a reality.

Matt, for your support and encouragement of my passion.

All of my friends, for putting up with me during the writing of this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER ONE: LITERATURE REVIEW

CHAPTER THREE: RESULTS AND DISCUSSION

LIST OF TABLES

CHAPTER ONE: LITERATURE REVIEW

CHAPTER TWO: DETAILED EXPERIMENTAL PROCEDURES

CHAPTER THREE: RESULTS AND DISCUSSION

LIST OF ABBREVIATIONS AND SYMBOLS

3’	3-prime
5’	5-prime
α	alpha
β	beta
χ^2	chi-square
©	copyright
˚C	degrees celcius
=	equal to
γ	gamma
>	greater than
<	less than
µ	micro (10^{-6})
µg	microgram
µg/l	microgram per litre
µl	microlitre
µM	micro molar
%	percentage
%C	percentage crosslinking
+	plus
±	plus-minus
®	registered trademark
×	times
× g	times gravity
™	trademark

A	adenosine
A (amino acid)	alanine
AA	acrylamide; C_3H_5NO
ADC	adenocarcinoma of the oesophagus
AP-1	activator protein 1
APS	ammonium persulphate; $(NH_4)_2S_2O_8$

ASIR	age standardised incidence rate
ATG	translation initiation site
ATP	adenosine triphosphate
ATP7B	ATPase, copper (Cu^{2+}) transporting, beta polypeptide gene
BAA	bis-acrylamide; N,N'-methylene-bis-acrylamide: $C_7H_{10}O_2N_2$
bp	base pair
BMI	body mass index
BSA	bovine serum albumin
C (amino acid)	cysteine
C	cytosine
$C_{13}H_{28}N_2Na_4O_{13}S$	xylene cyanol
$C_{19}H_{10}Br_4O_5S$	bromophenol blue
C/EBPα	CCAAT/enhancer binding protein alpha
CEPH	*Centre d'Etude du Polymorphisme Humain*
cm	centimetre
c-myc	myc proto-oncogene
CNS	central nervous system
CP	ceruloplasmin gene
CP	ceruloplasmin protein
Cp	mouse ceruloplasmin gene
CYBRD1	cytochrome b reductase 1 gene
CYBRD1	cytochrome b reductase 1 protein
D (amino acid)	aspartic acid
D'	coefficient of association
dATP	2'-deoxy-adenosine-5'-triphosphate
dCTP	2'-deoxy-cytidine-5'-triphosphate
DCT1	divalent cation transporter 1 protein
DCYTB	duodenal cytochrome b protein
ddH_2O	double distilled water
del	deletion

dGTP	2'-deoxy-guanosine-5'-triphosphate
dHPLC	denaturing high performance liquid chromatography
DMT1	divalent metal transporter 1 gene
DMT1	divalent metal transporter 1 protein
DNA	deoxyribonucleic acid
dNTP	2'-deoxy-nucleotide-5'-triphosphate
dTTP	2'-deoxy-thymidine-5'-triphosphate
E (amino acid)	glutamic acid
EDTA	ethylenediaminetetraacetic acid; $C_{10}H_{16}N_2O_8$
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor gene
ER	endoplasmic reticulum
ESE	exonic splice element
et al.	and others
EtBr	ethidium bromide; $C_{21}H_{20}BrN_3$
EtOH	ethanol; CH_3CH_2OH
F	forward primer
Fe	iron
Fe^{2+}	ferrous iron
Fe^{3+}	ferric iron
FNP1	ferroportin 1 protein
FOXC1	forkhead box protein C1
FOXD1	forkhead box protein D1
FOXL1	forkhead box protein L1
FOXM1a	forkhead box protein M1a
FOXM1b	forkhead box protein M1b
FTH	H-ferritin protein
FTL	L-ferritin protein

g	gram
G (amino acid)	glycine
G	guanosine
GATA1	GATA-binding protein 1
GATA2	GATA-binding protein 2
GATA3	GATA-binding protein 3
gDNA	genomic deoxyribonucleic acid
GI	gastrointestinal
GORD	gastro-oesophageal reflux disease
GR	glucocorticoid receptor
H^+	hydrogen ion
H_2NCHO	formamide
H_2O_2	hydrogen peroxide
H_3BO_3	boric acid
HAMP	hepcidin gene
HAMP	hepcidin protein
HCl	hydrochloric acid
HCP1	haem carrier protein 1 gene
HCP1	haem carrier protein 1
HEPH	hephaestin gene
HEPH	hephaestin protein
HEX-SSCP	heteroduplex single-strand conformation polymorphism analysis
HFE	high iron protein gene
HFE	high iron protein
Hfe	mouse high iron protein gene
HiNF-D	histone nuclear factor D
HiNF-M	histone nuclear factor M
HiNF-P	histone nuclear factor P
HJV	haemojuvelin gene
HJV	haemojuvelin protein
HMOX1	haem oxygenase 1 gene
HMOX1	haem oxygenase 1 protein
HNF	hepatocyte nuclear factor

HNF-1	hepatocyte nuclear factor 1
HNF-3α	hepatocyte nuclear factor 3 alpha
HNF-3β	hepatocyte nuclear factor 3 beta
HPV	human papilloma virus
HWE	Hardy-Weinberg equilibrium
IDT	Integrated DNA Technologies
in silico	refers to research conducted using computers only
in situ	Latin phrase meaning "in the place"
in vivo	Latin phrase meaning "in the body" or within a living organism
IRE	iron responsive element
IREG1	iron-regulated transporter 1 protein
IRP	iron regulatory protein
IRP1	iron regulatory protein 1
IRP2	iron regulatory protein 2
IVS	intervening sequence
kb	kilo base
kDa	kilo daltons
l	litre
LD	linkage disequilibrium
LOD	logarithm of odds
LOH	loss of heterozygosity
Ltd	limited
m	milli (10^{-3})
M	moles per litre/ molar
MAX	myc-associated factor X
mg	milligram
$MgCl_2$	magnesium chloride
mg/kg	milligram per kilogram
mg/ml	milligram per millilitre

min	minutes
ml	millilitre
mm	millimetre
mM	millimoles per litre/ millimolar
mRNA	messenger RNA
MTP1	metal transporter 1 protein
n	nano (10^{-9})
n	number of individuals
NaCl	sodium chloride
NaOH	sodium hydroxide
NF-Y	nuclear factor Y
ng	nanogram
ng/µl	nanogram per microlitre
$(NH_2)_2CO$	urea
NKX3-1	NK 3 homeobox 1
NRAMP2	natural resistance-associated macrophage protein 2
nt	nucleotide
O_2^-	superoxide free radical
OC	oesophageal cancer
OD	optical density
OH^-	hydroxyl free radical
p	short arm of chromosome
P	probability
p53	tumour suppressor 53 gene
p53	tumour suppressor 53 protein
PAA	polyacrylamide
PCR	polymerase chain reaction
PD	Parkinson's disease
pH	potential of hydrogen
pmol	pico mole

q	long arm of chromosome
R (amino acid)	arginine
R	reverse primer
r^2	correlation coefficient between two loci
Rb	retinoblastoma gene
RFLP	restriction fragment length polymorphism
ROS	reactive oxygen species
rSNP	regulatory single nucleotide polymorphism
RXRα	retinoid X receptor alpha
RXRγ	retinoid X receptor gamma
SCC	squamous cell carcinoma
SD	standard deviation
SF	serum ferritin
SfcI	*Escherichia coli* strain that carries the cloned *SfcI* gene from *Streptococcus faecium*
SLC40A1	solute carrier family 40 member 1 gene
SLC40A1	solute carrier family 40 member 1 protein
SNP(s)	single nucleotide polymorphism(s)
SOX9	sry-related high-mobility group box 9
SP1	specificity protein 1
SPI1	spleen focus forming virus proviral integration oncogene 1
SPIB	SPI-B transcription factor
SR	serine/arginine-rich protein
SRF	serum response factor
SRY	sex-determining region Y
SSCP	single-strand conformation polymorphism

T (amino acid)	threonine
T	thymidine
T3Rα	thyroid receptor hormone alpha
T_A	annealing temperature
$T1_A$	annealing temperature 1
$T2_A$	annealing temperature 2
Taq	*Thermus aquaticus*
TBE	tris-borate-EDTA buffer
TBI	transferrin-bound iron
TEMED	N,N,N',N'-tetramethylethylenediamine
TF	transferrin gene
TF	transferrin protein
TF(s)	transcription factor(s)
TFBS(s)	transcription factor binding site(s)
TFIID	transcription factor IID
TFPGA	tools for population genetics
TFR(s)	transferrin receptor(s)
TFR1	transferrin receptor 1 protein
TFR2	transferrin receptor 2 protein
Tm	melting temperature
TMF	TATA element modulatory factor
TNM	tumour-node-metastasis
Tris-HCl	tris hydrochloride [2-amino-2-(hydroxymethyl)-1,3propanediol-hydrochloride]
TS	transferrin saturation
U	enzyme activity unit
UK	United Kingdom
USA	United States of America
US	University of Stellenbosch
USF1	upstream regulatory factor 1 protein
USF2	upstream regulatory factor 2 gene
USF2	upstream regulatory factor 2 protein

UTR	untranslated region
UV	ultraviolet
V(amino acid)	valine
v	version
V	volt
vice versa	Latin phrase meaning “the other way round”
v/v	volume per volume
WA	Wales
w/v	weight per volume
WT	wild-type
Y (amino acid)	tyrosine
YY1	ying yang 1

CHAPTER 1

LITERATURE REVIEW

1.1 OESOPHAGEAL CANCER

1.1.1 Background

Oesophageal cancer (OC) is the 15th most common cancer in developed countries of the world and the 4th most common in developing countries (Crespi *et al*. 1994). This disease typically arises from a malignancy of the epithelium lining the oesophagus. The oesophagus consists of three sections – the upper, middle and lower. The wall of the oesophagus is made up of layers of muscle and is lined by the epithelium (Yang and Davis 1988).

OC characteristically originates in the lower layers of the cells lining the oesophagus, and grows steadily toward the outer layers of cells. The majority of tumours that develop in this region are malignant. These tumours are usually classified into two major subtypes, based on differences in tumour histology: adenocarcinoma (ADC) and squamous cell carcinoma (SCC). Occasionally small-cell carcinomas, that share many properties with small-cell lung cancer, may account for tumour malignancy in OC (Chen and Yang 2001).

ADC is often associated with a history of gastro-intestinal reflux and Barrett's oesophagus, and usually develops in the glandular tissue in the lower third of the oesophagus. It is characterised by replacement of the squamous epithelium in this region by columnar epithelium (Daly *et al*. 2000). In contrast, SCC develops in the squamous cells that line the upper portion of the oesophagus, and is relatively similar to head and neck cancers in its appearance (Yang and Davis 1988).

Both histological subtypes of OC have different pathological and aetiological characteristics with ADC being more prevalent in Caucasians and SCC in African-Americans and Blacks (Lu 2000). OC is an extremely aggressive form of cancer and patients with the disease have a

poor prognosis. The disease is most often seen in persons over the age of 25, with mortality rates increasing steadily with age (Blot 1994). It is also more common in men, with males being two to four times more likely to develop OC than females.

1.1.2 Diagnosis and development of OC

Patients with OC present with a variety of symptoms, such as dysphagia (difficulty to swallow), loss of weight, nausea and vomiting, coughing and hematemasis (vomiting of blood). A substantial loss of weight is characteristic of poor nutrition resulting from the presence of the cancer, and can be an indicator of the patient's prognosis (Fein *et al.* 1985). Tumour development can result in disruption of normal peristaltic events, which can lead to nausea, vomiting and regurgitation of food. Hematemasis is usually a result of the fragility of the tumour surface leading to bleeding (Ojala *et al.* 1982).

The majority of patients with SCC only begin to develop the above symptoms when the tumour has grown to a size large enough to result in a noticeable obstruction of the oesophagus. Due to this fact, the age at which patients are first diagnosed tends to be more advanced than in other common cancers and means that the prognosis is extremely poor, usually less than six months (Enzinger and Mayer 2003).

In contrast, ADC, which is the predominant cause of OC in the developed world, is generally caused by acid reflux from the stomach. These patients experience symptoms such as frequent heartburn and nausea. They tend to consult physicians earlier than patients with SCC, and as a result the disease is detected at an earlier stage and their prognosis is much improved. ADC is an extremely lethal form of OC, and although survival rates have

increased over the past few years, there is only about a 10% five year survival rate in most Western countries (Sundelöf *et al*. 2002).

Diagnostic tests are available to patients presenting with the above-mentioned symptoms to determine if OC may be the cause. The most commonly used of these tests is a barium swallow or oesophagram, which uses a series of X-rays to examine the oesophagus. Barium is swallowed and coats the lining of the oesophagus so that it is clearly visible on an X-ray (Levine *et al*. 1997).

Another common diagnostic test is oesophagoscopy (upper endoscopy). During this procedure, the inside of the oesophagus is examined using an endoscope. Any unusual growths are clearly noticeable as changes from the surrounding tissue. A tissue biopsy may also be performed at the same time (Kakushima and Fujishiro 2008).

The development of most cancers, including OC, may be linked to the spread of the cancer cells to the surrounding lymph nodes and subsequently to other areas of the body (Siewert *et al.* 2001). OC is unusual with respect to other cancers as it spreads to the lymph nodes in an extremely erratic manner, making predictions about disease progression very difficult. OC is classified according to the 2002 American Joint Committee on Cancer tumour-node-metastasis (TNM) classification system (Greene *et al.* 2002). This system classifies OC based on the characteristics of the primary tumour as well as tumour metastases. Different stages in the development of OC are widely recognized and utilized in determining the most effective treatment plan for the patient. These stages are as follows: *stage 0***,** the cancer, also called non-invasive cancer or high grade dysplasia, has not spread to other parts of the body; *stage I***,** the cancer occurs only in the top layer of cells lining the oesophagus; *stage II***,** the cancer has invaded deeper layers of the oesophageal lining and may have spread to neighbouring lymph

nodes; *stage III***,** the cancer has spread further into the wall of the oesophagus and to nearby tissues or lymph nodes; *stage IV***,** the cancer has spread to other parts of the body.

1.1.3 Demographics and aetiology of OC

OC occurs in high frequencies around the world and is one of the leading causes of cancer-related deaths worldwide. Incidence of OC shows marked geographic variation, occurring at high frequencies in so-called "oesophageal cancer belts". These regions are separated into the Asian belt, which is made up of countries such as Iran, Iraq, China, Japan and Turkey. In the Caspian Sea region of Iran for example, cancers of the gastrointestinal (GI) tract (including OC) account for over half of all cancer related deaths (Mohebbi *et al.* 2008). Other areas of high incidence include France, South America and South and East Africa (Parkin *et al*. 2005). Oesphageal cancer is common in Africa, particularly in the Transkei region of South Africa, which is thought to be the centre of the disease in Africa (Sammon 2007). An age-standardised incidence rate (ASIR) of 46.7 and 19.2/100,000 for males and females respectively was reported in this region (Makaula *et al*. 1996).

SCC occurs at a higher frequency than ADC, with the highest incidence of ADC in North America and France. Whilst the incidence of SCC worldwide has remained relatively constant over the last few decades, cases of ADC have increased markedly. The reported increase in the white male population of the USA is reported to be close to 10%, which makes it the fastest growing form of cancer in that specific population (Pisani *et al.* 1999). Asian countries such as China (Zhang *et al*. 2004) and Singapore (Fernandes *et al*. 2006) have also shown an increase in incidence.

As well as geographical differences between the two subtypes of OC, there has been found to be a difference in racial background. SCC is more prevalent in Blacks with ADC occurring

more often in Caucasians (defined as individuals from European descent). SCC is thought to be one of the leading causes of death among males of the Black population and the fourth most common cause of death among Coloured (defined as individuals of Mixed Ancestry) males in South Africa (Blot 1994).

OC also shows variation in incidence in differing age groups with an increased risk with increasing age. The average age of diagnosis in South Africa is around 60 years of age (Pisani *et al.* 1999).

1.1.4 Pathogenesis of OC

1.1.4.1 Environmental risk factors

The marked differences in the geographic and ethnic incidences of OC have been hypothesised to be due to variations in environmental factors in different populations originating from different regions (Marasas *et al*. 1988).

(i) Diet and nutrition

Particular diets from certain regions in the world may result in deficiencies of vitamins and micronutrients. Especially implicated are diets lacking vitamins such as B1 (riboflavin) and the mineral selenium. Individuals with lower levels of selenium have been shown to have an increased risk of developing OC (Wei *et al*. 2004, Cai *et al*. 2006).

In a global study, low intakes of fruit and vegetables were found to account for 20% of all cases of OC and 19% of cases of gastric cancers worldwide (Lock *et al*. 2004). This is

thought to be due to the fact that the anti-oxidants, minerals and micronutrients present in fruits and vegetables, suppress the action of carcinogens and prevent oxidative DNA damage (Farrow *et al.* 2000).

Areas of the Tanskei region of South Africa have diets based primarily on cereal grains, which are low in nutrients such as zinc and iron. As well as being relatively low in certain nutrients, food grains can also harbour fungal contamination. Fungal mycotoxins are a commonly known risk factor for the development of OC (Marasas *et al*. 1988). Fumonisin B1 is a product of *Fusarium moniliforme* and is found on corn and maize throughout the world (Sammon and Iputo 2006). In Iran (Shephard *et al*. 2002), China (Yoshizawa *et al*. 1994) and Transkei (Marasas 1979) the mycotoxin was found at higher levels than in other areas with a lower incidence of OC. It has been hypothesized that the fungi themselves, as well as their mycotoxins, are mutagenic and directly affect DNA synthesis *via* sphingolipid metabolism in the cell (Abnet *et al*. 2001). A study by Lim *et al*. (1996) showed that the administration of high doses of fumonisin B1 in rats was able to stimulate the proliferation of oesophageal cells.

Food infected with fungi may also contain N-nitrosamines, which have been shown to be carcinogenic to the oesophageal cells (Lijinsky *et al.* 1981). In combination with diethylnitrosamine, fumonisin B1 proved to be more toxic to oesophageal epithelial cells than alone (Myburg *et al.* 2002).

In addition to a diet based on cereal grains, the population of Tanskei routinely supplements their diet with wild herbs and vegetables. *Solanum nigrum* (garden nightshade) is one of the commonly used plants, and has been shown to be used more often in areas with a higher incidence of OC (Rose 1982). A study of rats fed on *S. nigrum* resulted in an increase in the cells lining the oesophagus (Purchase *et al.* 1975), and a case-control study conducted by

Sammon (1992) in the Transkei demonstrated that the risk for developing OC was higher for individuals consuming *S. nigrum* than for individuals that smoked.

Other potential dietary risk factors include high intake of dietary fat, cholesterol and animal protein, although these factors are not limited to OC alone (Mayne *et al*. 2001).

(ii) Alcohol consumption and tobacco use

The combined use of tobacco and alcohol has long been known to be associated with an increased risk of cancer development worldwide. OC is no exception, with the risk of disease development increasing with greater consumption of the tobacco product (Yu *et al*. 1988). In developed countries of the world with a high incidence of OC, alcohol and tobacco use (especially in combination) is the greatest risk factor (Day *et al*. 1994). The majority of areas in which OC is endemic, such as China, France and Italy, show a degree of association with tobacco use (Guo *et al*. 1994, Parkin *et al*. 1994). In South Africa a case control study by van Rensburg *et al*. (1985) showed a highly significant association ($P = 0.0001$) between cigarette smoking and OC. In the Black population of Soweto, South Africa, an increased risk for OC was observed for consumers of homegrown tobacco (either as hand-rolled cigarettes or chewing tobacco) when compared to cigarette smokers (Segal *et al*. 1988). In Transkei, tobacco cultivation, and its subsequent use, is lower in areas of low OC incidence and *vice versa* (Bradshaw and Schonland 1969). Other studies have demonstrated that 72% of OC patients in this region smoke (Sammon 1992) and 80% are consumers of traditional home-brewed African beer (Segal *et al*. 1988). Alcohol use irritates the lining of the oesophagus, leading to inflammation that eventually may cause malignant changes in the cells.

(iii) Viral risk factors

The human papilloma viruses (HPV) are a class of DNA viruses first shown to be implicated in the development of OC in a study conducted by Syrjanen (1982). To date, more than 70 different types of the virus have been identified, and the association between HPV and OC confirmed by techniques such as polymerase chain reaction (PCR).

In areas endemic for OC such as China and South Africa, the virus is frequently detected in patients with OC. In certain regions of China, 64% of OC patients were shown to have HPV DNA (Togawa *et al*. 1994). In South Africa, one study found evidence of HPV DNA in 10 out of 14 OC patients studied (Williamson *et al*. 1991). In Transkei, 46% of OC patients were shown to be positive for HPV DNA from viral subtypes normally regarded as a low risk for OC pathogenesis (Matsha *et al*. 2002).

In contrast, regions with a low incidence of OC (such as the USA and Northern parts of Europe) show relatively little HPV activity (Lam 2000).

(iv) Gastro-oesophageal reflux and Barrett's oesophagus

Gastro-oesophageal reflux disease (GORD) is widely accepted as the risk factor most responsible for ADC. It is characterized by the movement of the stomach contents into the lower region of the oesophagus (Lagergren *et al*. 1999). Other factors influencing the development of GORD are ulcers in the oesophagus, hernias or difficulty in swallowing. Barrett's oesophagus is a complication that arises due to long standing GORD (Goldblum 2003) and results in replacement of the stratified squamous epithelium of the oesophagus and oesophagogastric junction with different types of columnar epithelium. This occurs as a result of inflammation (oesophagitis) of the lining of the oesophagus from the acidic stomach

contents. Paull *et al.* (1976) described three different types of epithelium that occur in patients with Barrett's oesophagus: i) fundic-type, ii) cardiac-type (junctional) and iii) specialized columnar epithelium. Only the specialized columnar epithelium is susceptible to developing ADC (Chen and Yang 2001).

(v) Obesity

Obesity [body mass index (BMI) >30] is increasing at a rapid rate in western populations (11-15% in men and 15-25% in women), and numerous studies have shown that this increase parallels the rise in OC (Seidell 1997, Chow *et al.* 1998). Increased intra-abdominal pressure in obese individuals may be a cause of gastro-oesophageal reflux, along with decreased emptying of the stomach contents and a lower oesophageal sphincter pressure.

Chow *et al*. (1998) reported that having a BMI in the upper fourth of the scale compared with the lower fourth was associated with a dramatic increase in OC development. Lagergren *et al*. (1999) found similar results in a Swedish cohort.

(vi) Other environmental factors associated with OC

The consumption of hot beverages resulting in injury to the oesophagus accompanied by inflammation of the surrounding tissues has been implicated as a risk factor for OC (Yang and Wang 1993).

Individuals exposed to the by-products of the mining and industrial sector in the workplace show an increased risk of SCC development. These substances include liquids (eg. benzene and xylene), dust (eg. carbon black) and aromatic hydrocarbons (Parent *et al*. 2000).

Exposure to asbestos, a known carcinogen, has been linked to the development of many types of gastrointestinal cancers including OC (Selikoff *et al.* 1979). This arises predominately due to chronic inflammation of the tissues lining the upper respiratory tract.

1.1.4.2 Genetic risk factors

As previously discussed, OC occurs predominately in two forms, SCC and ADC. It is common knowledge that the underlying causes of complex diseases such as cancer are related to both genetic and environmental factors. In both SCC and ADC the genetic risk factors involved are poorly defined and understood. However, recent studies have led to the discovery of various chromosomal abnormalities and gene alterations that may shed some light on changes at a genetic level that may contribute to the development of OC.

Loss of heterozygosity (LOH) has proven to be an informative indicator of tumour-specific genetic alterations and is used in the analysis of human cancer cells as a method to define regions of the genome that may contain genes involved in tumour pathogenesis (Lasko *et al.* 1991). The tumour-suppressor function of the retinoblastoma (*Rb*) gene on chromosome 13q14 and the p53 tumour suppressor (*p53*) gene on chromosome 17p13 was elucidated using LOH as an approach (Huang *et al.* 1988, Chen *et al.* 1990). The most common genetic alterations occurring in OC that have been reported using LOH are allelic losses at chromosomes 3p, 5q, 9p, 13q, 17p, 7q, and 18q (Huang *et al.* 1992, Tarmin *et al.* 1994, Barrett *et al.* 1996, Shimada *et al.* 1996).

Allelic loss on chromosome 17p occurs at a high frequency in many human cancers (Hollstein *et al.* 1996). This loss often includes the *p53* locus which results in the inactivation

of the tumour suppressor function of this gene. With regard to OC, studies have shown increased frequencies of *p53* mutation in one allele of tumours that have retained both 17p alleles (Huang *et al.* 1993, Maesawa *et al.* 1994). This has led to the belief that the aberrant p53 protein exerts a dominant negative effect which is sufficient to interfere with the function of the remaining wild-type allele.

Alterations of genes involved in the regulation of cell proliferation are common in OC. These include amplifications of the epidermal growth factor (EGF) receptor gene (*EGFR*) (Slamon *et al.* 1987, Zhou *et al.* 1994) involved in growth signal reception, amplifications of *c-myc* (Lu *et al.* 1988) and overexpression of *cyclin D1* (Wang *et al.* 1994, Adelaide *et al.* 1995) involved in cell-cycle control.

To date, nonepidermolytic palmoplantar keratoderma, also referred to as tylosis, is the only recognised familial syndrome responsible for the predisposition of SCC to patients. Tylosis patients are at a 95% risk of developing SCC by the age of 70 (Ellis *et al.* 1994). This rare disorder is characterised by thickening of the oral mucosa as well as increased keratinisation of the palms of the hands and soles of the feet. It is caused by a genetic abnormality of chromosome 17q25 and is inherited in an autosomal dominant manner (Risk *et al.* 1994). It remains to be determined if this gene plays a role in the aetiology of sporadic OC.

There have also been studies that have reported genetic susceptibility of Barrett's oesophagus and GORD with ADC (Eng *et al.* 1993).

1.2 IRON AND OC

1.2.1 Mechanisms of iron carcinogenesis

Iron is an essential nutrient required for a variety of cellular functions. However, when present in an amount that exceeds the requirements of the body, iron can be highly toxic to cells and tissues and iron homeostasis must therefore be tightly regulated. In many cases, an excess of iron may initiate the process of carcinogenesis. The toxicity of iron is largely related to its ability to catalyze the formation of free radicals which attack and damage cellular components leading to cell death and tissue injury. The manner in which free radicals are formed is based on the chemistry of the Fenton and Haber-Weiss reactions. During these processes, catalysis of iron results in hydroxyl radicals (OH^-) from superoxide (O_2^-) and hydrogen peroxide (H_2O_2), which are termed reactive oxygen species (ROS) (Papanikolaou and Pantopolous 2005).

A surplus of iron may lead to an increase in the oxidative stress of the cell resulting in accelerated tissue damage as well as the oxidation of proteins, membrane lipids and nucleic acids. It has been demonstrated that the increased oxidative stress resulting from excess iron in the liver, pancreas and skin may lead to an elevated risk for carcinomas and sarcomas respectively (Weinberg 1999). Increased levels of iron are also required for the sustained proliferation of tumour cells. Hann *et al.* (1990) showed that supplementation with iron enhanced the growth of human hepatoma cells. Conditions of iron overload may also impair the cytotoxic activity of tumouricidal-activated macrophages against tumour cells by inhibiting their growth (Weiss *et al.* 1992). In normal circumstances, loss of iron from a target cell results in anti-tumour activity, which is reduced when there is excess iron present

(Huot *et al*. 1990). For this reason, iron can be said to have an indirect carcinogenic effect on cells.

Conditions of iron deficiency can be just as pathogenic to bodily tissues and cellular components as those of iron overload. Anaemia is a condition frequently associated with malignancies although the underlying mechanism of pathogenesis is not yet fully understood. Erythrocytes often have a shortened survival rate in cancer patients which is one of the causes of cancer-related anaemia. Erythropoiesis is unable to compensate for this due to the weakened response of the bone marrow of cancer patients (Zucker *et al*. 1974).

Characteristic changes in iron homeostasis are often observed in cancer patients. Some of these include elevated serum ferritin levels and decreased serum iron concentrations, which suggest a movement of iron toward the storage sites (Dorner *et al.* 1983). Raised levels of serum ferritin have been proven to correspond with tumour progression in head and neck cancers (Rosati *et al.* 2000).

1.2.2 Iron as a risk factor for OC

The role of iron as a potential risk factor for the development of OC was previously described in Black South African patients who had dietary iron overload, hypothesised to be as a result of drinking traditional beer brewed in non-galvanised steel drums (Bothwell *et al*. 1964, MacPhail *et al.* 1979, Isaacson *et al*. 1985). The involvement of iron in the development of other cancers such as liver cancer has also been previously studied and attributed to dietary iron overload (Mandishona *et al*. 1998). Studies have also shown that an increase in dietary iron plays a role in the development of OC in groups other than the Black South African population (Amer *et al*. 1990, Rogers *et al.* 1993). For example, a study conducted on a Danish population with primary haemochromatosis and iron overload demonstrated that these

patients were at an increased risk of developing certain cancers, including OC (Hsing *et al.* 1995). The study subjects were followed from the date of haemochromatosis diagnosis until the first date of cancer diagnosis or date of death, and the association between primary haemochromatosis and cancer was subsequently quantified.

Further studies using rat models showed that iron supplementation could be indicated as a risk factor for developing OC. The rats that were supplemented with iron had increased levels of inflammation, cell proliferation and ROS compared to the control animals. Increased tumour development, subsequent to the development of Barrett's oesophagus, in the lower regions of the oesophagus of the iron-supplemented rats was also noted (Goldstein *et al.* 1998, Chen *et al.* 1999, 2000).

1.3 IRON HOMEOSTASIS

1.3.1 Iron distribution and circulation

Iron is a trace element required by virtually all living organisms and is utilised in a variety of cellular and metabolic processes (Aisen *et al.* 2001). These include oxygen transport by the haem moiety of haemoglobin, electron transport on the cytochromes of the respiratory chain and various other enzymes, a few of which are involved in DNA synthesis (Lieu *et al.* 2001).

Within the adult human body, iron constitutes approximately 35 mg/kg of body weight in women and a slightly higher level of 45 mg/kg in men (Andrews 1999) and is one of the most abundant metals in the body. The vast majority of the total body iron, about 60 to 70%, is bound to the haemoglobin proteins of the erythrocytes which circulate in the blood stream. A further 10 to 15% of iron is present in other enzymes and the cytochromes, but in normal circumstances this level never exceeds 4 to 8 mg of iron. In the plasma approximately 1% of

iron is transported by forming a complex with transferrin, an 80 kDa protein that contains two iron binding sites (Emerit *et al.* 2001). Transferrin-bound iron (TBI) is utilised predominately by the bone marrow to produce the haemoglobin of the erythroid cells. The remainder of the total body iron, 20 to 30%, which is surplus to immediate cellular requirements, is stored in ferritin (Conrad *et al.* 1999). Due to the high demand for iron by the erythrocytes (20 mg per day) for erythropoiesis, the vast majority of iron comes from the destruction of old red blood cells by reticuloendothelial macrophages. This macrophage iron recycling results in the release of haem molecules from the haemoglobin into circulation (Bottomley *et al.* 1995).

1.3.2 Dietary iron uptake

The regulation of the stores of iron within the body takes place primarily through the absorptive process in the duodenum and jejunum of the small intestine. This intestinal absorption of dietary iron is important in maintaining the iron balance as the body has no physiologic pathway to regulate iron excretion. Iron loss takes place to a small extent (1 mg/day in an adult human) by excretion of iron in the urine and bile, sweating and the recurrent loss of cells from the skin and gut. Due to the regular blood loss during menstruation and childbirth, women lose additional iron from the high concentrations contained in the haemoglobins (Andrews *et al*. 1999). The loss of bodily iron *via* these processes is kept in balance by the absorption of 1 to 2 mg of iron from the diet daily. The relationship between the excretion and the absorption of iron leads to the maintenance of a relatively constant amount of stored iron throughout life.

As stated previously, the regulation of iron takes place predominately at the level of absorption in the intestine. Specifically, dietary iron is taken up across the brush border of the intestinal enterocytes and subsequently released across the basolateral membrane into the

circulation. Iron exists in two forms, the ferric (Fe^{3+}) and the ferrous (Fe^{2+}) forms. Due to the fact that ferric iron is rendered insoluble at a pH level greater than 3, and ferrous iron remains soluble at pH 7, the absorption of ferrous iron is more efficient than that of ferric iron. Inorganic iron ingested from the diet exists in the oxidised ferric (Fe^{3+}) form and therefore has to be reduced to the ferrous (Fe^{2+}) at the apical membrane of the enterocytes before absorption of iron in the intestine can occur (Conrad *et al.* 1999).

1.3.2.1 Non-haem iron uptake

The small intestine is the site where the absorption of all iron from the diet occurs. The cells located on the intestinal villus, the enterocytes, are highly specialized cells that control the absorption of dietary iron, as well as its transfer to the circulation. The brush border ferrireductase enzyme, cytochrome b reductase 1 (CYBRD1), also referred to as duodenal cytochrome b (DCYTB), mediates the reduction of the ferric iron to ferrous iron (McKie *et al.* 2001). The divalent cation transporter 1 (DCT1), also known as divalent metal transporter 1 (DMT1) or natural resistance-associated macrophage protein 2 (NRAMP2), transports the ferrous iron across the apical membrane, and into the lumen, of the enterocytes (Fleming *et al.* 1997). NRAMP2 acts as a proton-coupled divalent cation transporter (Gunshin *et al.* 1997), and the low pH of the gastric environment provides a proton-rich environment which facilitates this transport.

The inorganic iron released from haem (refer to section 1.3.2.2) or imported *via* DMT1 into the cytosol of the enterocyte enters the labile iron pool. It may then be stored within the cell as ferritin, or transported across the basolateral membrane into the circulation. The solute carrier family 40 (iron-regulated transporter) member 1 protein (SLC40A1), also referred to as the solute carrier family 11 (proton-coupled divalent metal ion transporter) member 3

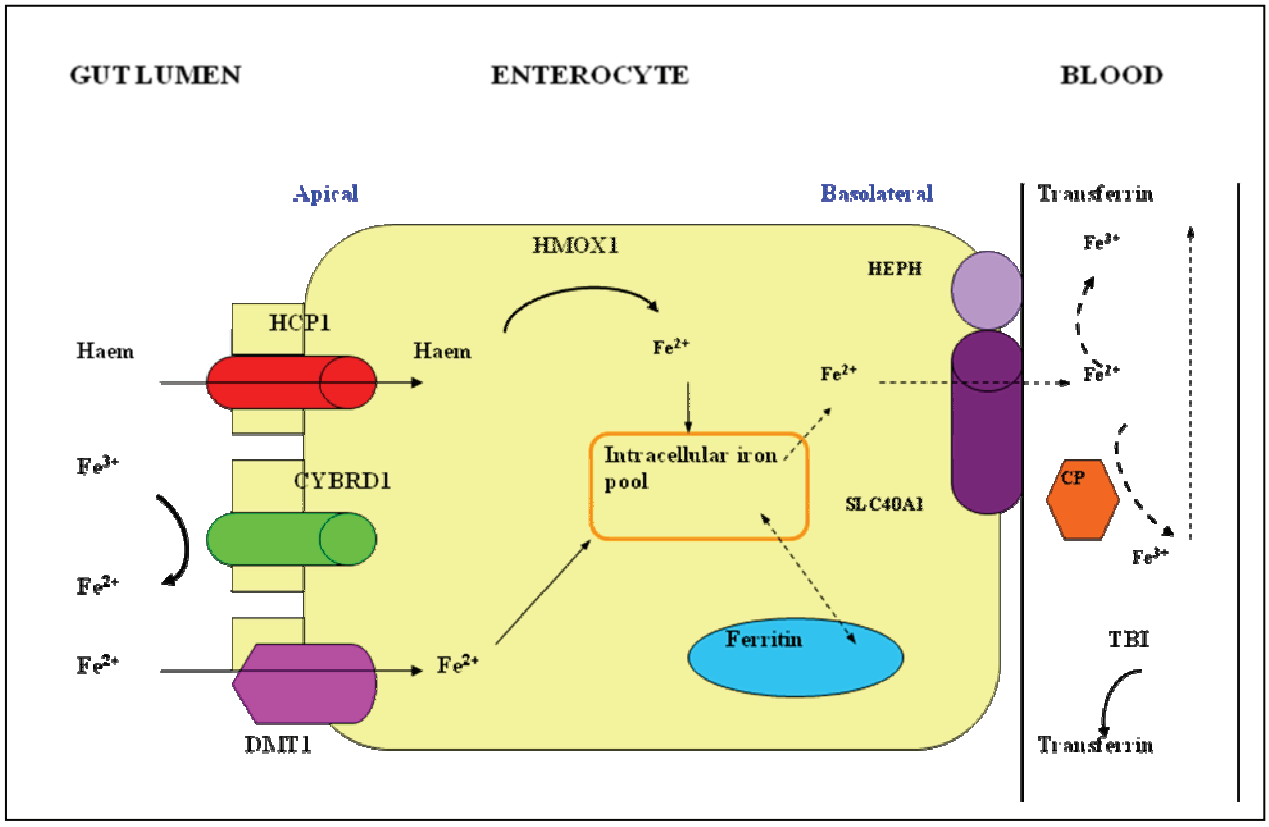
protein (SLC11A3), or ferroportin 1 (FPN1), or the iron-regulated transporter 1 (IREG1) or metal transporter 1 (MTP1), is responsible for mediating the transport of iron across this membrane (Donovan *et al.* 2000). SLC40A1 works in conjunction with the proteins ceruloplasmin (CP) and hephaestin (HEPH). CP and HEPH are homologous multicopper ferroxidase proteins responsible for the oxidation of ferrous iron to ferric iron. HEPH is a membrane-bound protein located on the basolateral membrane of the enterocytes and CP is located primarily in the plasma (Harris *et al.* 1998). These two proteins are thought to function as aids in iron transport by creating ion gradients that favour the export of iron from cells (McKie *et al.* 2000). CP in particular is hypothesised to aid with the binding of iron in the ferric state to its plasma transporter, transferrin (Harris *et al.* 1998).

1.3.2.2 Haem iron uptake

Iron from haemoglobin contained in food is more efficiently absorbed than inorganic iron (Majuri and Grasbeck 1987). For this reason, haem enters the enterocyte in a different pathway than that of inorganic iron (Conrad et al. 1999). In the intestinal lumen, haem is enzymatically cleaved from haemoglobin or myoglobin, and enters the enterocyte as a metalloporphyrin (reviewed by Anderson *et al.* 2005). A haem carrier protein-1 (HCP1) has been identified and is believed to bind to haem and transport it across the apical membrane of the enterocytes (Shayeghi *et al.* 2005). It is thought that the haem-HCP1 complex enters the cell *via* receptor-mediated endocytosis and progresses to the endoplasmic reticulum (ER) (Shayeghi *et al.* 2005). Haem oxygenase 1 (HMOX1), located on the surface of the ER, degrades the haem and releases inorganic iron, which is either stored as ferritin or enters the circulation across the basolateral membrane (refer to section 1.3.2.1). Recently, it has been demonstrated that HCP1 may have a more defined role as a folate transporter in intestinal

tissues (Qiu *et al*. 2006). However, HCP1 is expressed in many tissues other than the intestine and subsequently its role in haem transport cannot be disregarded.

Figure 1.1 A schematic representation of dietary iron uptake by the enterocyte.



Legend to Figure 1.1

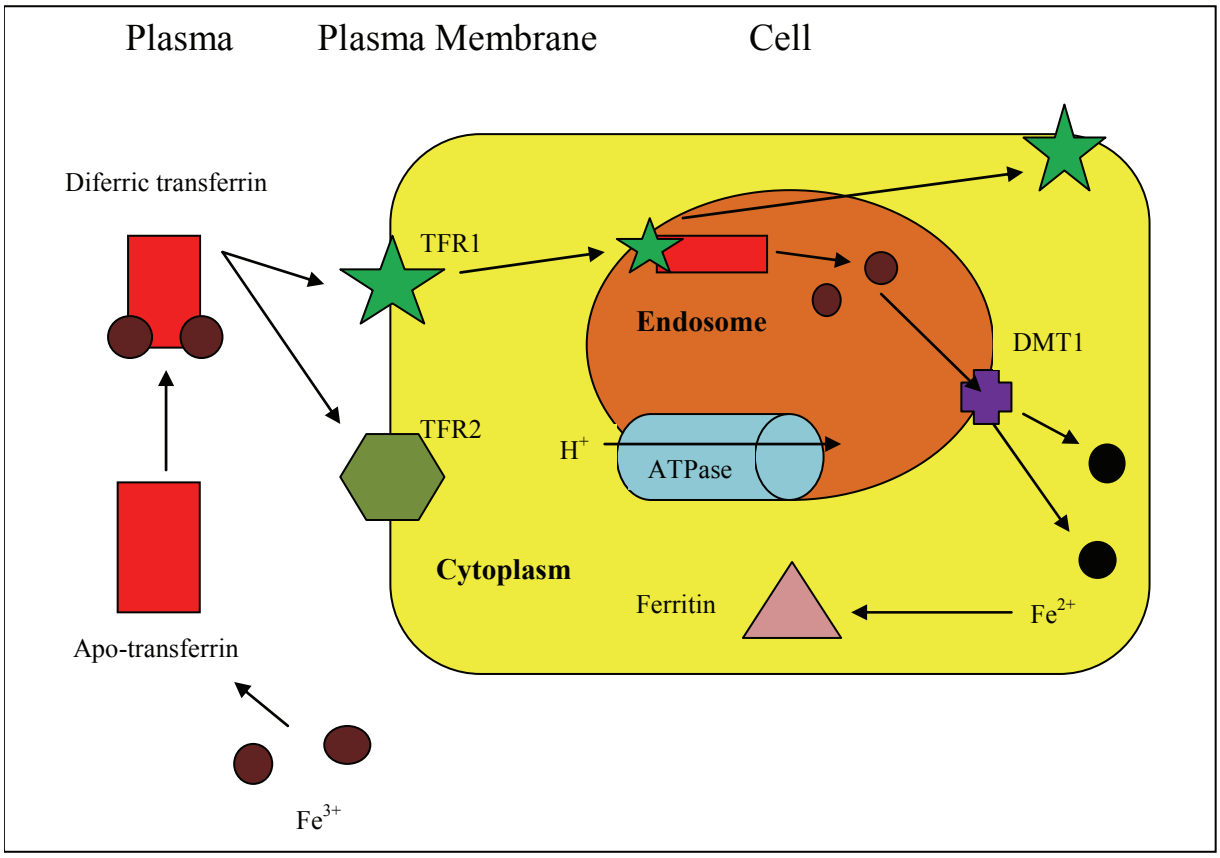
Dietary iron in the intestinal lumen is reduced by CYBRD1 from the ferric form (Fe^{3+}) to the ferrous form (Fe^{2+}). Fe^{2+} is transported across the apical membrane by DMT1. Haem is enzymatically cleaved from haemoglobin and transported into the enterocyte *via* HCP1. HMOX1 releases ferrous iron from haem. The intracellular iron is either stored as ferritin or transported out of the cell by SLC40A1, which is located on the basolateral membrane. HEPH (membrane-bound) and CP (in the plasma) facilitate the export of iron by oxidising iron from Fe^{2+} to Fe^{3+}, which subsequently binds to transferrin. **Abbreviations:** CP, ceruloplasmin; CYBRD1, cytochrome b reductase 1; DMT1, divalent metal transporter-1; Fe^{3+}, ferric iron; Fe^{2+}, ferrous iron; HCP1, haem carrier protein-1; HEPH, hephaestin; HMOX1, haem oxygenase-1; SLC40A1, solute carrier family 40 (iron-regulated transporter) member 1; TBI, transferrin-bound iron. Adapted from Trinder *et al*. 2002.

1.3.3 Cellular iron uptake

Iron is transported in the plasma bound to serum transferrin, which has a high affinity for ferric iron. Once bound to transferrin, iron is no longer toxic, as it is unable to generate free radicals in this state (Hoefkens *et al.* 1996). Transferrin is a glycoprotein consisting of two globular domains, each of which has a high-affinity binding site for an iron molecule (Yang *et al.* 1984). Transferrin exists in a mixture of three different states, iron-free (apo-transferrin), one iron molecule bound (monoferric transferrin) and two iron molecules bound (diferric transferrin).

Two types of transferrin receptors (TFRs) have been identified and isolated, TFR1 and TFR2 (Kawabata *et al.* 1999). These receptors are located on the surface of cells and bind with a high affinity to diferric transferrin in a manner that is dependent on pH. The majority of cells, except mature erythrocytes, express TFR1 (Davies *et al.* 1981). In contrast to this ubiquitous expression profile, TFR2 is predominately expressed by the hepatocytes of the liver where TFR1 expression is relatively low (Kawabata *et al.* 1999). At the normal physiological pH of the body, TFRs have a high affinity for binding to diferric serum transferrin. The complex of TFR and transferrin is internalised by the process of receptor-mediated endocytosis. Once inside the cell, the endosomal lumen acidifies to a pH of around 5.5 *via* an ATPase proton pump, and as a result the binding of iron to transferrin is weakened and the iron is released. The transferrin molecules bound to the TFRs return to the surface of the cell where the neutral pH of the blood facilitates the release of apo-transferrin into the circulation to be re-used. Once the ferric iron has been released from transferrin, it is reduced to ferrous iron before it passes into the cytoplasm of the cell *via* DMT1 located on the endosomal membrane (Fleming *et al.* 1998). The iron that is now in the cytoplasm of the cell can be utilised by the cell or used for the synthesis of haem.

Figure 1.2 A schematic representation of cellular iron uptake.



Legend to Figure 1.2

In the plasma, ferric iron is bound by apo-transferrin to form diferric transferrin. In most cells, diferric transferrin binds to TFR1 (or TFR2 in the hepatocytes of the liver) on the cell surface. This complex is internalised *via* receptor-mediated endocytosis. An ATPase proton pump decreases the pH of the endosome and iron is released from transferrin. Iron is then transported over the endosome membrane into the cytoplasm *via* DMT1. Apo-transferrin TFR complexes are recycled back to the cell surface for another cycle of iron uptake. **Abbreviations:** ATP, Adenosine Triphosphate; DMT1, divalent metal transporter 1; Fe^{3+}, ferric iron; Fe^{2+}, ferrous iron; H^{+}, proton; pH, percentage hydrogen; TFR1, transferrin receptor 1; TFR2, transferrin receptor 2. Adapted from Lieu *et al*. 2001.

1.3.4 Iron storage

Iron that is not utilised immediately by the body for metabolic processes is stored as a reserve in the event that the levels of free body iron become low. In healthy individuals, approximately 20 to 30% of the total body iron is stored. Free iron is extremely toxic and aggregates to form toxic precipitates. To prevent this, iron is stored as ferritins and haemosiderins in reticuloendothelial macrophages and hepatocyte cells of the liver (Sargent *et al.* 2005).

1.3.4.1 Hepatic iron storage

The hepatocyte cells of the liver are the main site of iron storage within the body. Iron gains entry to these cells bound to transferrin *via* the TFRs located on the cell surface. If the iron levels within the cell exceed the cellular needs, the surplus is stored as ferritin predominately with a small amount stored as haemosiderin (Trinder *et al.* 2002).

Ferritin exists as two subunits consisting of heavy (H for heavy or heart) and light (L for light or liver) chains that can store up to 4500 molecules of ferric iron (Theil 1998). H-ferritin has the ferroxidase ability to oxidise Fe^{2+}, the predominant form of iron in the cytoplasm, to Fe^{3+}, the preferred iron form for storage. Ferritin synthesis is induced in the presence of free iron and repressed in conditions of iron deficiency (Zahringer *et al.* 1976, Ke *et al.* 1998).

The mechanism by which iron is released from ferritin is poorly understood to date; although it has been hypothesised that ferritin is degraded by lysosomes or protoeosomes in order to provide iron for systemic requirements (Aisen *et al.* 2001).

Haemosiderins are poorly defined in comparison to ferritin. They are thought to be iron-protein complexes that form an insoluble iron storage system, and are derived from the degradation of ferritin. In normal homeostatic conditions, haemosiderin can be detected in low amounts in the body tissues. However, during primary and secondary iron overload haemosiderin levels increase dramatically (Sargent *et al.* 2005). Haemosiderin has been shown to be less effective than ferritin at generating free radicals *via* the Haber-Weiss-Fenton reaction, which could suggest a reason for the increase in haemosiderin levels during iron overload (O'Connell *et al.* 1986).

1.3.4.2 Reticuloendothelial iron storage

There are two mechanisms by which reticuloendothelial (RE) macrophages acquire iron. The first is *via* the TFR-transferrin complex (Testa *et al.* 1991) on the cell surface and the second is through the process of phagocytosing old erythrocytes (Deiss 1983). Haem contained in the erythrocytes is cleaved enzymatically by haem oxygenase (HMOX1) whereby it is either released, with the aid of SLC40A1 as previously described, into the plasma to be bound to transferrin or it is stored as ferritin within the macrophages.

1.3.5 Regulation of iron homeostasis

As previously mentioned, the body has no fixed pathway to deal with the excretion of iron that exceeds the levels needed by the body. In order to prevent conditions such as iron overload or anaemia, complicated control pathways exist to monitor the amount of iron that is absorbed or stored within the body. For this reason the majority of the proteins mentioned in relation to iron metabolism are under extremely tight genetic control and are up- or down-

regulated depending on the levels of iron within cells and the availability of iron to cells. These control mechanisms involve changes at the level of transcription, translation, as well as post-translational modifications at the cellular level. With regard to the organism as a whole, iron homeostasis is dependent on mechanisms that monitor iron stores, the rate of iron use by the erythrocytes (Finch 1994) and hypoxia (Trinder *et al.* 2002).

In order to explain the regulation of iron, two models have been proposed by Pietrangelo (2004) which include the crypt-programming model and the hepcidin model. These two models and post-translational control of iron regulation will briefly be discussed.

1.3.5.1 Crypt programming model

Intestinal cells of the duodenum called crypt cells are precursor cells that migrate onto the intestinal villi and differentiate to become mature enterocytes. These precursor cells are believed to sense the iron requirements of the body and regulate the absorption of dietary iron by the absorptive villus enterocytes accordingly.

This model proposes that the crypt cells absorb iron from the plasma so that their intracellular iron levels match that of the body iron stores. This means that they are able to regulate the intestinal absorption of iron from the gut lumen as they migrate up the villi and mature into absorptive cells on the brush border themselves (Oates *et al*. 2000).

TFR1 and TFR2 proteins mediate the uptake of transferrin-bound iron (TBI) from the plasma, and are found on the basolateral membrane of crypt cells. The high iron protein (HFE) is a heterodimeric membrane protein with the ability to form associations with TFR1 and is highly expressed in the crypt cells (Parkkila *et al.* 1997). Experiments by Waheed *et al.* (1999) demonstrated that by binding to TFR1 and modifying its expression, HFE could change the iron sensory function of the precursor cell. Other experiments have shown that

wildtype HFE binds to TFR1 and lowers the binding of TBI therefore lowering the levels of iron uptake. This leads to a decrease in intracellular iron concentrations and ferritin levels and an increase in the number of transferrin receptors that are present on the cell surface (Feder *et al*. 1997, Waheed *et al.* 1997).

The *Hfe* knockout mouse model used in a study by Trinder *et al.* in 2002 showed that TFR1-mediated uptake of iron from the plasma is decreased by the mutant *HFE*. These results add support to the hypothesis of the crypt cell model by demonstrating that normal HFE is responsible for regulating the uptake of TBI from the plasma by inducing TFR1 expression, but are as of yet not conclusive evidence.

1.3.5.2 The Hepcidin model

Hepcidin is a peptide protein synthesized predominately by the hepatocyte cells of the liver, with expression levels being dependant on the stage of hepatocyte differentiation (Papanikolaou and Pantopoulos 2005). It has antimicrobial properties and is thought to play an important role in maintaining iron homeostasis, mainly as a regulator of iron stores, but it is also able to act as an erythroid regulator (Ganz 2003).

A study by Nicolas *et al.* (2001) using upstream stimulatory factor 2 (USF2) knockout mice unexpectedly demonstrated that the mice progressively developed iron overload in the tissues of the liver and pancreas, whilst the RE macrophages showed no increase in iron levels. These symptoms appeared to mimic those of haemochromatosis patients. It was found during further analysis that a recombination event had removed both *USF2* and *HAMP* genes, and that it was in fact the resulting deficiency of hepcidin that was responsible for the observed phenotype. In later studies by the same group, over-expression of hepcidin-1 in transgenic mice was shown to result in severe iron-deficiency anaemia (Nicolas *et al.* 2002).

Low levels of hepcidin trigger an increased absorption of iron from the intestine as well as the release of iron from macrophages. In contrast, an excess of hepcidin results in a decrease of dietary iron uptake and the retention of iron by macrophages (Andrews 2008). In response to hypoxia and anaemia the expression of hepcidin is decreased, irrespective of the level of body iron stores. Increases in hepcidin levels have also been noted in humans and mice with inflammation, which has resulted in the theory that hepcidin may be responsible for the anaemia of chronic disease (Nicolas *et al*. 2002).

Hepcidin appears to regulate iron efflux from cells by binding to ferroportin and internalising it (Nemeth *et al*. 2004). Rising levels of hepcidin, as a result of iron overload or inflammation, lower the rate of export of iron from macrophages and enterocytes. In conditions such as haemochromatosis when hepcidin expression is decreased, ferroportin is able to function normally and iron is released from the intestinal cells and macrophages (Siah *et al*. 2006).

1.3.5.3 Post-translational control

Post-translational regulation involves iron regulatory proteins (IRPs) and iron responsive elements (IREs). IRP1 and IRP2 are responsible for the post-translational control of iron homeostasis by binding to IREs (Hentze and Kuhn 1996). IREs are located in the 5' and 3' untranslated regions (UTRs) of mRNA that encodes proteins involved the regulation of iron homeostasis (Andrews 2008). IRPs that bind to IREs in the 5' UTRs of mRNA have been shown to block translation, and IRPs binding to IREs in the 3' UTRs stabilise the mRNA and therefore increase translation (Ganz and Nemeth 2006). Proteins that are under the control of IREs include those involved in iron storage, iron export, iron uptake and haem synthesis.

The regulation of the binding of IRPs to IREs can be attributed to a number of factors. For example, in cases of iron overload, an iron sulphur cluster assembles in IRP1 inhibiting binding to IREs (Pantopoulos 2005). With regard to IRP2, proteosomes bind to a specific iron-dependant domain in the presence of iron and degrade it, therefore preventing any interaction with IREs.

1.4 GENES INVOLVED IN IRON HOMEOSTASIS

There have been several genes implicated in the homeostasis of iron in the body. The majority code for the proteins that play a role in the storage and transport of iron. A summary of the genes discussed in previous sections is illustrated in Table 1.1. below.

Table 1.1 Genes involved in iron metabolism.

Gene	Protein	Chromosomal location	Protein function
CP	Ceruloplasmin (CP)	3q21-24	Serum ferroxidase
CYBRD1	Cytochrome b reductase 1 (CYBRD1)	2q31	Reduction of Fe^{3+} to Fe^{2+} in gut lumen
DMT1	Divalent metal transporter 1 (DMT1)	12q13	Transports iron from gut lumen into enterocyte; from endosome to cytoplasm
Ferritin	H-Ferritin (FTH)	11q13	Oxidation of Fe^{2+} to Fe^{3+} to be stored in ferritin
	L-Ferritin (FTL)	19q13.3	Sequesters iron in the nuclear core of the ferritin protein
HAMP	Hepcidin	19q13	Systemic iron regulation
HCP1	Haem carrier protein 1	17q11.2	Transport of haem across apical membrane of enterocyte
HEPH	Hephaestin (HEPH)	Xq11-12	Membrane-bound ferroxidase
HFE	High iron protein (HFE)	6p21	Binds to TFR, reducing affinity for transferrin therefore reducing iron uptake
HJV	Haemojuvelin (HJV)	1q21	Modulator of hepcidin expression
HMOX1	Haem oxygenase 1 (HMOX1)	22q12	Oxidation of haem to Fe^{2+}
IRP1	Iron regulatory protein 1 (IRP1)	9p13-22	Bind to IREs of mRNA to post-translationally control protein expression
IRP2	Iron regulatory protein 2 (IRP2)	15q25.1	Bind to IREs of mRNA to post-translationally control protein expression
SLC40A1	Solute carrier family 40 (iron regulated transporter) member 1 (SLC40A1)	2	Transports iron at basolateral membrane of enterocyte
TF	Transferrin (Tf)	3q21	Iron transport protein in the plasma and extracellular fluid
TFR1	Transferrin receptor 1 (TFR1)	3q29	Iron uptake molecule of transferrin-bound iron into cells
TFR2	Transferrin receptor 2 (TFR2)	7q22	Exact function unknown. Homolog of TFR1

Adapted from Brittenham *et al.* 2000.

The association of the gene ceruloplasmin with OC has been investigated in this study and therefore only this gene (in relation to other iron genes) will be discussed in more detail.

1.4.1 Ceruloplasmin

1.4.1.1 Identification and mapping

Larsen (1977) demonstrated that ceruloplasmin (*CP*) and the transferrin (*TF*) gene were linked in cattle (LOD score of 11.3). Weitkamp (1983) found a peak LOD score of 3.5 for linkage of *CP* to TF in human families with a recombination frequency of about 10-15%. These findings were conclusive evidence that *CP* and *TF* were linked. Once *TF* was mapped to chromosome 3q21-25, a linkage group consisting of *CP* was mapped to the same region. This was achieved by Southern blot analysis of human-mouse somatic cell hybrids and *in situ* hybridisation (Naylor *et al*. 1985, Yang *et al*. 1986). In these studies, other sites of hybridisation on different chromosomal sites suggest that there may be *CP*-like DNA sequences in other areas of the human genome.

1.4.1.2 Structure of *CP* gene and protein

The *CP* gene consists of 19 exons and spans approximately 50 kb (Daimon *et al*. 1995). *CP* codes for the ferroxidase protein, ceruloplasmin (CP). CP is comprised of a single polypeptide chain of 1046 amino acids, with a molecular mass of 132 kDa (Takahashi *et al.* 1984). CP is a blue alpha-2-glycoprotein that binds 90-95% of plasma copper and has 6 to 7 copper ions per molecule.

1.4.1.3 Function

In most organisms, ferroxidases are the proteins that are essential for maintaining iron homeostasis (Sargent *et al.* 2005). CP has been shown to participate in iron homeostasis in swine afflicted with anaemia, as the condition is reversed in these animals following injection with ceruloplasmin (Ragan *et al*. 1969). CP is located primarily in the plasma and is responsible for driving iron transport from stores in various tissues (Cherukuri *et al.* 2005).

The CP enzyme is synthesized in the liver and it catalyzes the oxidation of ferrous iron (Fe^{2+}) to ferric iron (Fe^{3+}), thereby creating an ion gradient favouring iron export from the cells (Sargent *et al.* 2005). Transferrin (TF) is the main iron transport molecule in the plasma, but it can only carry iron in the ferric state, therefore CP acts as an aid in iron transport in the plasma. Enzymatic oxidation of ferrous iron by CP has been hypothesized as an important step in the formation of TF (Osaki *et al*. 1966), *via* the incorporation of iron into apo-transferrin. CP can also facilitate iron release from macrophages. CP is highly expressed in cells of the liver, brain and other tissues. It is not expressed in cells of the small intestine and therefore does not play an important role in iron release from intestinal enterocytes. This has been demonstrated in patients with aceruloplasminemia, and in ceruloplasmin null mutant mice that develop iron accumulation mainly in the liver and pancreas, but show no signs of abnormal intestinal iron absorption (Harris *et al*. 1999). These findings demonstrate an important physiologic role for CP in determining the rate of iron efflux from the cells in which iron is stored.

1.4.1.4 Ceruloplasmin and disease

(i) Wilson-Menkes disease and aceruloplasminaemia

CP was originally thought to be the gene responsible for Wilson-Menkes disorder, a rare disease characterized by accumulation of copper in the tissues and defective loading of CP with copper ions (Danks 1989). Harris *et al*. (1995) investigated a number of potential Wilson disease patients, a few of whom later turned out not to have the disease, but did display the characteristic molecular degeneration and low levels of serum CP. They identified a patient with a 5 bp insertion in exon 7 of *CP*, which resulted in a frame-shift mutation and truncated open reading frame. The mutation was confirmed in the heterozygous state in the patient's daughter, leading to the identification of aceruloplasminaemia.

Aceruloplasminaemia is an autosomal recessive disorder characterized by progressive neurodegeneration of the retina and basal ganglia and diabetes mellitus. Due to mutations in *CP*, iron accumulates in the pancreas, liver and brain. Accumulation in the eye may lead to retinal degeneration.

Wilson-Menkes disease, on the other hand, has subsequently been found to be caused by mutations in the *ATP7B* gene on chromosome 13, which is a putative copper-transporting P-type ATPase (Ala *et al.* 2007).

(ii) Parkinson's disease

Parkinson's disease (PD) is a neurological condition, characterized by the presence of Lewy bodies and regional areas of neurodegeneration. In the majority of PD patients, ferric iron is found to be accumulated in the brain (Sofic *et al.* 1988). The transferrin/iron ratio is also decreased in PD patients, which suggests that there is a problem with the movement of iron (Loeffler *et al.* 1995). It is still not clearly understood what causes the ultimate neuronal death in PD patients, although new evidence suggests that increased oxidative stress may play a role. Iron may contribute to this by increasing the formation of free radicals within cells of the central nervous system (CNS).

CP is a highly effective antioxidant because of its ability to oxidise highly toxic ferrous iron to the nontoxic ferric form, which helps to prevent damage to proteins, lipids and DNA (Gutteridge 1992). CP is expressed by astrocyte cells of the brain and cerebellum (Patel *et al.* 2000) in an alternatively spliced glycosylphosphatidylinositol (GPI)-anchored form, and not in the secreted form produced by the hepatocytes of the liver (Patel and David 1997). This GPI-anchored form is thought to play a role in iron homeostasis and antioxidant activity within the CNS. The importance of *CP* is illustrated in patients with aceruloplasminemia who have massive iron deposits in the brain and liver which lead to neurodegeneration and motor incoordination. In a study by Patel *et al.* (2002), it was shown that *Cp* knockout mice had an accumulation of iron as well as increased free radical injury to the CNS. Levels of CP have been reported to be reduced in the cortex of Alzheimer's patients (Connor *et al.* 1993) and the ferroxidase activity of CP is reduced in the cerebrospinal fluid in PD (Boll *et al.* 1999). It is therefore a possibility that disruptions to *CP*, either at the gene or the protein level, may contribute towards the process of neurodegeneration by increasing the levels of ferrous iron and subsequently the production of damaging free radicals.

(iii) Cancer

CP acts as an acute phase protein which is up-regulated in response to injury and inflammation. In this manner, it contributes in a major way to the antioxidant defence system of human plasma. It does so by a variety of mechanisms which include the inhibition of iron-dependent lipid peroxidation and OH formation from H_2O_2 *via* its ferroxidase activity, scavenging H_2O_2 and O_2^- and the inhibition of copper-induced lipid peroxidation by its ability to bind copper ions (Mukhopadhyay *et al.* 1998).

Various studies have demonstrated that levels of CP are raised in patients with cancer (Boz *et al.* 2005, Doustjalali *et al.* 2006). An example of this was demonstrated in a study by Pousset *et al.* (2001) using transgenic mice that developed hepatocellular carcinomas. These mice were found to have extremely elevated levels of serum CP which could not be attributed exclusively to inflammation when compared to other studies. The mechanism underlying this is poorly understood, although it is thought that an increase in CP levels may be explained by a protective response to an increase in circulating Fe^{2+} unbound to transferrin. It may also be an attempt by liver hepatocytes to control excessive tumour proliferation by reducing the free iron available to the tumour cells (Hann *et al.* 1992).

To date no conclusive evidence exists that links CP to the development of specific cancers. However, from the evidence presented above, it is becoming increasingly clear that CP plays an important role in the pathogenesis of cancer, particularly in its action as a potent antioxidant in response to infection and inflammation. Disruptions to the *CP* gene and/or protein may therefore have important clinical consequences in the aetiology of this complex disease. Further studies to elucidate the relationship between *CP* and cancer are therefore warranted.

1.5 OBJECTIVES OF THIS STUDY

In order to gain a greater understanding of the pathogenesis of oesophageal cancer, the identification of modifier genes responsible for OC susceptibility is of great importance. The information that can be gathered from the identification of these potential modifier genes can lead to improvements in early detection of OC, which in turn may lead to advancements in the treatment and counselling to individuals with OC.

The aims and objectives of this study are as follows:

1) The mutation analysis of the gene Ceruloplasmin (*CP*), involved in iron metabolism, in patients with Oesophageal Cancer (OC) by:
    - PCR amplification of the promoter and coding regions of this gene.
    - Heteroduplex single-strand conformation polymorphism (HEX-SSCP) analysis, restriction fragment length polymorphism (RFLP) analysis and semi-automated DNA sequencing analysis of the amplified regions to identify any novel and/or previously described variants in this gene.

2) Statistical analysis of the variants identified to determine statistically significant associations that may exist between specific variants, or groups of variants, and OC pathogenesis and/or susceptibility.

3) To determine if the effects of these variants in this gene are associated with its role in iron metabolism and possibly disease expression by *in silico* analysis.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

This project has gained ethical approval from the Research and Ethics/Biosafety Committee, Faculty of Health Sciences, University of Stellenbosch, with reference number N07/06/147.

2.1.1 Study cohort

The study cohort for this study included 184 (96 patients and 88 controls) individuals from the Black Xhosa-speaking population of South Africa. All individuals are of Central African descent and observe cultural traditions originating from Xhosa tribes of South Africa.

Patient samples were recruited from collaborating clinicians, who were responsible for clinical assessment. Patients with OC present with symptoms such as dysphagia (difficulty to swallow), loss of weight, nausea and vomiting, coughing and hematemasis (vomiting of blood). These patients were referred to the Provincial Hospital, Port Elizabeth for a barium swallow followed by biopsies in the theatre to confirm the presence of squamous cell carcinoma (SCC) or adenocarcinoma (ADC) on the basis of the histology. All of the patients in this study presented with SCC.

The control individuals in this study consist of 88 unrelated, healthy individuals from the same population. These individuals were interviewed for a potential family history of relevant diseases and early screening for OC was performed using a Nabeya capsule for brush biopsy. Recruitment of the control individuals took place at the Mbekweni and Mpeko Clinics, Transkei.

A whole blood sample was obtained with written informed consent from all of the study participants.

2.1.2 Patient demographics

The patient cohort used in this study consisted of 96 Black South African individuals from the Transkei region of South Africa. The patient cohort included 48 (50%) males and 48 (50%) females, with a mean age of 59 years [standard deviation (SD) ± 13 years] and 66 years [standard deviation (SD) ± 12 years] respectively (Table 2.1). Demographic information on alcohol consumption and the smoking status of the patients is summarised in Table 2.1. Smoking status was restricted to cigarette smoking (including shop-bought and home-grown cigarettes). Alcohol consumption was restricted to beer (including shop-bought and home-made beer).

Table 2.1 Patient demographics.

Demographic	OC Patients	
	Males (n=48)	**Females (n=48)**
Average Age	59±13	66±12
Smoking		
Yes	9 (19%)	4 (8%)
Stopped	27 (56%)	18 (36%)
Never	4 (8%)	13 (27%)
Unknown	8 (17%)	13 (37%)
Alcohol consumption		
Yes	31 (65%)	25 (52%)
Stopped	2 (4%)	0
Never	6 (13%)	10 (21%)
Unknown	9 (19%)	13 (27%)

2.1.3 Body iron status

In the current study, the transferrin saturation (TS) and the serum ferritin (SF) levels were measured in the patient and control cohorts according to standard methodology. Iron overload parameters were defined as the percentage transferrin saturation (%TS) > 45 and/or serum ferritin exceeding 200 μg/l in females and 300 μg/l in males (Looker and Johnson 1988, Adams and Chakrabarti 1998).

The iron status of each individual was classified into one of four groups (see Table 2.2): iron deficiency (SF < 20 μg/l), normal serum ferritin levels (females: SF between 20 and 200 μg/l and males: SF between 20 and 300 μg/l), raised serum ferritin levels (females: SF > 200 μg/l and males: SF > 300 μg/l) with %TS < 45 and raised serum ferritin levels with %TS > 45.

Table 2.2 Iron status of the OC patients included in this study.

	Normal iron levels (Male: SF<300μg/l; Female: SF<200μg/l)	**Iron deficient (SF<20μg/l)**	**Increased iron levels (Male: SF>300μg/l; Female: SF>200μg/l)**	
			%TS<45	**%TS>45**
Male (n=48)	31 (64.5%)	0	17 (35.4%)	0
Female (n=48)	25 (52.08%)	3 (6.25%)	19 (39.5%)	1 (2.08%)

2.2 DETAILED EXPERIMENTAL PROCEDURES

Refer to Appendix 1 for a list of all chemicals/reagents and their respective suppliers used in this study.

2.2.1 Total genomic DNA isolated from whole blood

Total genomic DNA (gDNA) of the patient and control groups used in this study was available from a previous study (Human 2007). gDNA extractions were performed using whole blood samples collected in tubes containing ethylenediamine tetraacetic acid (EDTA) as the preservative, using a modified protocol of the salting out technique as described by Miller *et al.* (1988).

2.2.2 Polymerase chain reaction (PCR) amplification

2.2.2.1 Oligonucleotide primers

The oligonucleotide primers used in this study were designed using the published *CP* reference sequence (Ensembl reference number ENSG00000047457 for the promoter region and GenBank reference number NM_000096 for the coding region) with the Primer3 v0.2 program (Rozen and Skaletsky 2000). The reference sequence of the *CP* promoter and coding region, with the relevant positions of the primers designed, is illustrated in Appendix 2. The DNA samples were amplified using overlapping primers designed for the promoter region of the gene (Table 2.3) and intronic primers designed for the coding region of the gene (Table 2.4)

Table 2.3 Oligonucleotide primers used for PCR amplification of the *CP* gene promoter region.

Primer Name	Forward Primer (5'-3')	T_m (˚C)	Reverse Primer (5'-3')	T_m (˚C)	Product Size (bp)	$T1_A$ (˚C)	$T2_A$ (˚C)	PCR Cycle
CPP1[a]	CCTGTTAGGCTCTGCTAGTT	60	CTTATGGGACCACCATCACA	60	267	56		A
CPP2[a]	AGAAATAGTCATGCACCAC	54	TGCATTGTTAGGCTATTTTG	54	311	60	55	B
CPP3[a]	CAGGTTTGTAGCCTAGGAGC	62	ATGCTCCCTTTGTTCCTCTG	60	282	56		A
CPP4[a]	CTCAGAACGTATCCCTGTCAT	62	TTGCCAGGCTTCTCTGACTG	62	275	57		A
CPP5[a]	CAGAGGAACAAAGGGAGCAT	60	GGAGCCTGAGAAGAAATGAAG	62	318	57		A
CPP6[a]	CAGTCAGAGAAGCCTGGC	58	GTGACTTACGTGTCAACAG	56	466	56		A

Abbreviations: 5', 5-prime; 3', 3-prime; ºC, degrees Celsius; bp, base pair; *CP*, Ceruloplasmin gene; CPP1-6, Ceruloplasmin promoter fragments 1 to 6; PCR, polymerase chain reaction; $T1_A$, annealing temperature 1; $T2_A$, annealing temperature 2; T_m, melting temperature [$T_m = 2(nA+nT)+4(nG+nC)$]. References: [a]M Hallendorff (2008).

Table 2.4 Oligonucleotide primers used for PCR amplification of the *CP* gene coding region.

Exon	Primer Name	Forward Primer (5'-3')	T_m (˚C)	Reverse Primer (5'-3')	T_m (˚C)	Product Size (bp)	T_A (˚C)	PCR Cycle
1	CP1[b]	AGGCTCCAAGAAGGGGAAA	58	CAAGCCCACATTTTGCAG	54	288	62	A
2	CP2A[b]	TGGAGGCATCCCTACAACAG	62	TAGGGCCTAAAAACCCAAGC	60	280	62	A
	CP2B[b]	GGCCCAGATAGAATTGGGAGA	64	AGCTAAAAGGCACTTCTATCG	60	281	60	A
3	CP3A[b]	AACACATCCCCAAGGATCAC	60	CACAATTGCCATCTCCTTCC	60	257	61	A
	CP3B[b]	CTTCCTAGGGGCCATCTACC	64	GCCCCTGTCTTTTGGTCATA	60	269	60	A
4	CP4[b]	CATGCTAAAAGTTTAGTCTTG	56	ATGTGAGGGAATAAGCTTAGC	60	297	61	A
5	CP5A[b]	GAGGTAACTCTCACTATCTC	58	TTGTGTCAATACGGTAGTTC	56	224	61	A
	CP5B[b]	CTCCCAGGACTCTCCATGGTG	68	CACCTTTTTCAGCTGACTGC	60	273	61	A
6	CP6[b]	ACCCGAGCAGTGTTTACAGG	62	GTGCGGGGGAGAGCATATT	60	300	60	A
7	CP7[b]	GTTCTACTTCAACCCCAGCA	60	CCCATGGGAAGAGTAAACCA	60	230	62	A
8	CP8[b]	GCATGTTACATACCAAGGAAGG	64	CCTGCCTTCAGTTTTTGCAG	60	290	60	A
9	CP9[b]	GGCTCAAATGACCACGTTAG	60	TTTTTCCCCAGTTGGACTTAC	60	291	61	A
10	CP10[b]	TGCACATGGAAGTCTTCTGC	60	GGTAGATTGGTGGATGATGC	60	242	60	A
11	CP11A[b]	GGGCAAGCCTGAATAGGTCT	62	GCTGTGTCTCTCCATTCTCC	64	281	60	A
	CP11B[b]	ATCAGCCGGGTCTCACTATG	62	GCTTGGGGAAGGGATAAGTT	60	294	60	A
12	CP12[b]	CAGAGGGCACTAGCAAAGAAG	64	TTGGGAATCAGAGTCTGGAG	60	283	60	A
13	CP13[b]	CATTAAGACAAAACTAACC	50	CCCACACCTGATAAACTGGAG	64	214	56	A
14	CP14[b]	CCATGGGGGAGAGAAACAAG	62	CAGCCTGTTAAAATGCACCA	58	270	61	A
15	CP15[b]	GTTCCTGGGTCCTAGTTATC	60	GGGAAGTGTCACAAAACAAATG	62	255	62	A
16	CP16A[b]	CCTGGGTGCAAAGTCTCAGT	62	CTTTCTCGGGGTGATCAGAG	62	249	62	A
	CP16B[b]	CAATCCCAGAAGGAAACTGG	60	CAAAGTGAGGCAGAAGTGGT	60	276	61	A
17	CP17[b]	ATCCTGAAAAGTAACATAAAACC	60	GCACCGGCTGTACTCTTTG	60	295	61	A
18	CP18[b]	TTCAATATCCCTGAGCTGAACA	62	TGGCTTCTAGAATTACTACCTGGA	68	283	61	A
19	CP19[b]	ATGAATCAGGAGTAAAGTAAC	56	TTCCCCCACAAATGTACAAAG	60	279	61	A

Abbreviations: 5', 5-prime; 3', 3-prime; ºC, degrees Celsius; bp, base pair; *CP*, Ceruloplasmin gene; CP1-19, Ceruloplasmin coding fragments 1 to 19; PCR, polymerase chain reaction; $T1_A$, annealing temperature 1; T_m, melting temperature [$T_m = 2(nA+nT)+4(nG+nC)$]. References: [b]This study.

All primers were manufactured by Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, South Africa) and Integrated DNA Technologies (IDT). The melting temperature (T_m) was calculated for each of the primers using the following equation as described by Thein and Wallace (1986): $T_m = 2(nA + nT) + 4(nG + nC)$.

2.2.2.2 PCR reaction and DNA amplification

PCR conditions were optimized for each specific primer pair and are indicated in Tables 2.3 and 2.4. PCR amplification of the promoter and coding region was performed in a GeneAmp® 2700 PCR System (Perkin Elmer Applied Biosystems, Warrington, WA, Great Britain) in reactions with a total volume of 25 µl, consisting of 0.5 U *Taq* polymerase enzyme (BIOTAQ™ DNA polymerase, Bioline Ltd., London, England), 50 ng DNA template, 0.2 mM of each dNTP (dATP, dCTP, dTTP, dGTP) (Fermentas), 10 pmol of each primer, 1 × Buffer (Bioline) and 1.5 mM of magnesium chloride ($MgCl_2$). For every round of PCR amplification performed, a negative control (reaction without gDNA template) was included to indicate the possibility of contamination in the PCR reactions.

Two different PCR cycle programs, namely A and B, were utilised for the amplification of the gene (see Tables 2.3 and 2.4) and are described below:

Cycle A

An initial denaturation step at 95˚C for 5 minutes, followed by 35 cycles of denaturation at 95˚C for 30 seconds, annealing at a temperature optimized for each primer set (indicated as the $T1_A$ of each amplicon in Table 2.3 and as the T_A in Table 2.4) for 45 seconds and an

elongation step of 72˚C for 30 seconds. This was followed by a final extension step of 72˚C for 10 minutes.

Cycle B

An initial denaturation step at 95˚C for 5 minutes, followed by 10 cycles of denaturation at 95˚C for 30 seconds, annealing at a temperature optimized for the primer set (indicated as the $T1_A$ of the amplicon in Table 2.3) for 45 seconds and an elongation step of 72˚C for 30 seconds. This was then followed by 30 cycles of the same conditions but with an annealing temperature as specified in Table 2.3 as $T2_A$, followed by a final extension step of 72˚C for 10 minutes.

2.2.3 Agarose gel electrophoresis

Agarose gel electrophoresis was used to test for successful amplification of the PCR products. Following PCR amplification, each sample was electrophoresed on a 2% (w/v) horizontal agarose gel [consisting of 4g agarose in 200 ml 1 × Tris-Borate-EDTA (TBE) (90 mM Tris-HCl, 90 mM boric acid (H_3BO_3) and 0.1 mM EDTA, pH 8.0) and 0.01% (v/v) ethidium bromide (EtBr)].

The PCR product (5µl) was mixed with an equal volume of cresol red loading buffer [2 mg/ml cresol red and 35% (w/v) sucrose] and loaded into the wells of the gel. The size of each specific fragment was compared against an appropriate molecular size marker [100 bp ladder (Generuler™, Fermentas)] loaded alongside the PCR samples. Resolution of the PCR products was performed at 120 V for 1 hour in 1 × TBE buffer solution. The DNA was

visualised by ultraviolet light transillumination and captured using the MultiGenius Bio Imaging System (Syngene, Cambridge, England).

2.2.4 Heteroduplex single-strand conformation polymorphism (HEX-SSCP) analysis

HEX-SSCP analysis (Kotze *et al.* 1995) was performed on a 16 cm vertical gel apparatus using the Hoefer SE 600 Basic slab gel apparatus. The PCR products of the promoter amplicons and the various exons were resolved on a 12% polyacrylamide (PAA) gel supplemented with urea [$(NH_2)_2CO$] [gel consisting of 7.5% (w/v) $(NH_2)_2CO$, 1.5 × TBE (135 mM Tris-HCl, 135 mM H_3BO_3 and 2 mM EDTA, pH 8.0), 12% (w/v) PAA [of a 40% stock solution with 1%C [99 acrylamide (AA): 1 bisacrylamide (BAA)], 0.1% (w/v) ammonium persulphate (APS) and 0.1% (v/v) TEMED].

A 0.75 mm gel was cast and allowed to polymerise. Gels were placed in a Hoefer SE 660 electrophoresis tank filled with fresh 1 × TBE buffer. The upper buffer chamber was filled with 1.5 × TBE buffer. Prior to loading, 15 µl of bromophenol blue ($C_{19}H_{10}Br_4O_5S$) loading buffer [consisting of 95% formamide (H_2NCHO) (v/v), 20 mM EDTA, 0.05% (w/v) xylene cyanol ($C_{13}H_{28}N_2Na_4O_{13}S$) and 0.05% (w/v) bromophenol blue] was added to 20 µl of each of the PCR products; heat denatured for ten minutes at 95ºC and then immediately cooled to 4˚C by placing on ice. Approximately 15 µl of the denatured PCR product was loaded onto the gel using a Hamilton syringe. Electrophoresis was performed at 250 V (4ºC) for 18 hours. Following electrophoresis, visualisation of the DNA fragments on the PAA gels was achieved by placing them in an EtBr solution [0.01% (v/v)] for 10 minutes, followed by 3 minutes of destaining in ddH_2O. The DNA fragments were then subjected to ultraviolet light transillumination and photographed with the Multigenius Bio Imaging System (Syngene, Cambridge, UK).

2.2.5 Restriction fragment length polymorphism (RFLP) analysis

In order to be able to distinguish clearly between the different genotypes of the T83 polymorphism in exon 2 (amplicon 2A, see Table 2.4) of the *CP* gene, PCR products were subjected to RFLP analysis using the *SfcI* enzyme (recognition site: 5' C↓TRYAG 3'; New England Biolabs Inc., Beverly, USA). This variant abolishes the *SfcI* restriction site, and therefore generates a single 280 bp product when present in the homozygous state following digestion. The wild-type homozygous state produces two fragments of 242 bp and 38 bp as the restriction site is uninterrupted in these samples, therefore allowing digestion to take place. The heterozygous state of the variant produces three fragments of 280 bp, 242 bp and 38 bp.

The PCR products together with the *SfcI* enzyme were digested overnight in 20 µl reactions consisting of 10 µl PCR product, 2 U of the enzyme and 1 × buffer, in a 37˚C water bath.

Electrophoresis of the digested PCR products was performed on a 3% (w/v) horizontal agarose gel [consisting of 6g agarose in 200 ml 1 × TBE and 0.01% (v/v) EtBr]. The digested product (10 µl) was mixed with 5 µl cresol red loading buffer and loaded onto the gel. A molecular size marker [100 bp ladder (Generuler™, Fermentas)] was loaded alongside the samples in order to determine the correct size of the digested fragments. Resolution of the digested products was performed in 1 × TBE running buffer at room temperature at 80 V for the first hour, followed by 100 V for a further 2.5 hours. The DNA was visualised by UV light transillumination and captured as described in section 2.2.3.

2.2.6 Semi-automated DNA sequencing analysis

2.2.6.1 DNA Purification

Purification of the PCR products showing sequence variation upon HEX-SSCP analysis was performed using the Sigma GenElute™ PCR Clean-Up Kit (Sigma kit constituents not available). Briefly, a GenElute Miniprep Binding Column was placed within a collection tube and 500 µl of Column Preparation Solution was added. After a 1 minute centrifugation step at 12000 × g (Centrifuge 5415D, Eppendorf), the eluate was discarded. A mixture of 500 µl Binding Solution and 100 µl PCR product was then transferred to the column and centrifuged at 15000 × g for the duration of 1 minute. The flow-through was discarded and 500 µl of Wash Solution added to the column, followed by further centrifugation at 15000 × g for 1 minute. The flow-through was once again discarded followed by a further centrifugation step at 15000 × g for 2 minutes. The flow-through and collection tube were subsequently discarded and a new collection tube used with the column. After the addition of 10 µl sterile SABAX water to the column, it was centrifuged at 15000 × g for 1 minute and the subsequent eluate contained the purified PCR product. The DNA was stored at 4˚C.

2.2.6.2 Cycle sequencing reaction and electrophoresis

The cycle sequencing reaction consisted of a total volume of 7 µl, made up of 10 ng purified PCR product, 3.3 pmol primer and 1 µl of the BigDye® Terminator v3.1 Cycle Sequencing mix (Applied Biosystems, Warrington, WA, Great Britain). The same primers used for the PCR amplification were used (see Tables 2.3 and 2.4).The reaction was carried out using the GeneAmp® PCR System 2700 (Applied Biosystems). The following cycle sequencing

program was used: 25 cycles of denaturation at 96˚C for 10 seconds, annealing at 50˚C for 5 seconds and an extension step at 60˚C for 4 minutes. Samples were subsequently sent to an analytical facility for post-cycle sequencing clean-up and electrophoresis of the sample on an ABI Prism 3130XL Genetic Analyser (Applied Biosystems).

Analysis of the DNA sequences and chromatograms was performed by alignment with the published reference (wild-type) sequence (accession numbers listed in Section 2.2.2.1 and Appendix 2) using the BioEdit Sequence Alignment Editor v7.0.9.0 (Hall 1999), as well as visual inspection of the electropherograms to detect differences from the reference sequence.

2.3 STATISTICAL ANALYSIS

The allele- and genotype frequencies of the variants detected were determined in both the patient and control groups by allele counting. The total carrier frequency, including heterozygotes and homozygotes, was counted to estimate gene variants. Statistical differences between the patient and control groups were tested for by chi-squared (χ^2) analysis. Probability (*P*) values smaller than 0.05 ($P < 0.05$) were regarded as statistically significant. These analyses were conducted using Microsoft Office 2007 software and all results were verified using STATISTICA [StatSoft, Inc. (2003) STAT (data analysis software system), version 6]. Departure from Hardy-Weinberg equilibrium (HWE) in the patient and control groups was also tested for in the case of each variant that was identified. This was completed using a chi-squared (χ^2) goodness-of-fit test using the Tools for population genetics association studies (TFPGA) program version 1.3 (Miller 1997).

Haplotype analysis and linkage disequilibrium (LD) analysis was performed on each of the variants identified (Haploview 4.0, Barrett *et al.* 2005). Default parameters were applied to test for LD [the coefficient of association (D'): where D'=1 for perfect linkage disequilibrium; logarithm of the likelihood odds ratio, measure of confidence (LOD): LOD>3; correlation coefficient between two loci (r^2): $r^2>0.8$ where $r^2 =1$ for prefect linkage disequilibrium]. The default block definition was applied (Gabriel *et al*. 2002) upon haplotype analysis.

2.4 BIOINFORMATIC ANALYSIS

In silico analysis was performed on each of the variants identified in the promotor and coding regions in this study. Several bioinformatic databases are available for *in silico* analysis of the promotor region, of which the following two were used in this study: JASPAR CORE (Sandelin *et al.* 2004) and TRANSFAC®7 (Wingender *et al.* 2001). From TRANSFAC®7, two programs were used, namely PATCH and MATCH™ (v1.0) (Kel *et al.* 2005). The default parameters were employed in the use of both of these databases. In order to compare the putative transcription factor binding sites (TFBS) identified by these databases, the reference sequence (Ensembl, www.ensembl.org) was compared with the variant sequence. Using this analytical method, the creation or abolishment of a particular TFBS due to the particular nucleotide variation detected in this study could be determined.

The ESEfinder (ESE-Exonic Splice Element) program was utilised for *in silico* analysis of the exonic variants identified in this study (Cartegni *et al.* 2003). ESEfinder was employed to determine the possible effect of the exonic variants on gene splicing. Putative ESEs in input sequences are searched for using weight matrices which correspond to the motifs of four different human serine/arginine-rich (SR) proteins.

CHAPTER 3

RESULTS AND DISCUSSION

Presented in the form of a full-length manuscript in preparation for future publication in a scientific journal

Molecular Genetic Analysis of *Ceruloplasmin* in Oesophageal Cancer

ABSTRACT

OC is a disease characterised by the development of malignant tumours in the epithelial cells lining the oesophagus. It demonstrates marked ethnic variation, with SCC being more prevalent in the Black population and ADC occurring more often in Caucasians. The aetiology of this complex disease has been attributed to a variety of factors, including an excess of iron (resulting in increased tumourigenesis), oesophageal injury and inflammation (due in part to Barrett's oesophagus and smoking among others). The aim of this study was to determine if genetic variations identified in the *CP* gene (implicated in iron homeostasis) contribute in any way to OC pathogenesis or susceptibility. The study cohort consisted of 96 unrelated OC patients from the Black Xhosa-speaking South African population and 88 population-matched control individuals. The promoter and coding regions of the *CP* gene were analysed for DNA sequence variation using heteroduplex single-strand conformation polymorphism (HEX-SSCP) analysis, restriction fragment length polymorphism (RFLP) analysis and semi-automated bidirectional DNA sequencing analysis. Fourteen previously described (5'UTR-567C→G, 5'UTR-563T→C, 5'UTR-439C→T, 5'UTR-364delT, 5'UTR-354T→C, 5'UTR-350C→T, 5'UTR-282A→G, V223, Y425, R367C, D544E, IVS4-14C→T, IVS7+9T→C and IVS15-12T→C) and four novel (5'UTR-308G→A, T83, V246A and G633) variants were identified. Statistical analysis revealed that two of the novel variants were significantly associated with OC in this study and could, therefore, potentially contribute to disease susceptibility. This is the first study to examine *CP* with respect to OC in the Black South African population. As such, these findings should serve to further our understanding of the relationship between iron metabolism and disease pathogenesis.

INTRODUCTION

OC is one of the leading causes of cancer-related deaths worldwide; it is the 15th most common cancer in developed nations and the 4th most common in developing countries such as South Africa (Crespi *et al*. 1994). The Transkei region of South Africa is thought to be the centre of the disease in Africa (Sammon 2007), with an age standardised incidence rate (ASIR) of 46.7/100 000 for males and 19.2/100 000 for females previously being reported (Makaula *et al*. 1996). OC shows clear geographic variation and occurs at a high incidence in certain areas of the world, which are termed "oesophageal cancer belts" (Blot 1994, Parkin *et al*. 2005). Two subtypes of OC exist, SCC and ADC, and both demonstrate marked racial variation which is characteristic of the disease. ADC is more prevalent in Caucasians and SCC occurs more frequently in Blacks. In South Africa, SCC is thought to be the leading cause of death among Black males and the 4th most common cause of death in Coloured males (Blot 1994). In addition, OC incidence rates also increase markedly with age, with the average age of disease diagnosis in South Africa being 60 years of age (Lagergren 2008).

The combined use of tobacco and alcohol is one of the most important risk factors in the aetiology of OC in Western countries (Parkin 2001). Other risk factors include oesophageal injury and inflammation, infectious agents, nutrition, nitrosamines, mycotoxins and exposure to environmental toxins (Marasas 1979, Syrjanen 1982, Blot 1994, Parent *et al.* 2000, Matsha *et al.* 2002, Goldblum 2003).

The relationship between iron and OC was previously described during a study of Black South African OC patients who were shown to have iron overload as a consequence of routinely drinking traditional beer brewed in non-galvanised steel drums (Bothwell *et al.* 1964, MacPhail *et al.* 1979, Isaacson *et al.* 1985). It was demonstrated that excess dietary iron can act as a potential risk factor for the development of OC, and further studies have

shown that iron is involved in the development of other cancers such as liver cancer (Mandishona *et al.* 1998). In addition, studies using rat models showed that iron supplementation could be indicated as a risk factor for developing OC (Goldstein *et al.* 1998, Chen *et al.* 1999, 2000). Studies have also shown that an increase in dietary iron plays a role in the development of OC in groups other than the Black South African population (Amer *et al.* 1990, Rogers *et al*. 1993). It has also been demonstrated that loss of iron from a target cell results in anti-tumour activity, which is reduced when there is excess iron present (Huot *et al.* 1990). For this reason, iron can be said to have an indirect carcinogenic effect on cells. Increased levels of iron can also inhibit the growth of tumouricidal-activated macrophages (Weiss *et al.* 1992).

In order to further investigate iron as a risk factor for OC, this study focused on the analysis of the regulatory and coding regions of the *CP* gene, which is involved in iron homeostasis, in patients diagnosed with OC. CP is a ferroxidase enzyme synthesized in the liver which catalyses the oxidation of ferrous iron (Fe^{2+}) to ferric iron (Fe^{3+}), thereby creating an ion gradient favouring iron export from the cells (Sargent *et al.* 2005). Therefore, CP which is located primarily in the plasma is responsible for driving iron transport from stores in various tissues (Cherukuri *et al*. 2005).

In this study, we aim to demonstrate the relationship between genes involved in iron metabolism (specifically *CP*) and the development of OC, by identifying gene variations that could potentially contribute toward iron dysregulation and subsequent disease pathogenesis. It is anticipated that the results obtained from this study will lead to a greater understanding of the role that iron homeostasis plays in the aetiology of OC.

MATERIALS AND METHODS

The detailed experimental procedures employed in this study, as well as information regarding the study participants and ethical approval for this study, are described in detail in the Materials and Methods section (see Chapter 2).

RESULTS

3.1 MUTATION ANALYSIS

The *CP* gene implicated in iron metabolism was screened in Black South African OC patients as well as population-matched control individuals. Mutation analysis of the entire *CP* gene revealed several variants (previously described and novel) following HEX-SSCP analysis. RFLP analysis was performed for specific variants where HEX-SSCP analysis proved inconclusive. All of the variants detected were verified by semi-automated bidirectional DNA sequencing analysis to confirm their presence. The positions of all the variants identified in this study are indicated on the *CP* reference sequence in Appendix 2.

The PCR reaction for amplicon 13 (see Table 2.4) of the *CP* coding region could not be optimised and therefore mutation analysis for this region is incomplete.

3.1.1 Promoter variants

Mutation analysis of the *CP* promoter region resulted in the detection of six previously described single nucleotide polymorphisms (5'UTR-567C→G, 5'UTR-563T→C, 5'UTR-439C→T, 5'UTR-354T→C, 5'UTR-350C→T and 5'UTR-282A→G), one previously described single base pair deletion (5'UTR-364delT) and one novel single nucleotide substitution (5'UTR-308G→A).

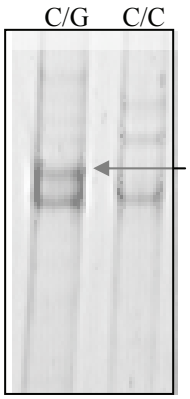
HEX-SSCP analysis of fragment 2 (see Table 2.3) of the promoter region revealed two previously described variants:

The first was a C to G substitution (5'UTR-567C→G) (RefSNP ID: rs34053109) detected 567 nucleotides upstream of the translation initiation site (ATG). This previously described variant was observed only in the heterozygous state and only in two of 86 (2.3%) of the population-matched controls and none of the OC patients (Figure 3.1).

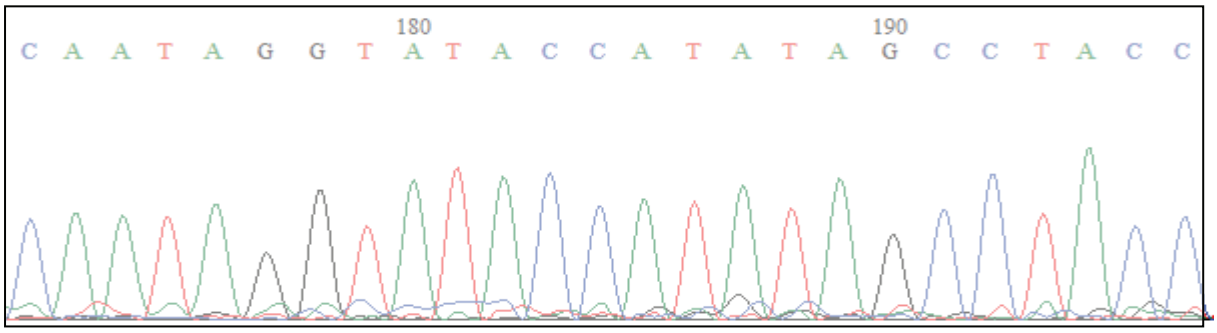
The second previously described variant in fragment 2 of the *CP* promoter results from a T to C transition 563 nucleotides upstream of the initiating ATG (5'UTR-563T→C) (RefSNP ID: rs17838834). It was detected in the heterozygous state in 47 of 92 (51.1%) of the patients and in 38 of 86 (44.2%) of the controls. In the homozygous state, the variant was detected in seven of 92 (7.6%) of the patients and three of 86 (3.5%) of the controls (Figure 3.2).

Figure 3.1 Schematic representation of the 5'UTR-567C→G variant in the *CP* promoter.

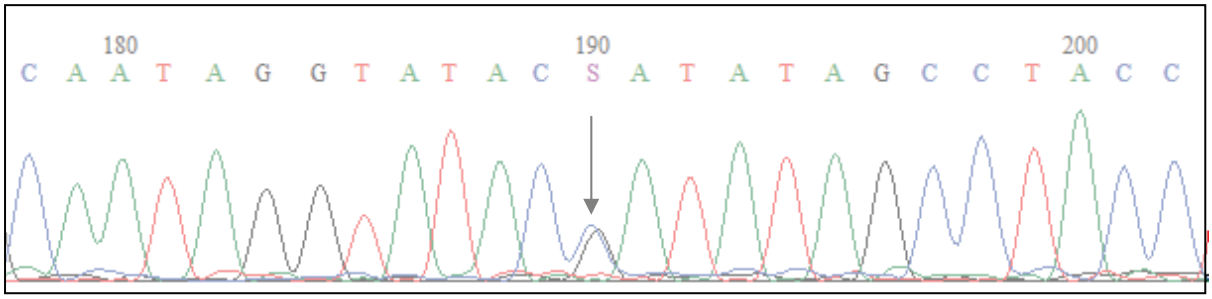
A.



B.(i)

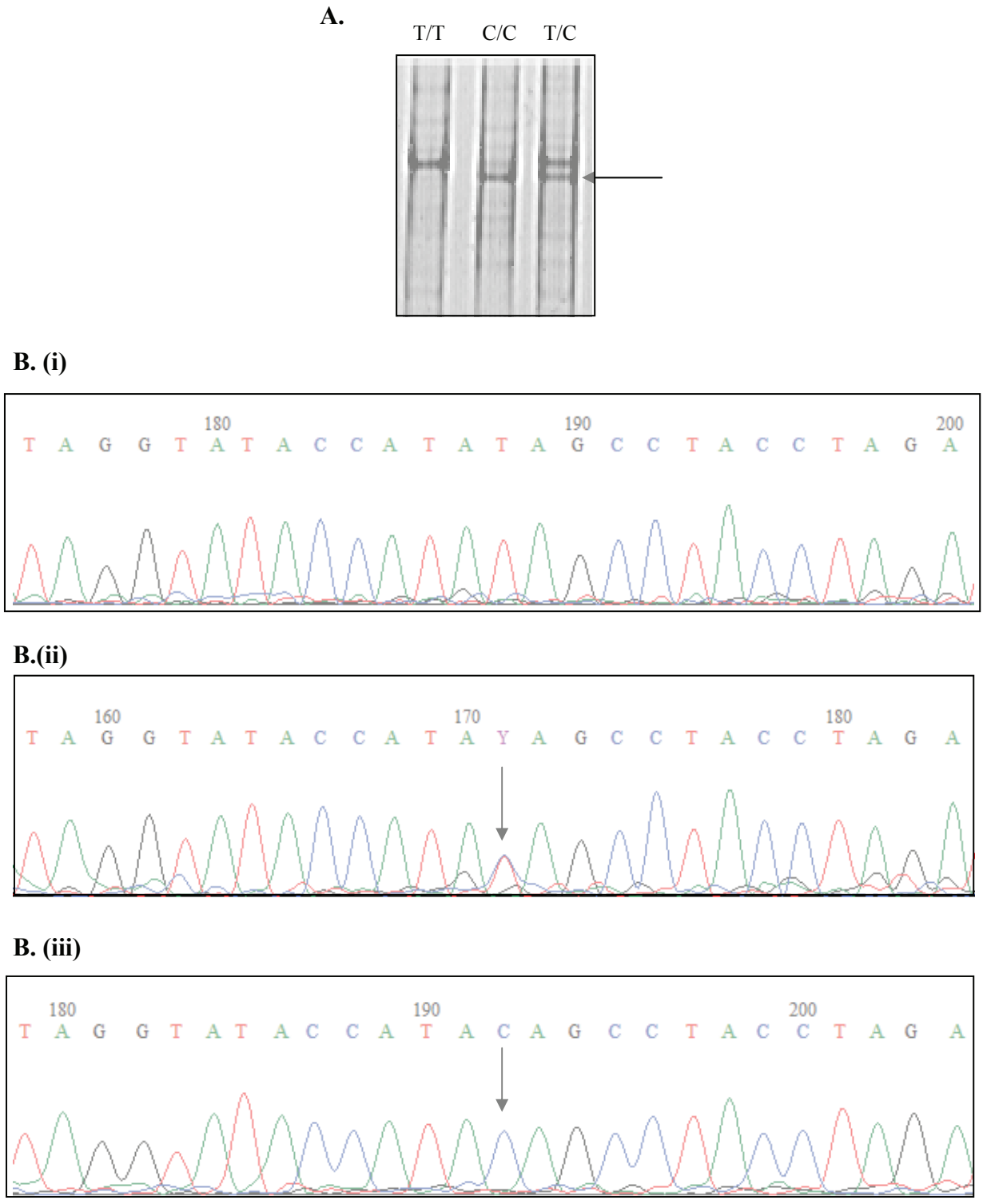


B.(ii)



Legend to Figure 3.1
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrow indicates aberrant band. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the 5'UTR-567C→G variant in the heterozygous state. Arrow indicates the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

Figure 3.2 Schematic representation of the 5'UTR-563T→C variant in the *CP* promoter.



Legend to Figure 3.2
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrow indicates aberrant band. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the 5'UTR-563T→C variant in the heterozygous state (iii) the 5'UTR-563T→C variant in the homozygous state. Arrows indicate the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

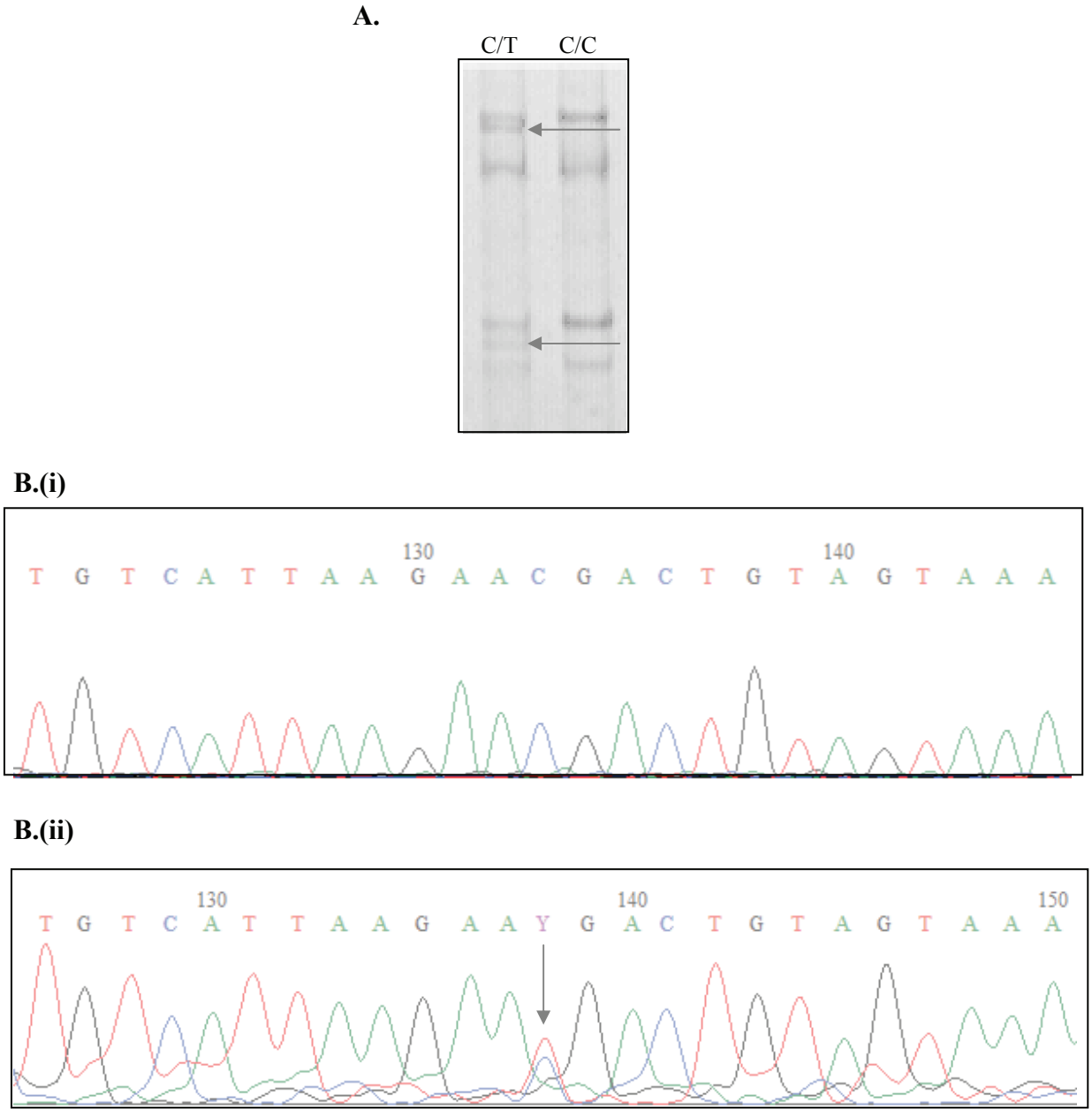
HEX-SSCP analysis of fragment 3 (see Table 2.3) of the *CP* promoter revealed four previously identified variants:

The first, a single base pair change at nucleotide position 5'UTR-439C→T was initially identified as a T to C transition. The common allele was initially identified upon inspection of a reference sequence (www.ensembl.org) and then cross referenced with another reference sequence from a different database (www.genatlas.org). However, in this study, it was found that the variant allele at position 5'UTR-439T→C had a higher incidence than the allele denoted on the reference sequence. After taking the ancestral allele from the HapMap project (The International HapMap Consortium, 2003) (http://www.hapmap.org) into consideration it was decided to annotate the variant using the common allele as the one with the higher frequency. This variant was subsequently annotated as 5'UTR-439C→T (RefSNP ID: rs701749). This variant was detected only in the heterozygous state, in one of 79 (1.3%) of the OC patients and three of 79 (3.8%) of the controls (Figure 3.3).

The second previously described variant in fragment 3 was 5'UTR-364delT (RefSNP ID: rs17838833), which involves a deletion of the T nucleotide 364 nucleotides upstream of the translation initiation site. It was identified only in the heterozygous state and was limited to two of 85 (2.4%) of the patients (Figure 3.4).

A previously described T to C substitution (5'UTR-354T→C) (RefSNP ID: rs17838832) was identified in the heterozygous state in 40 of 80 (50%) of the patients and 33 of 79 (41.8%) of the controls. The homozygous state was detected in six of 80 (7.5%) of the patients and in six of 79 (7.6%) of the controls (Figure 3.5).

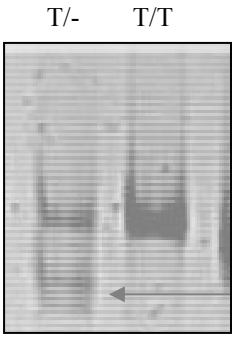
Figure 3.3 Schematic representation of the 5'UTR-439C→T variant in the *CP* promoter.



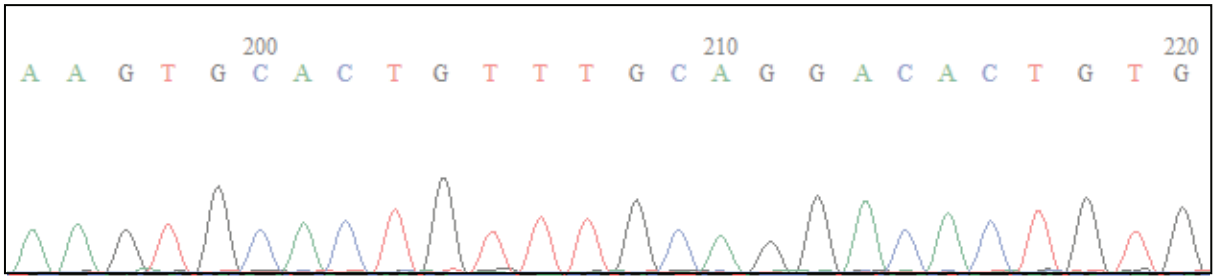
Legend to Figure 3.3
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrows indicate aberrant bands. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the 5'UTR-439C→T variant in the heterozygous state. Arrow indicates the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

Figure 3.4 Schematic representation of the 5'UTR-364delT variant in the *CP* promoter.

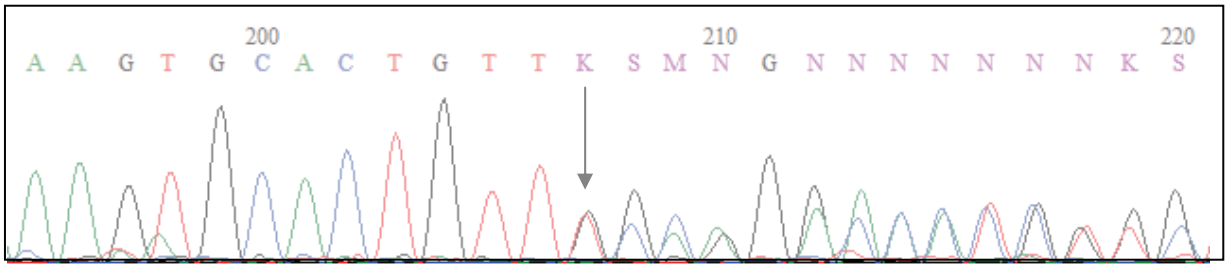
A.



B.(i)

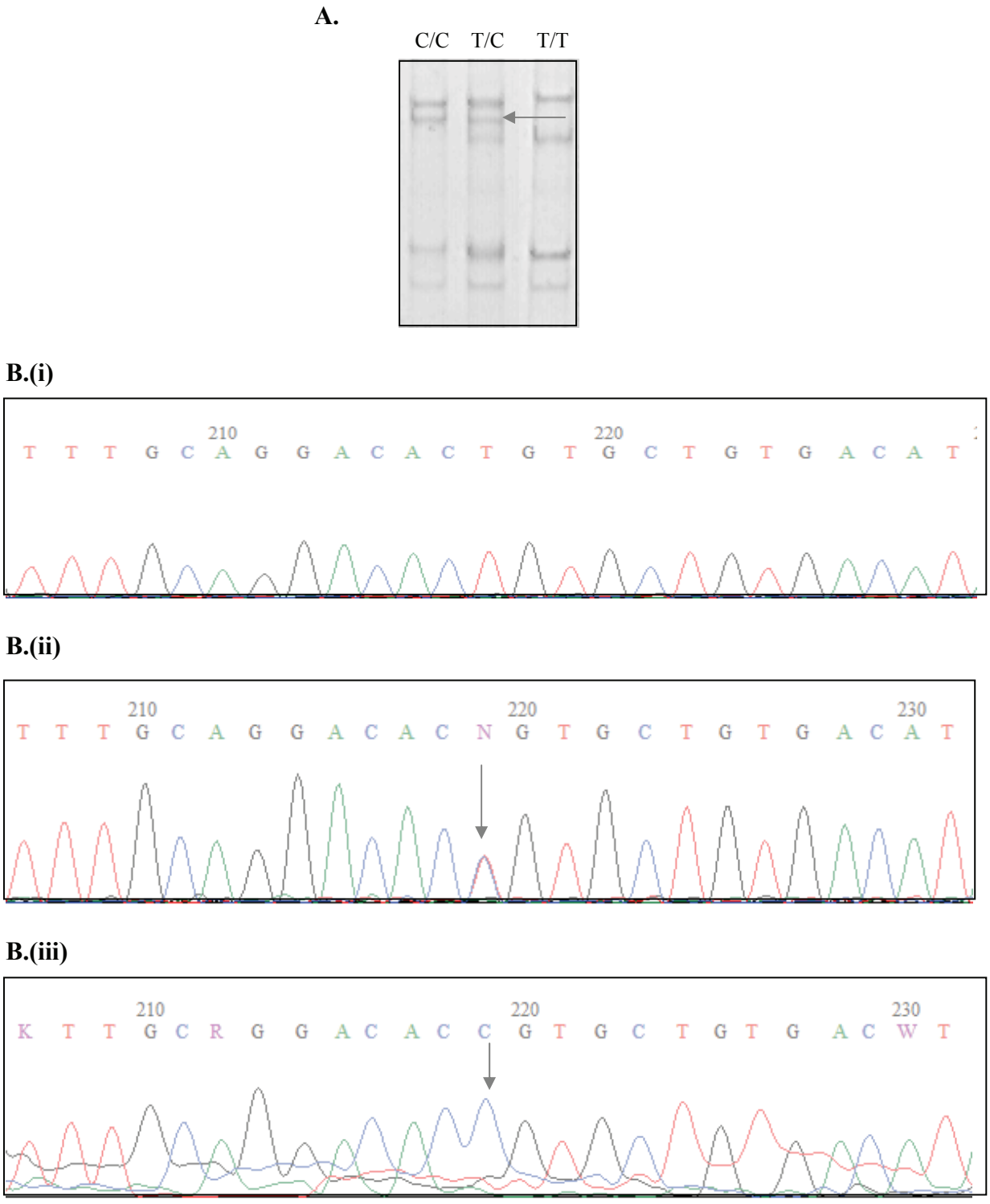


B.(ii)



Legend to Figure 3.4
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrow indicates aberrant band observed in the heteroduplex. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the 5'UTR-364delT variant in the heterozygous state. Arrow indicates the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

Figure 3.5 Schematic representation of the 5'UTR-354T→C variant in the *CP* promoter.



Legend to Figure 3.5
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrow indicates aberrant band. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the 5'UTR-354T→C variant in the heterozygous state (iii) the 5'UTR-354T→C variant in the homozygous state. Arrows indicate the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

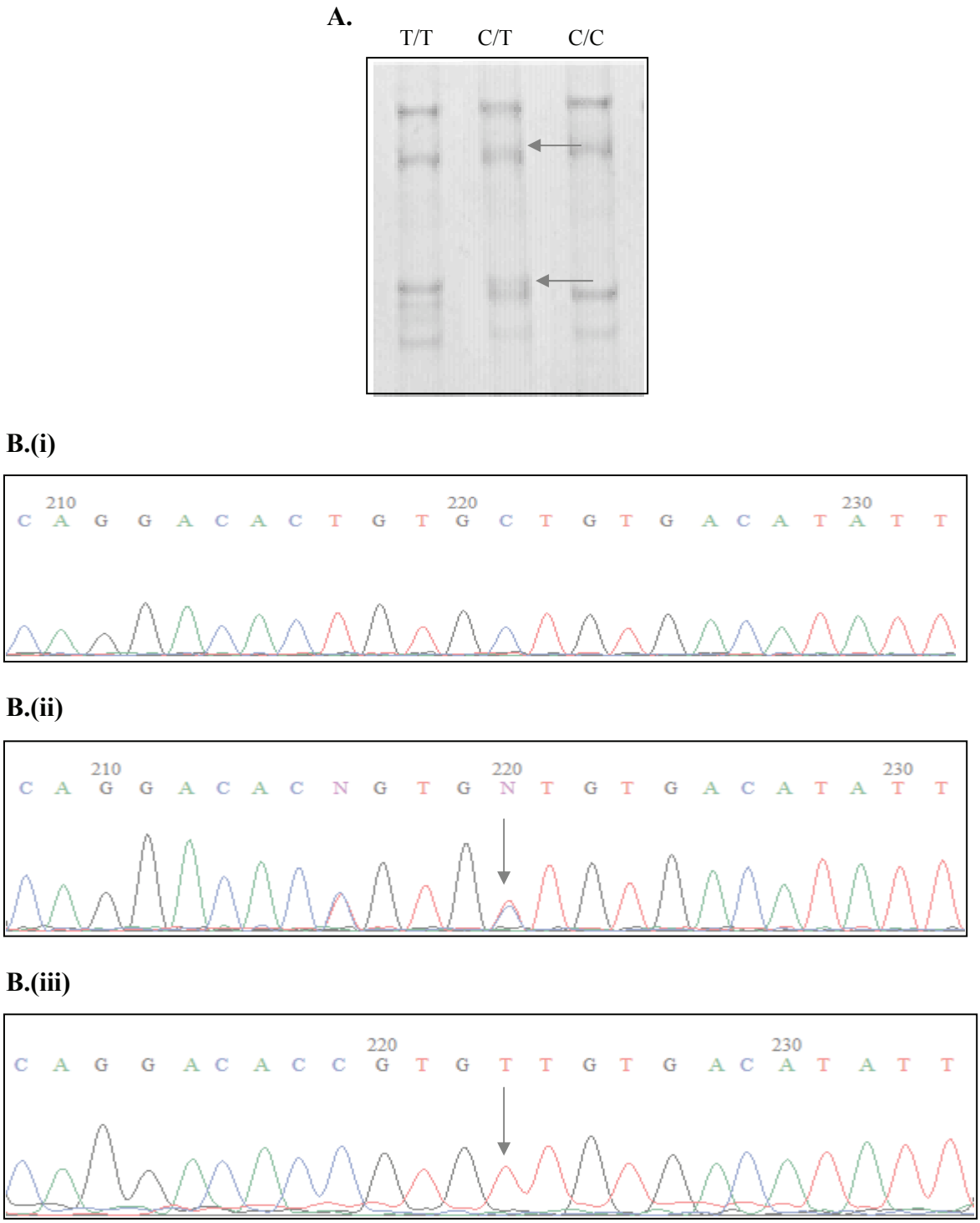
Following HEX-SSCP analysis of promoter fragment 3, a fourth previously identified variant was observed. The 5'UTR-350C→T variant (RefSNP ID: rs34334174) results from a C to T substitution and was identified in this study in the heterozygous state in 16 of 80 (20%) of the patients and 11 of 78 (14.1%) of the controls. In the homozygous state, this variant was detected in three of 80 (3.8%) of the patients and in only one of 78 (1.3%) of the control individuals (Figure 3.6).

HEX-SSCP analysis of fragment 4 (see Table 2.3) revealed two variants, one novel and one previously described.

The previously described variant was identified 282 nucleotides upstream of the ATG (5'UTR-282A→G) (RefSNP ID: rs17838831) and results from an A to G substitution. This variant presented in the heterozygous state in 20 of 53 (37.7%) of the patients and 20 of 64 (31.3%) of the population-matched controls. In the homozygous state, it was detected in one of 53 (1.9%) of the patients and four of 64 (6.3%) of the controls (Figure 3.7).

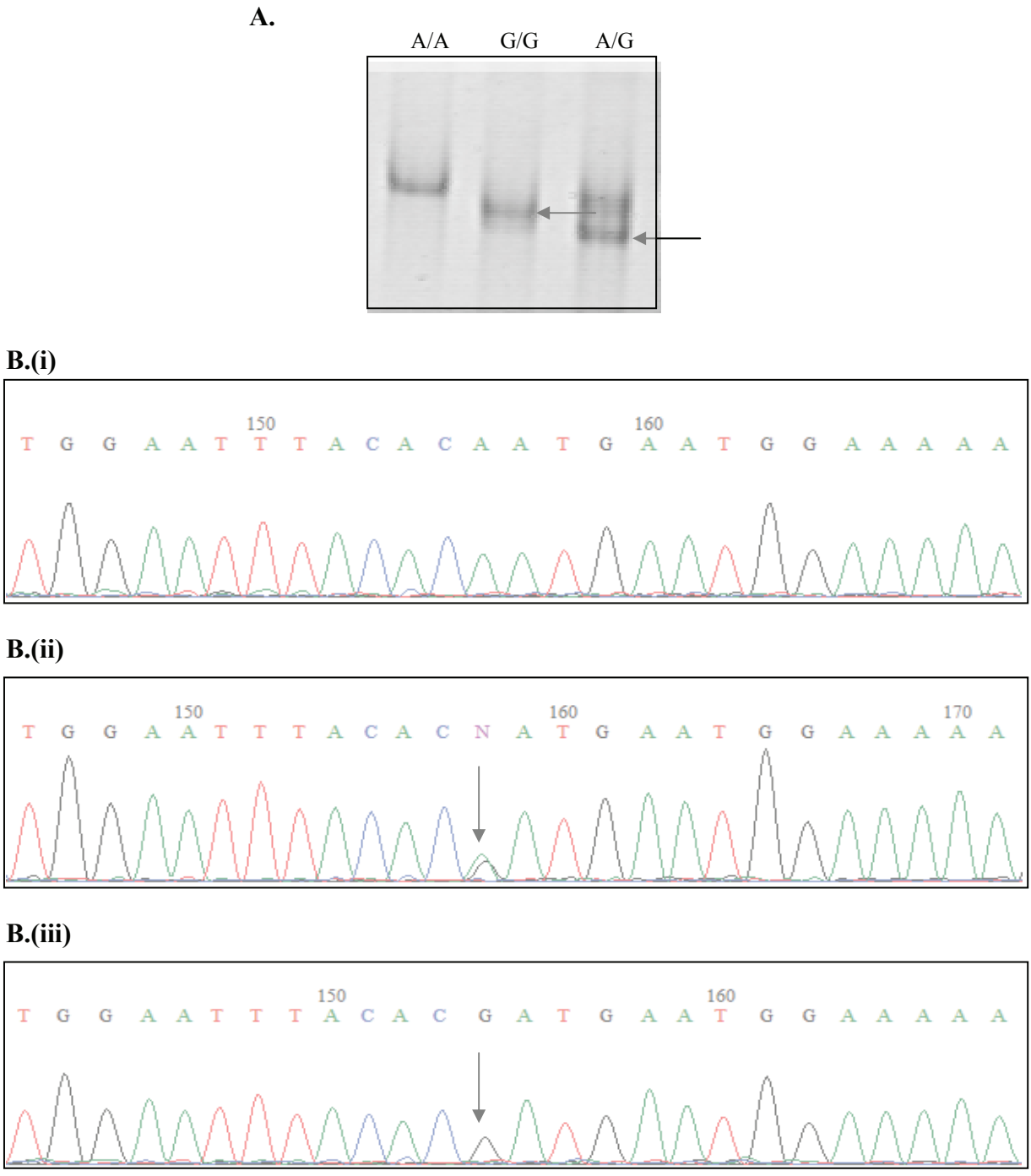
The novel promoter variant occurs as a result of a G to A substitution 308 nucleotides upstream of the translation initiation site (5'UTR-308G→A). This novel polymorphism presented only in the heterozygous state and was not identified in the control individuals. It was detected in five of 52 (9.6%) of the OC patient individuals (Figure 3.8).

Figure 3.6 Schematic representation of the 5'UTR-350C→T variant in the *CP* promoter.



Legend to Figure 3.6
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrows indicate aberrant bands. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the 5'UTR-350C→T variant in the heterozygous state (iii) the 5'UTR-350C→T variant in the homozygous state. Arrows indicate the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

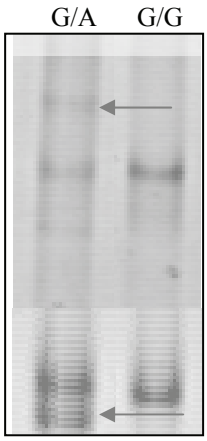
Figure 3.7 Schematic representation of the 5'UTR-282A→G variant in the *CP* promoter.



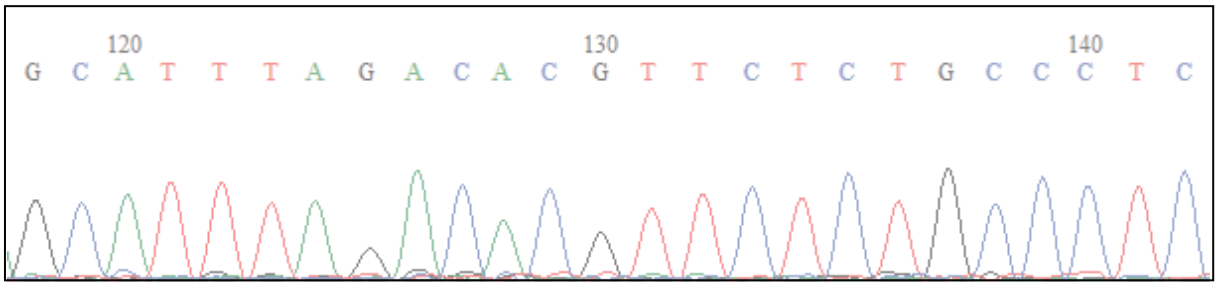
Legend to Figure 3.7
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrow indicates aberrant band observed in the heteroduplex. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the 5'UTR-282A→G variant in the heterozygous state (iii) the 5'UTR-282A→G variant in the homozygous state. Arrows indicate the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

Figure 3.8 Schematic representation of the novel 5'UTR-308G→A variant in the *CP* promoter.

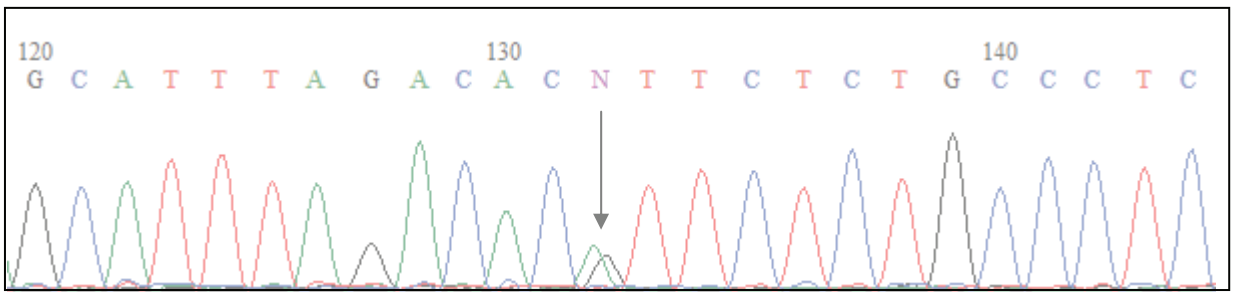
A.



B.(i)



B.(ii)



Legend to Figure 3.8
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrows indicate aberrant bands observed in the SSCP and heteroduplex. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the 5'UTR-308G→A variant in the heterozygous state. Arrow indicates the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

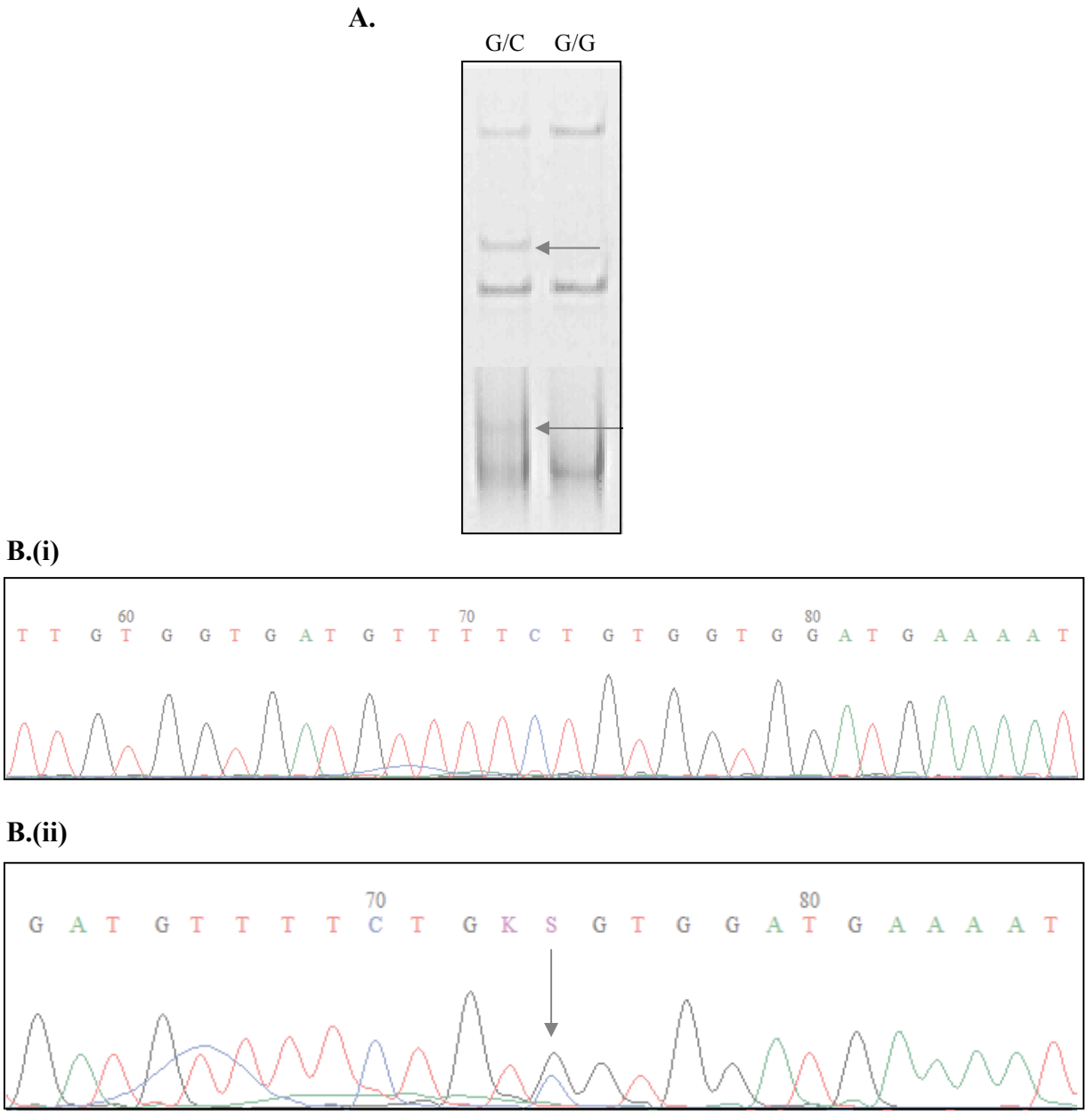
3.1.2 Exonic Variants

Four previously described (V223, R367C, Y425 and D544E) and three novel (T83, V246A and G633) variants were identified in the coding regions of the *CP* gene following HEX-SSCP analysis. The T83 variant was identified by RFLP analysis due to difficulties distinguishing between banding patterns on the HEX-SSCP gels.

The first previously described exonic variant occurs at amino acid position 223 (V223) (RefSNP ID: rs35438054) and does not result in the replacement of the valine (V) amino acid. This synonymous variant was identified in exon 4 of the *CP* gene, and results from a G to C transversion at nucleotide position 661 (g.661 G→C). Following HEX-SSCP analysis, this variant presented only in the heterozygous state, in one of 95 (1.1%) of the patients and in one of 79 (1.3%) of the controls (Figure 3.9).

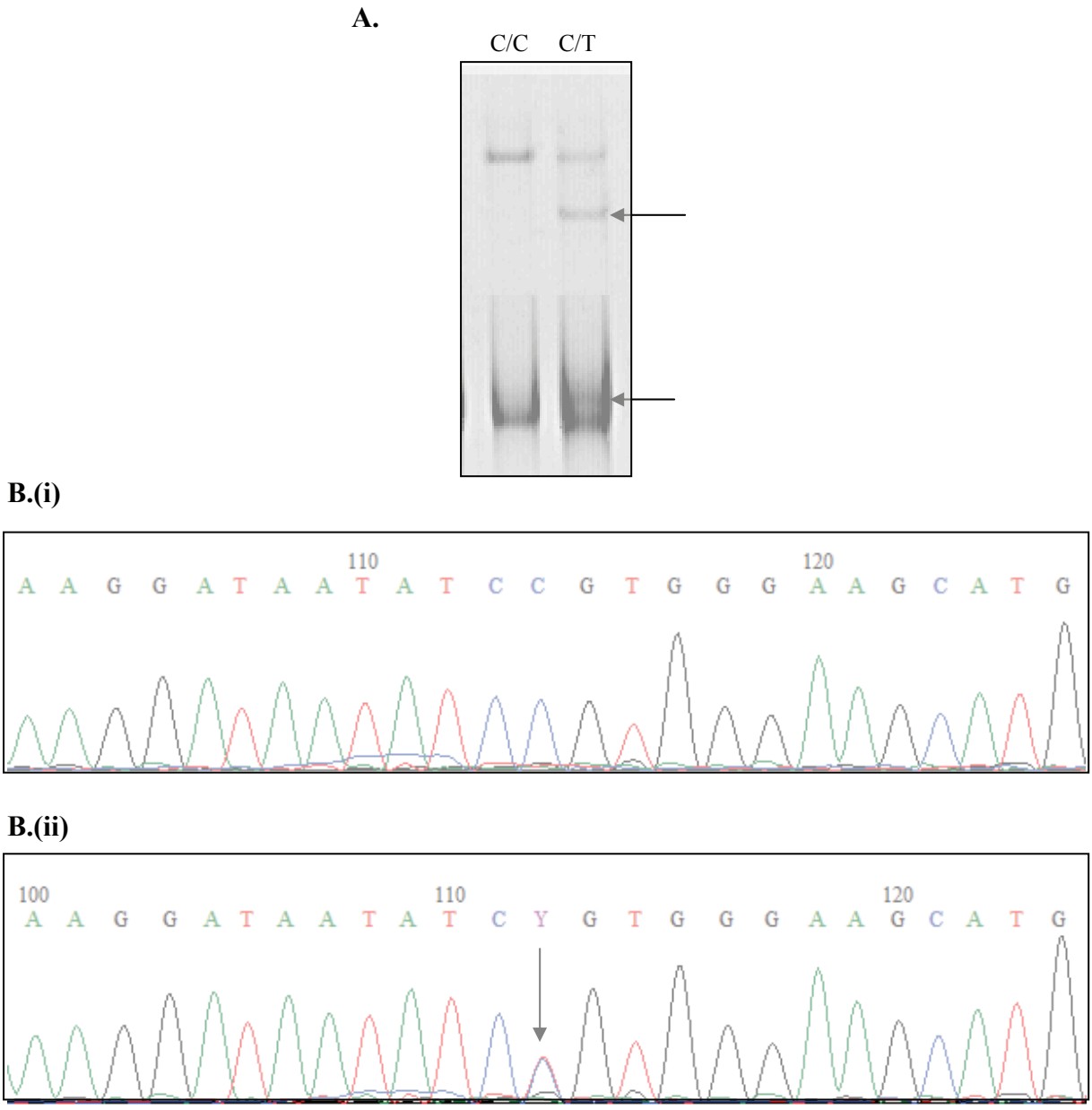
A second previously described single nucleotide substitution was detected following HEX-SSCP analysis of exon 6. This missense mutation is a C to T transition at nucleotide position 1091 (g.1091 C→T) that results in the replacement of arginine (R) with cysteine (C) at amino acid position 367 (R367C) (RefSNP ID: rs34624984). The R367C variant was identified in the heterozygous state in three of 95 (3.2%) of the patient group and in one of 87 (1.1%) of the control group (Figure 3.10). The homozygous state was not observed.

Figure 3.9 Schematic representation of the V223 variant in the *CP* coding region.



Legend to Figure 3.9
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrows indicate aberrant bands observed in the SSCP and heteroduplex. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the V223 variant in the heterozygous state. Arrow indicates the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

Figure 3.10 Schematic representation of the R367C variant in the *CP* coding region.

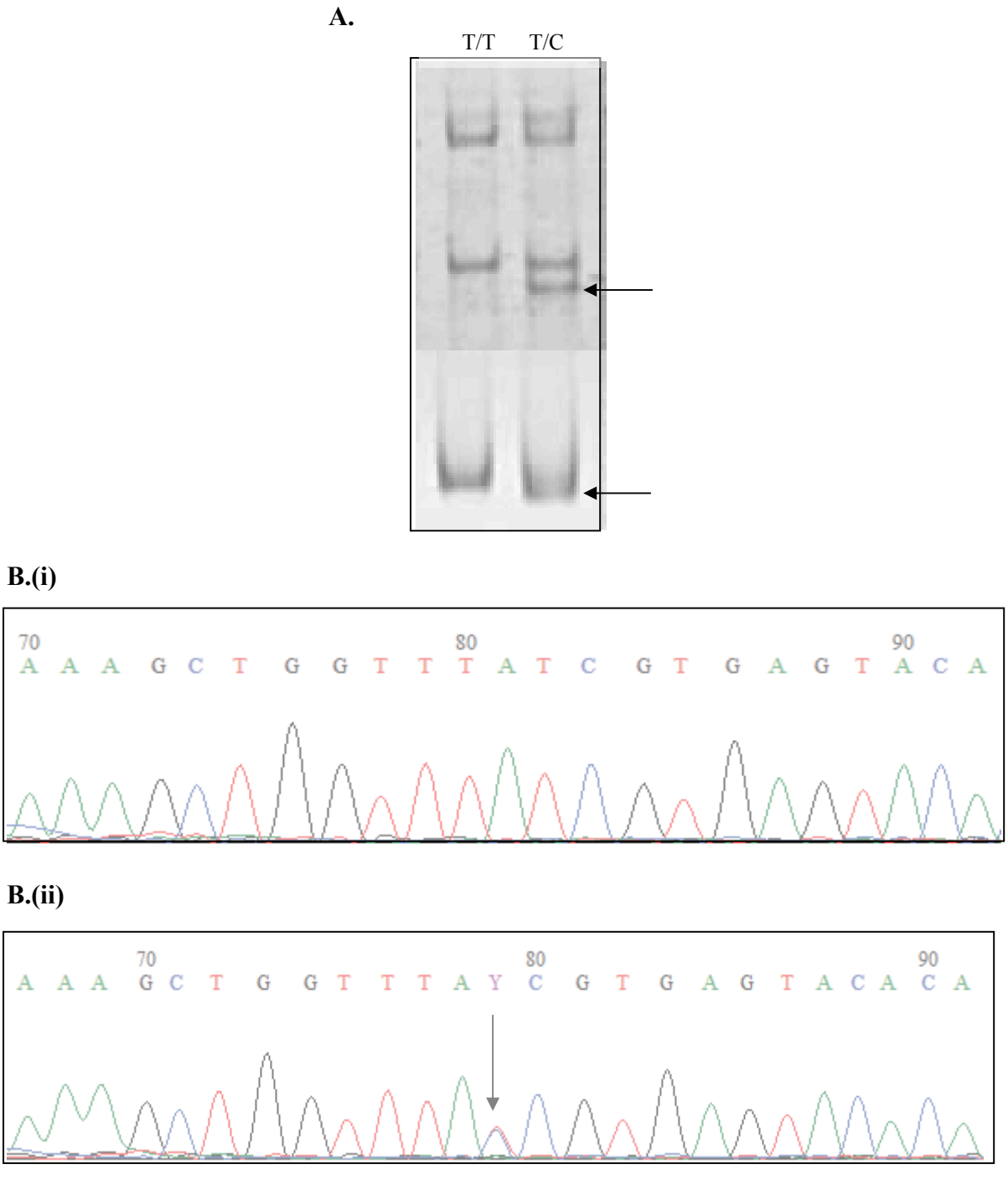


Legend to Figure 3.10
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrows indicate aberrant bands observed in the SSCP and heteroduplex. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the R367C variant in the heterozygous state. Arrow indicates the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

HEX-SSCP analysis of exon 7 of the *CP* gene revealed a single previously described synonymous variant occurring from a T to C substitution at nucleotide position 1268 (g.1268 T→C). This variant does not result in the replacement of tyrosine (T) at amino acid position 425 (Y425) (RefSNP ID: rs34237139). Y425 was restricted to the heterozygous state and was detected in only one of 91 (1.1%) of the OC patients (Figure 3.11).

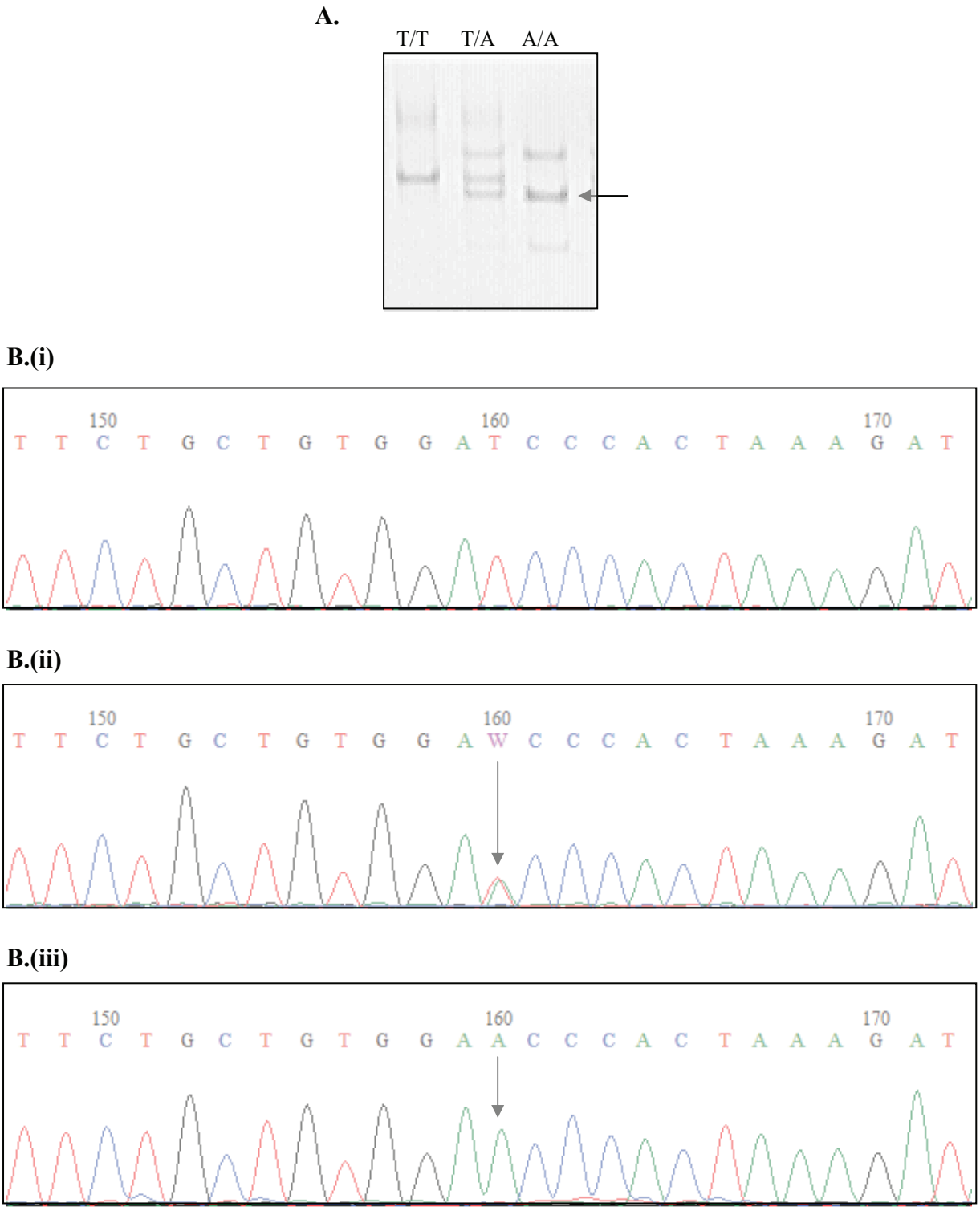
HEX-SSCP analysis of exon 9 of the *CP* gene revealed the presence of a previously identified single base pair substitution, initially annotated as an A to T transversion, based upon the reference sequence employed. However, in this study, it was observed that the T allele occurred in a higher frequency than the A allele, and after considering the ancestral allele (HapMap) and international incidences of these alleles, it was decided to annotate the T allele as the common allele. Therefore the variant was annotated as a T to A substitution at nucleotide position 1625 (g.1625 T→A). This base pair substitution results in the replacement of aspartic acid (D) with glutamic acid (E) at amino acid position 544 (D544E) (RefSNP ID: rs701753). This missense mutation was observed in the heterozygous state in 42 of 95 (44.2%) of the patients and 29 of 81 (35.8%) of the controls. The homozygous state was detected in one of 95 (1.1%) of the patients and seven of 81 (8.6%) of the controls (Figure 3.12).

Figure 3.11 Schematic representation of the Y425 variant in the *CP* coding region.



Legend to Figure 3.11
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrows indicate aberrant bands observed in the SSCP and heteroduplex. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the Y425 variant in the heterozygous state. Arrow indicates the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

Figure 3.12 Schematic representation of the D544E variant in the *CP* coding region.



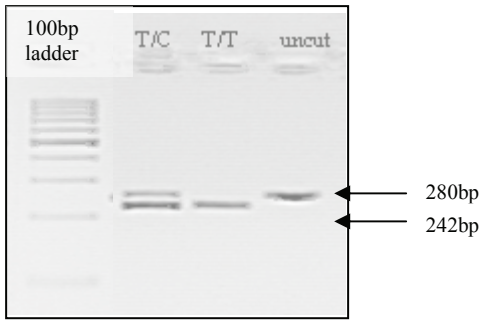
Legend to Figure 3.12
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrow indicates aberrant band. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the D544E variant in the heterozygous state (iii) the D544E variant in the homozygous state. Arrows indicate the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

A novel synonymous variant was identified after mutation analysis of exon 2. This variant occurs from a T to C substitution at nucleotide position 249 (g.249 T→C), resulting in no change of the amino acid threonine (T) at position 83 (T83). Results from HEX-SSCP analysis of exon 2 were uninformative and the patient and control samples were therefore subjected to RFLP analysis for genotyping of this variant. The T83 variant abolishes a *SfcI* recognition site (5' C↓TRYAG 3'). The homozygous wild type genotype results in two DNA fragments of 242bp and 38bp respectively, the heterozygous genotype results in three DNA fragments of 280bp, 242bp and 38bp, while the homozygous variant genotype results in a single DNA fragment of 280bp. The 38bp DNA fragment was not visible on a 3% (w/v) agarose gel but the 280bp and 242bp DNA fragments were able to be used for genotyping purposes. This novel variant was restricted to the heterozygous state and was identified in three of 92 (3.3%) of the patients and only one of 85 (1.2%) of the control individuals (Figure 3.13).

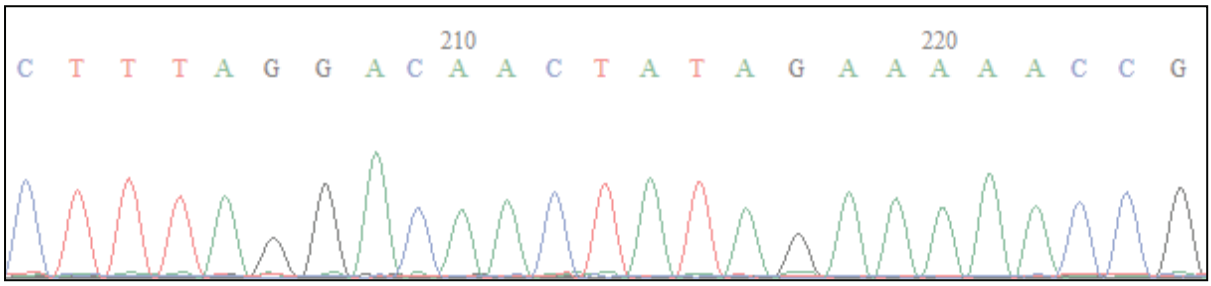
The second novel exonic variant identified in this study is caused by a T to C substitution that occurs at nucleotide position 729 (g.729 T→C) and results in the replacement of valine (V) with alanine (A) at amino acid position 246 (V246A). This missense variant was also identified in exon 4 of the *CP* gene following HEX-SSCP analysis and was restricted to both the heterozygous state and the patient group, where it was detected in two of 95 (2.1%) of the individuals (Figure 3.14).

Figure 3.13 Schematic representation of the novel T83 variant in the *CP* coding region.

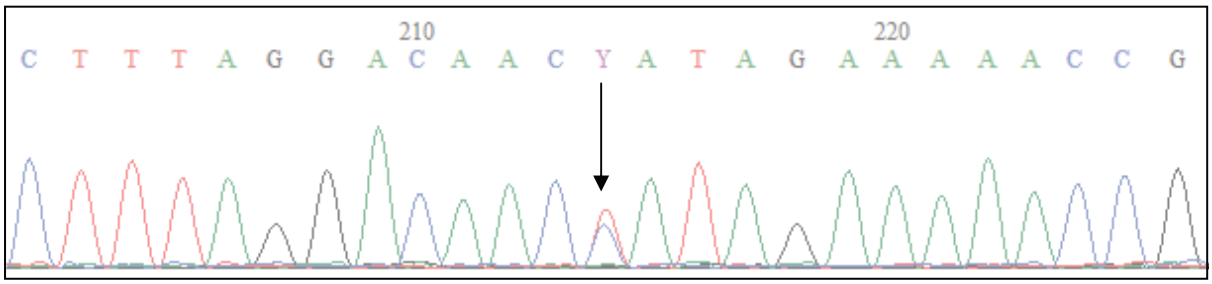
A.



B.(i)

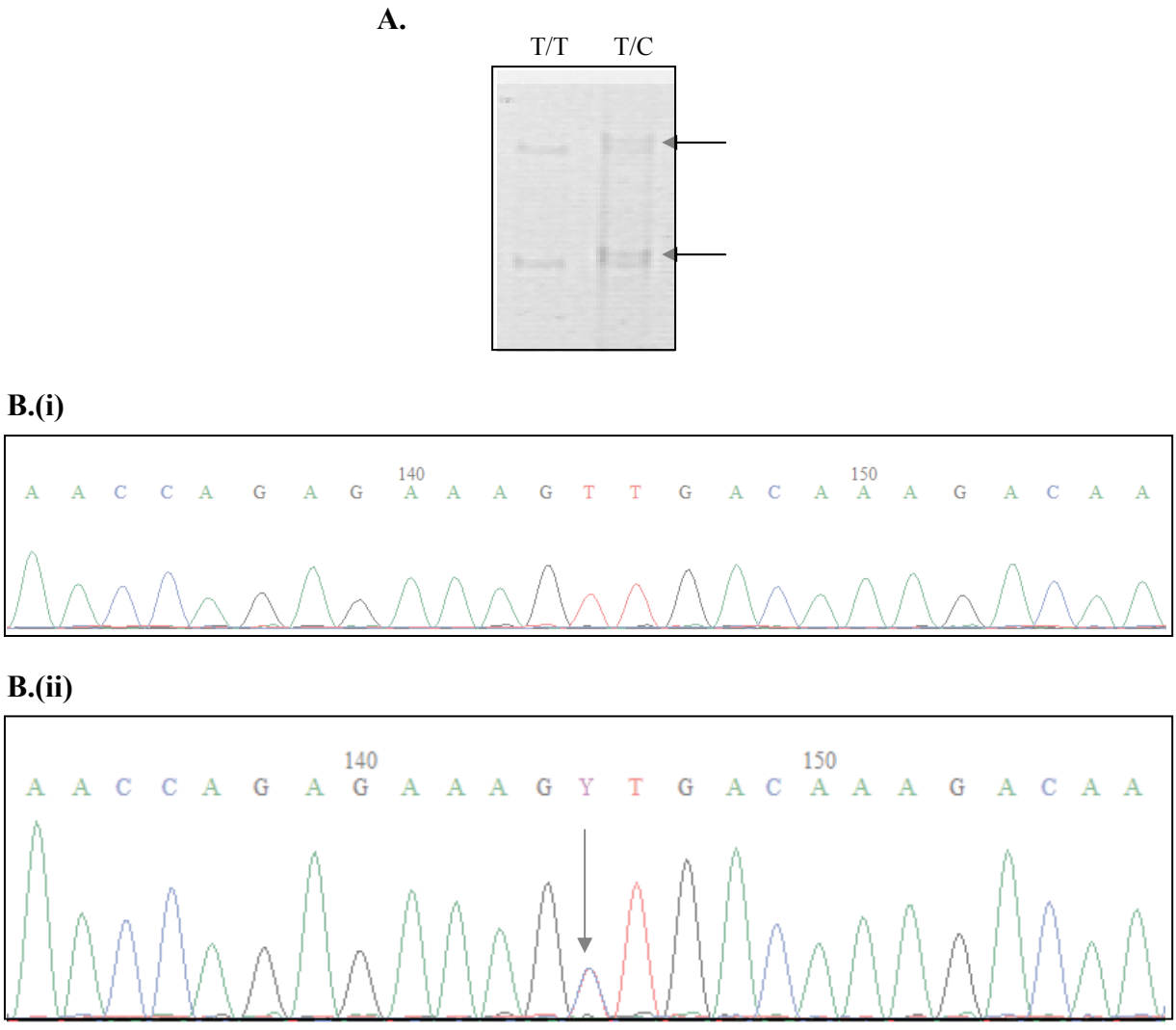


B.(ii)



Legend to Figure 3.13
A. 3% agarose gel stained with EtBr and visualised using ultraviolet light transillumination. Arrows indicate size of bands observed following RFLP analysis. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the T83 variant in the heterozygous state. Arrow indicates the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

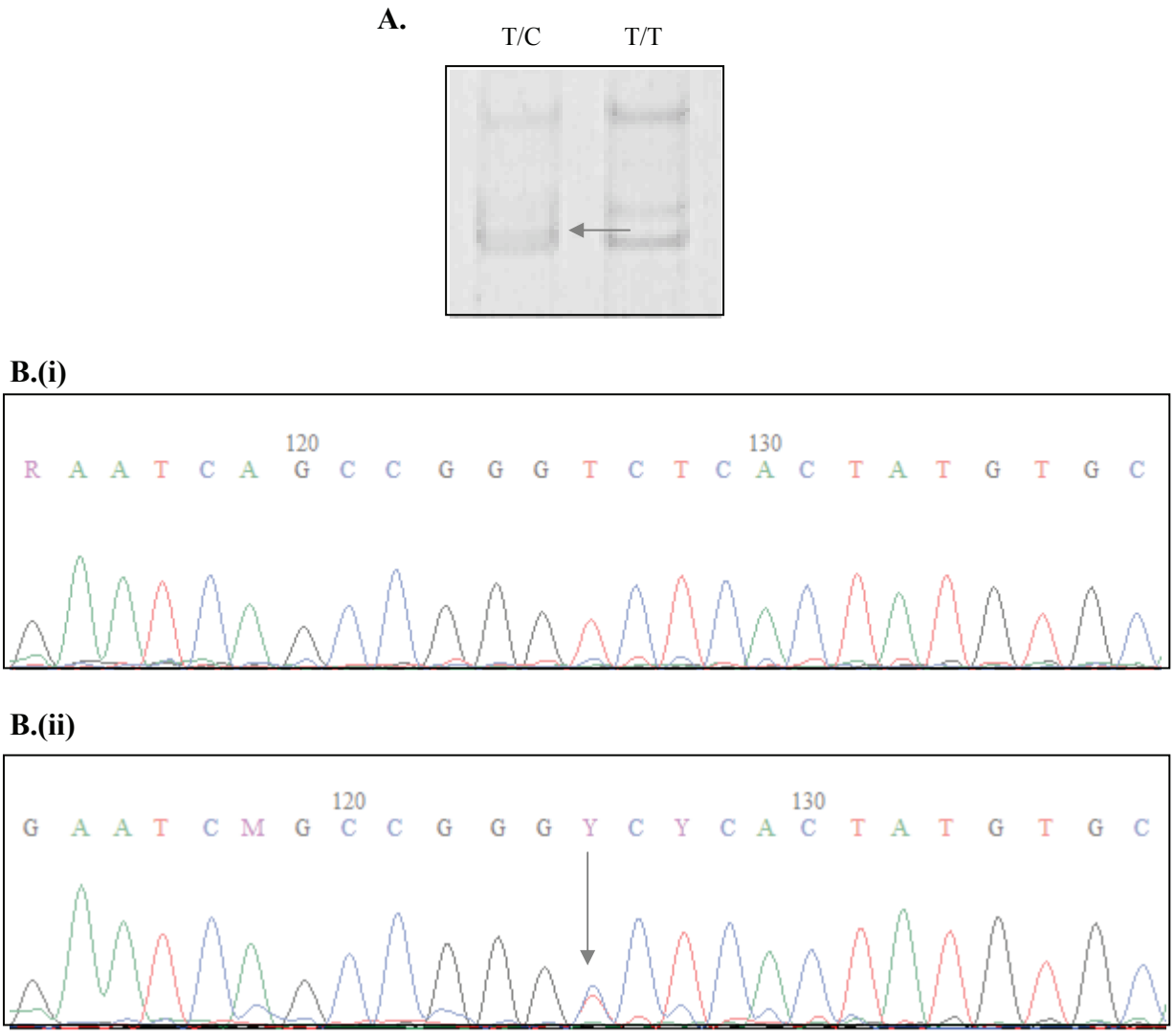
Figure 3.14 Schematic representation of the novel V246A variant in the *CP* coding region.



Legend to Figure 3.14
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrows indicate aberrant bands. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the V246A variant in the heterozygous state. Arrow indicates the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

The final novel variant detected in the *CP* coding region was identified after HEX-SSCP analysis of exon 11. It occurs from a T to C transition at nucleotide position 1891 (g.1891 T→C). This polymorphism does not result in the replacement of glycine (G) at amino acid position 633 (G633). The heterozygous state of this synonymous variant was observed in 13 of 88 (14.8%) of the patient group (Figure 3.15). It was not detected in the homozygous state and was absent from the control group.

Figure 3.15 Schematic representation of the novel G633 variant in the *CP* coding region.



Legend to Figure 3.15
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrow indicates aberrant band. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the G633 variant in the heterozygous state. Arrow indicates the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

3.1.3 Intronic Variants

Three previously described intronic variants (IVS4-14C→T, IVS7+9T→C and IVS15-12T→C) were identified following HEX-SSCP analysis of the non-coding regions flanking the exonic regions screened in this study.

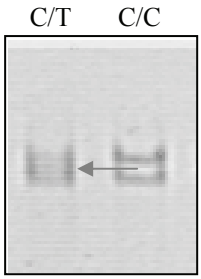
The first intronic variant was observed after HEX-SSCP analysis of exon 5. The C to T substitution occurs 14 nucleotides upstream from the start of exon 5 (IVS4-14C→T) (RefSNP ID: rs34067682). It was observed exclusively in the heterozygous state and was identified in 37 of 95 (38.9%) of the patients and in 26 of 84 (31%) of the controls (Figure 3.16).

The second previously described non-coding variant was a T to C transition in intron 7, located 9 nucleotides downstream of the end of exon 7 (IVS7+9T→C) (RefSNP ID: rs35272481). This variant was detected in only one of 91 (1.1%) of the patients and only in the heterozygous state (Figure 3.17).

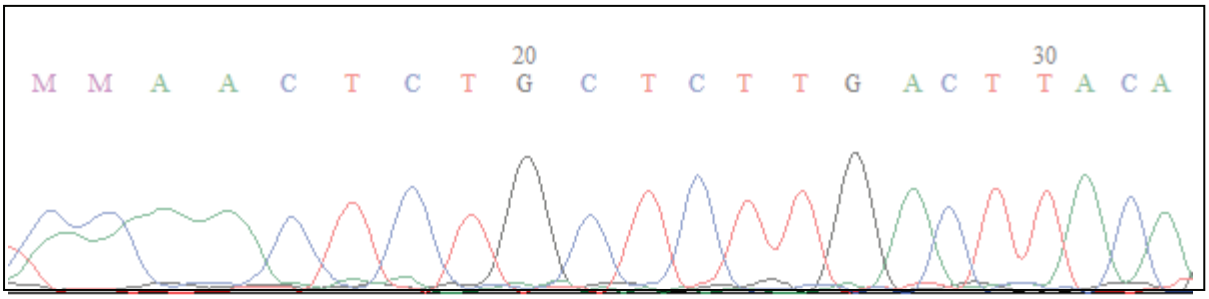
The third intronic variant, a T to C substitution, was identified after HEX-SSCP analysis of exon 16. It is located in intron 15, 12 nucleotides upstream of the start of exon 16 (IVS15-12T→C) (RefSNP ID: rs16861582). This variant was observed at high frequencies in both the patient and control groups. It was identified in 36 of 91 (39.6%) of the patients in the heterozygous state and in 44 of 88 (50%) of the controls. In the homozygous state it was found in 15 of 91 (16.5%) of the patients and in 15 of 88 (17%) of the control individuals (Figure 3.18).

Figure 3.16 Schematic representation of the IVS4-14C→T variant in the *CP* non-coding region.

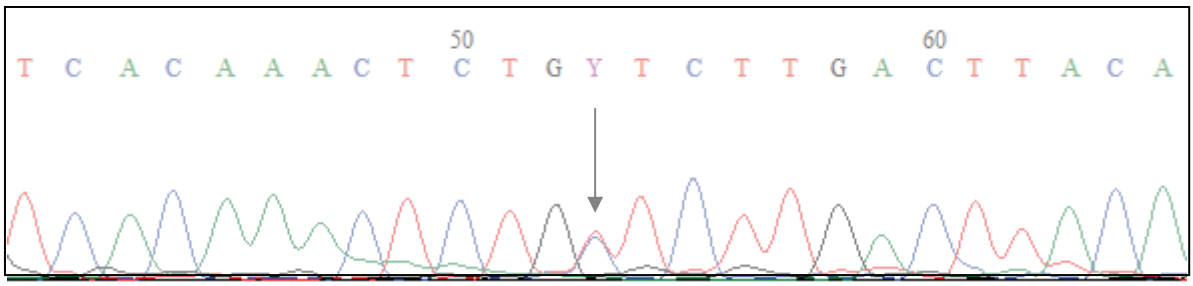
A.



B.(i)

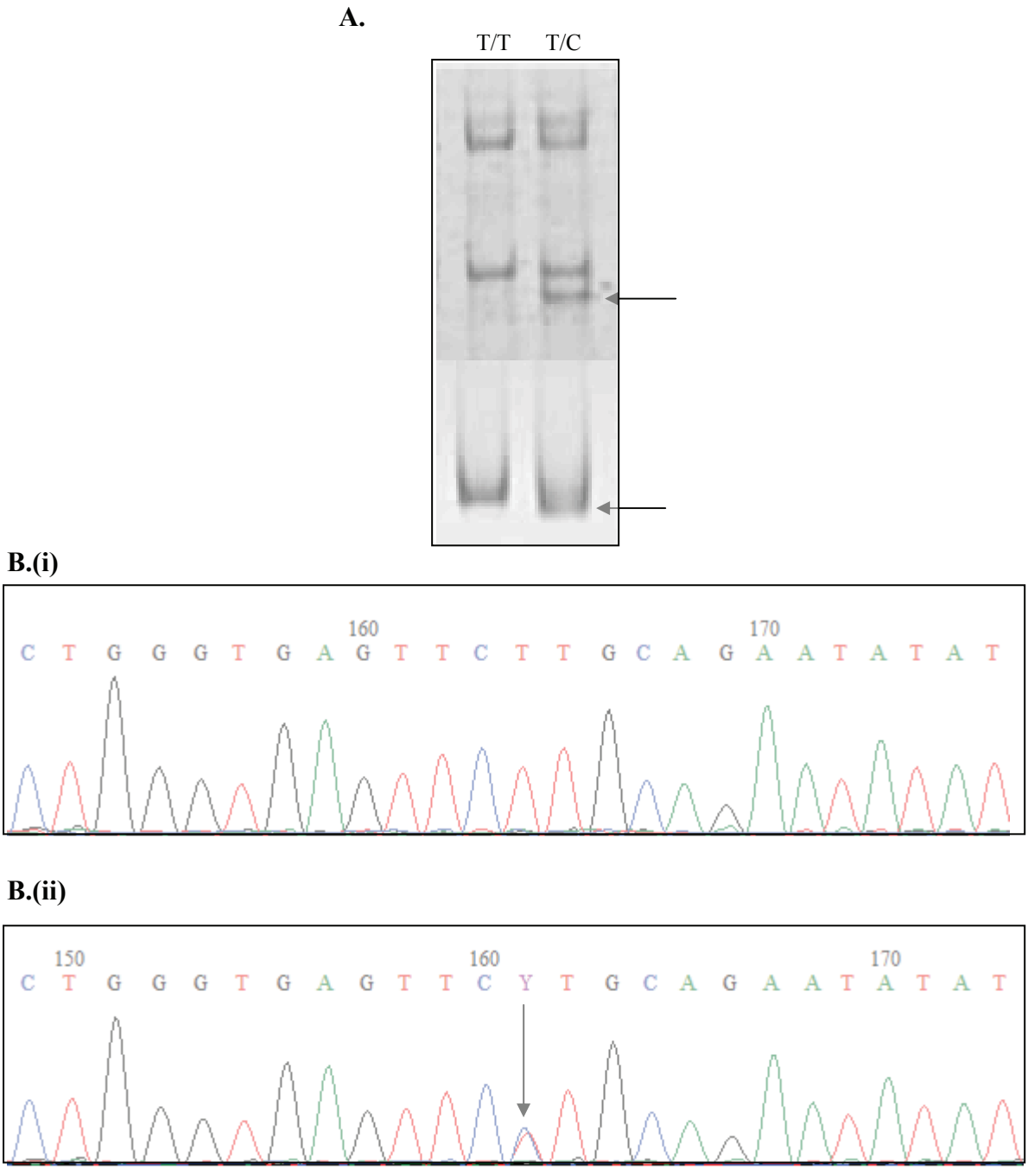


B.(ii)



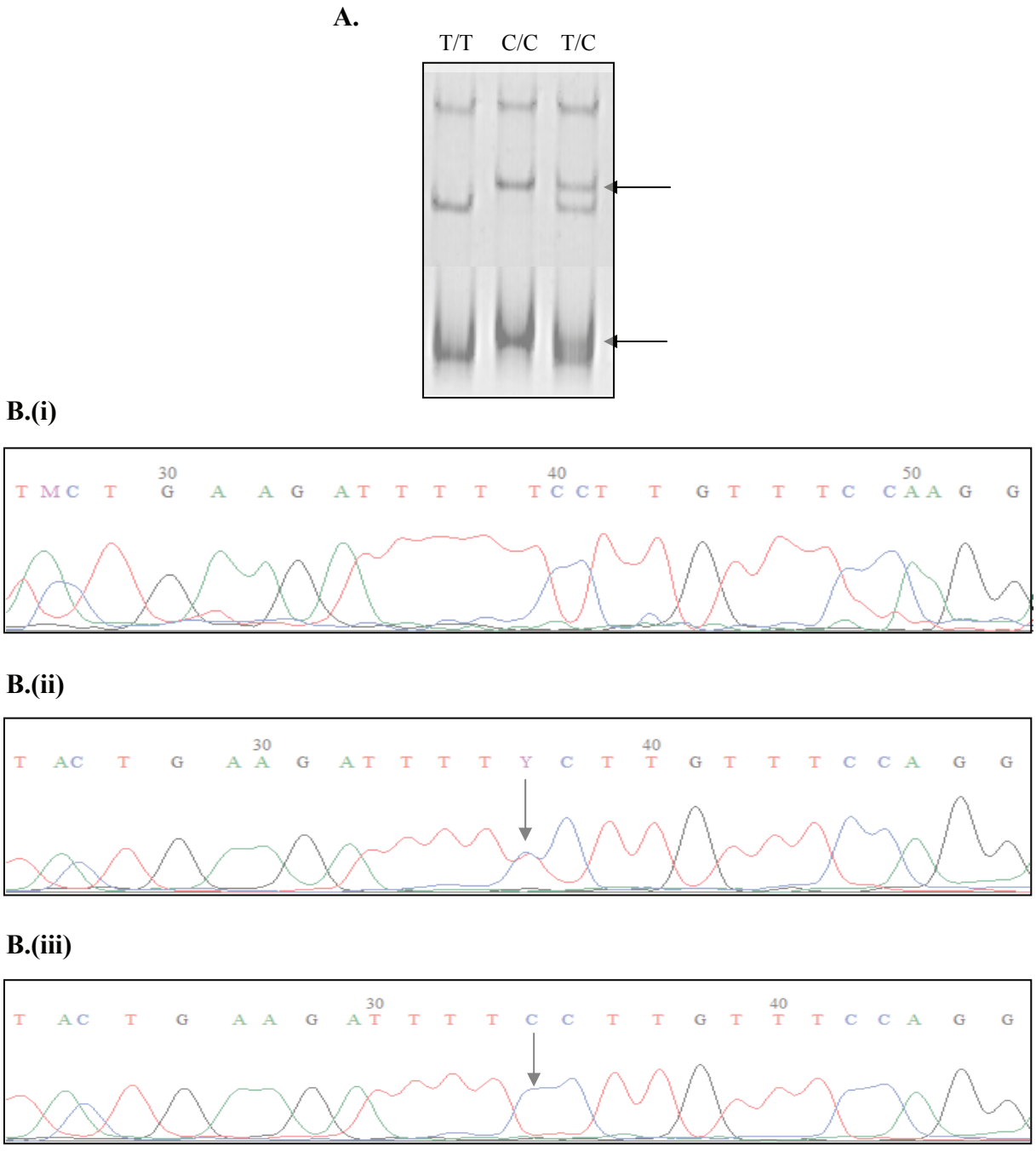
Legend to Figure 3.16
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrow indicates aberrant band. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the IVS4-14C→T variant in the heterozygous state. Arrow indicates the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

Figure 3.17 Schematic representation of the IVS7+9T→C variant in the *CP* non-coding region.



Legend to Figure 3.17
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrows indicate aberrant bands observed in the SSCP and heteroduplex. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the IVS7+9T→C variant in the heterozygous state. Arrow indicates the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

Figure 3.18 Schematic representation of the IVS15-12T→C variant in the *CP* non-coding region.



Legend to Figure 3.18
A. HEX-SSCP gel stained with EtBr and visualised using ultraviolet light transillumination. Arrows indicate aberrant bands observed in the SSCP and heteroduplex. **B.** Sequencing electropherograms indicating (i) the wild type sequence and (ii) the IVS15-12T→C variant in the heterozygous state (iii) the IVS15-12T→C variant in the homozygous state. Arrows indicate the point of variation, *green* adenine or A, *blue* cytosine or C, *black* guanine or G, *red* thymine or T.

3.2 STATISTICAL ANALYSIS

The variants identified in this study were subjected to statistical analysis. The polymorphic allele and genotype frequencies observed in the patients were compared to those of the population-matched control individuals and are shown in Tables 3.1 and 3.2. A Chi-squared (χ^2) goodness-of-fit test using 2×2 (STATISTICA 8.0) or 2×3 contingency tables (Microsoft Excel Software), was employed to determine the probability (*P*) values for each identified variant in the patient and control groups (Tables 3.1 and 3.2). A *P* value of less than 0.05 ($P<0.05$) was considered to be statistically significant (highlighted in yellow in Tables 3.1 and 3.2).

The patient and control groups were tested for deviation from Hardy-Weinberg Equilibrium (HWE) by utilising the χ^2 goodness-of-fit test in the Tools for population genetics association studies (TFPGA) program v1.3 (Miller 1997). A probability value of greater than 0.05 ($P>0.05$) was used to identify if a particular group conformed to HWE. The *P* values for the patient and control groups for each particular variant locus are shown in Tables 3.3 and 3.4.

Table 3.1 Genotypic and polymorphic allele frequencies of variants identified in the *CP* promoter region in the Black South African population.

Variant	Study cohort	*n*	Genotype			*P*	*2n*	Variant Allele[a]	*P*
			CC	CG	GG			G	
5'UTR-567C→G[b,e]	Patients	84	1.00	0.00	0.00	**0.157**	168	0.00	**0.159**
*rs34053109	Controls	85	0.98	0.02	0.00		170	0.01	
			TT	TC	CC			C	
5'UTR-563T→C[c]	Patients	90	0.41	0.51	0.08	**0.229**	180	0.33	**0.127**
*rs17838834	Controls	85	0.52	0.45	0.04		170	0.23	
			CC	CT	TT			T	
5'UTR-439C→T[b]	Patients	79	0.99	0.01	0.00	**0.305**	158	0.01	**0.308**
*rs701749	Controls	78	0.96	0.04	0.00		156	0.02	
			TT	T-	--			-	
5'UTR-364delT[b,d]	Patients	79	0.98	0.03	0.00	**0.157**	158	0.01	**0.159**
*rs17838833	Controls	78	1.00	0.00	0.00		156	0.00	
			TT	TC	CC			C	
5'UTR-354T→C[c]	Patients	80	0.43	0.50	0.08	**0.321**	160	0.33	**0.481**
*rs17838832	Controls	78	0.50	0.42	0.08		156	0.30	
			CC	CT	TT			T	
5'UTR-350C→T[c]	Patients	80	0.76	0.20	0.04	**0.284**	160	0.14	**0.125**
*rs34334174	Controls	78	0.85	0.14	0.01		156	0.08	
			GG	GA	AA			A	
5'UTR-308G→A[b,d]	Patients	52	0.90	0.10	0.00	**0.011**	104	0.05	**0.012**
[#]This study	Controls	64	1.00	0.00	0.00		128	0.00	
			AA	AG	GG			G	
5'UTR-282A→G[c]	Patients	53	0.60	0.38	0.02	**0.539**	106	0.21	**0.835**
*rs17838831	Controls	64	0.63	0.31	0.06		128	0.22	

*RefSNP identification number; [#]This study; [a]Allele frequencies of only the polymorphic allele denoted; [b]variants identified only in the heterozygous state; [c]variants identified in both the heterozygous and homozygous states; [d]variants identified in only the patient group; [e]variants identified only in the control group. Abbreviations: 5', 5 prime end; A, adenine; C, cytosine; del, deletion; G, guanine; *n*, number of individuals; *2n*, number of alleles; *P*, probability value; T, thymine; UTR, untranslated region.

Table 3.2 Genotypic and polymorphic allele frequencies of variants identified in the *CP* exonic and intronic regions in the Black South African population.

Exon/ Intron	Variant	Study cohort	*n*	Genotype			*P*	*2n*	Variant Allele[a]	*P*
2	**T83**[b]			TT	TC	CC			C	
	[#]This study	Patients	92	0.97	0.03	0.00	**0.357**	184	0.02	**0.360**
		Controls	84	0.99	0.01	0.00		168	0.01	
4	**V223**[b]			GG	GC	CC			C	
	*rs35438054	Patients	95	0.99	0.01	0.00	**0.937**	190	0.01	**0.937**
		Controls	85	0.99	0.01	0.00		170	0.01	
4	**V246A**[b,d]			TT	TC	CC			C	
	[#]This study	Patients	95	0.98	0.02	0.00	**0.179**	190	0.01	**0.180**
		Controls	85	1.00	0.00	0.00		170	0.00	
4	**IVS4-14C→T**[b]			CC	CT	TT			T	
	*rs34067682	Patients	84	0.69	0.31	0.00	**0.959**	168	0.12	**0.963**
		Controls	83	0.69	0.31	0.00		166	0.12	
6	**R367C**[b]			CC	CT	TT			T	
	*rs34624984	Patients	94	0.97	0.03	0.00	**0.356**	188	0.02	**0.359**
		Controls	86	0.99	0.01	0.00		172	0.01	
7	**Y425**[b,d]			TT	TC	CC			C	
	*rs34237139	Patients	91	0.99	0.01	0.00	**0.327**	182	0.01	**0.328**
		Controls	87	1.00	0.00	0.00		174	0.00	
7	**IVS7+9T→C**[b,d]			TT	TC	CC			C	
	*rs35272481	Patients	91	0.99	0.01	0.00	**0.327**	182	0.01	**0.328**
		Controls	87	1.00	0.00	0.00		174	0.00	
9	**D544E**[c]			TT	TA	AA			A	
	*rs701753	Patients	95	0.55	0.44	0.01	**0.380**	190	0.23	**0.463**
		Controls	81	0.56	0.36	0.09		162	0.27	
11	**G633**[b,d]			TT	TC	CC			C	
	[#]This study	Patients	88	0.85	0.15	0.00	**0.0003**	176	0.07	**0.0004**
		Controls	84	1.00	0.00	0.00		168	0.00	
15	**IVS15-12T→C**[c]			TT	TC	CC			C	
	*rs16861582	Patients	93	0.28	0.56	0.16	**0.453**	186	0.44	**0.616**
		Controls	88	0.33	0.51	0.16		176	0.41	

[*]RefSNP identification number; [#]This study, [a]Allele frequencies of only the polymorphic allele denoted; [b]variants identified only in the heterozygous state; [c]variants identified in both the heterozygous and homozygous states, [d]variants identified in only the patient group. Abbreviations: C, cytosine; IVS, intervening sequence; *n*, number of individuals; *2n*, number of alleles; *P*, probability value; T, thymine.

Table 3.3 The Probability (*P*) values at variant loci in the *CP* gene promoter region were tested for departure from HWE. OC patients and control individuals from the Black South African population were analysed.

Variant	Study cohort	*P*
5'UTR-567C→G	Patients	-
	Controls	0.994
5'UTR-563T→C	Patients	0.363
	Controls	0.313
5'UTR-439C→T	Patients	0.998
	Controls	0.985
5'UTR-364delT	Patients	0.994
	Controls	-
5'UTR-354T→C	Patients	0.459
	Controls	0.964
5'UTR-350C→T	Patients	0.374
	Controls	0.794
5'UTR-308G→A	Patients	0.936
	Controls	-
5'UTR-282A→G	Patients	0.563
	Controls	0.791

-, HWE not calculated as polymorphic allele not detected in respective group. Abbreviations: 5', 5 prime end; A, adenosine; C, cytosine; del, deletion; G, guanine; *P*, Probability value; T, thymidine; UTR, untranslated region.

Table 3.4 The Probability (*P*) values at variant loci in the *CP* gene coding region were tested for departure from HWE. OC patients and control individuals from the Black South African population were analysed.

Exon/Intron	Variant	Study cohort	*P*
2	**T83**	Patients	0.987
		Controls	0.999
4	**V223**	Patients	0.999
		Controls	0.999
4	**V246A**	Patients	0.995
		Controls	-
4	**IVS4-14C→T**	Patients	0.245
		Controls	0.239
6	**R367C**	Patients	0.988
		Controls	0.999
7	**Y425**	Patients	0.999
		Controls	-
7	**IVS7+9T→C**	Patients	0.999
		Controls	-
9	**D544E**	Patients	0.062
		Controls	0.762
11	**G633**	Patients	0.756
		Controls	-
15	**IVS15-12T→C**	Patients	0.433
		Controls	0.882

-, HWE not calculated as polymorphic allele not detected in respective group. Abbreviations: C, cytosine; del, deletion; *P*, Probability value; T, thymidine.

3.2.1 The HapMap project

The HapMap project is in the process of developing a map of DNA sequence variation in the human genome that is available for public use. The HapMap study involves 270 individuals from three populations, European, Asian and African. The European population is represented by samples from the *Centre d'Etude du Polymorphisme Humain* (CEPH), the Asian population by unrelated Japanese and Han Chinese individuals and the African population by family groups from Nigeria. The aim of the project is to identify single nucleotide polymorphisms (SNPs) in these populations as well as the degree of association among them. Although not all populations are present in this study, the HapMap project is useful as a starting point when identifying variants in other populations.

The polymorphic allele frequencies of the known variants identified in the South African Black control group (unaffected individuals) were compared to the allele frequencies for Sub-Saharan Africans available on HapMap (Tables 3.5 and 3.6). Although the Black South Africans screened in this study do not display complete genetic similarity to the Sub-Saharan Africans on HapMap, the comparison of the allele frequencies does provide a minimum validation of the accuracy of the results obtained. The allele frequencies for the variants identified in the Black South African OC patients have also been included for reference in Tables 3.5 and 3.6, but should not be compared to the HapMap data as the allele frequencies available on this database refer to the general population and not a select group of individuals with a particular disease. HapMap European allele frequencies have also been provided.

Table 3.5 HapMap allele frequencies for known variants identified in the promoter region of the *CP* gene in this study.

Variant	Polymorphic Allele Frequencies			HapMap Population Group
	SA Black Patients	SA Black Unaffected Individuals	HapMap Unaffected Individuals	
5'UTR-567C→G	0.00	0.01	0.02	SS African
			0.00	European
5'UTR-563T→C	0.33	0.23	0.07	SS African
			0.07	European
5'UTR-439C→T	0.01	0.02	0.00	SS African
			0.11	European
5'UTR-364delT	0.01	0.00	0.03	SS African
			0.09	European
5'UTR-354T→C	0.33	0.30	0.05	SS African
			0.07	European
5'UTR-350C→T	0.14	0.08	0.03	SS African
			0.00	European
5'UTR-282A→G	0.21	0.22	0.13	SS African
			0.18	European

Abbreviations: 5', 5 prime end; A, adenine; C, cytosine; del, deletion; G, guanine; SA, South African; SS African, Sub-Saharan African; T, thymine; UTR, untranslated region.

Table 3.6 HapMap allele frequencies for known variants identified in the coding regions of the *CP* gene in this study.

Exon/Intron	Variant	Polymorphic Allele Frequencies			HapMap Population Group
		SA Black Patients	SA Black Unaffected Individuals	HapMap Unaffected Individuals	
4	**V223**	0.01	0.01	0.08	SS African
				0.00	European
4	**IVS4-14C→T**	0.12	0.12	0.06	SS African
				0.07	European
6	**R367C**	0.02	0.01	0.02	SS African
				0.00	European
7	**Y425**	0.01	0.00	0.04	SS African
				0.00	European
7	**IVS7+9T→C**	0.01	0.00	0.04	SS African
				0.00	European
9	**D544E**	0.23	0.27	0.63	SS African
				0.93	European
15	**IVS15-12T→C**	0.44	0.41	0.58	SS African
				0.21-0.32	European

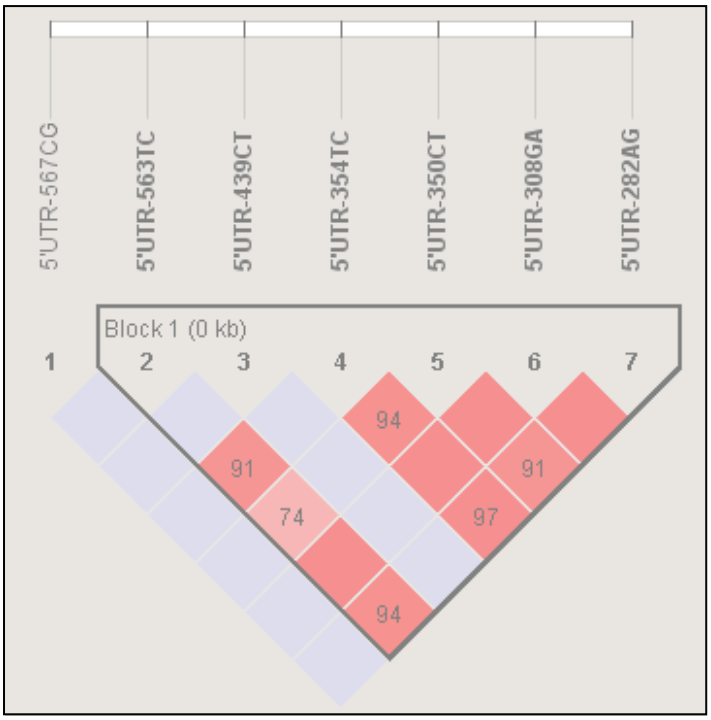
Abbreviations: C, cytosine; del, deletion; IVS, intervening sequence; SA, South African; SS African, Sub-Saharan African; T, thymine.

3.2.2 Haplotype analysis

LD and haplotype analysis, using the Haploview 4.0 software, was performed on all of the variants identified in this study using association and case-control studies (Barrett *et al.* 2005). The LD test was applied, and a haplotype block consisting of the variant alleles of the promoter variants 5'UTR-563T→C, 5'UTR-439C→T, 5'UTR-354T→C, 5'UTR-350C→T, 5'UTR-308G→A and 5'UTR-282A→G was predicted (D'=0.947, LOD=30.81, r^2=0.875) (Figure 3.19). This haplotype spans a genomic region of 281 bp. Haplotype association frequencies demonstrated that this haplotype is significantly associated (*P*=0.01) with OC in the Black South African population.

No other haplotypes were predicted to exist with any of the other variants identified in this study.

Figure 3.19 LD plot of the statistically significant haplotype predicted using Haploview 4.0.



Abbreviations: 5', 5 prime end; A, Adenine; C, Cytosine; G, Guanine; T, Thymine; UTR, Untranslated region.

3.3 BIOINFORMATIC ANALYSIS

3.3.1 *In Silico* analysis of promoter variants

All the variants identified in the 5'UTR region of the *CP* gene were subjected to *in silico* analysis in order to determine if they disrupted or created any putative transcription factor binding sites (TFBS) predicted to occur within this regulatory region. This was achieved by comparing the published *CP* promoter reference sequence (see Appendix 2) with the variant sequence using the appropriate computer software. Identification of the same TFBS by more than one of the programs utilised increases the likelihood that the motif exists. The TRANSFAC®7 database (Wingender *et al.* 2001) was used for analysis of the promoter region in this study. Within this database, the PATCH and MATCH™ programs were employed to analyse the identified promoter variants. These variants were also analysed using the non-redundant JASPAR CORE database (Sandelin *et al.* 2004). For both of these two software programs, the default settings and parameters were employed to determine putative TFBS.

The results obtained through *in silico* analysis of the promoter variants identified in the *CP* gene are shown in Table 3.7.

Table 3.7 Predicted TFBS in the promoter region of the *CP* gene.

Variant	Putative Transcription Factor Binding Sites					
	PATCH		MATCH™		JASPAR CORE	
	Abolished	Created	Abolished	Created	Abolished	Created
5'UTR-567C→G	YY1	GATA1	-	HNF-1	FOXL1, YY1	GATA3
5'UTR-563T→C	YY1, HiNF-D, HiNF-M, HiNF-P, TFIID, TMF, SRF	RXRα, RXRγ, T3Rα	-	-	FOXL1	FOXC1, SP1
5'UTR-439C→T	-	-	-	AP-1	SPI1	SRY, NKX3-1
5'UTR-364delT	FOXM1a, FOXM1b, HNF-3α, HNF-3β	GR	HNF-3β	-	SRY, FOXD1	-
5'UTR-354T→C	SP1	c-myc	-	-	-	-
5'UTR-350C→T	GR, SPI1, NF-Y	-	-	GR, C/EBPα	-	-
5'UTR-308G→A	p53	GR	-	-	MAX, USF1, SPI1	FOXL1,
5'UTR-282A→G	C/EBPα, HNF-1, HNF-3α	-	C/EBPα	-	HLF, FOXL1, SRY, SOX9	GATA2, GATA3, SPIB

Abbreviations: 5', 5 prime end; AP-1, Activator protein 1; C/EBPα, CCAAT/enhancer binding protein alpha; c-myc, Myelocytomatosis viral oncogene c; FOXC1, Forkhead box protein C1; FOXD1, Forkhead box protein D1; FOXL1, Forkhead box protein L1; FOXM1a, Forkhead box M1a; FOXM1b, Forkhead box M1b; GATA1, GATA-binding protein 1; GATA2, GATA-binding protein 2; GATA3, GATA-binding protein 3; GR, Glucocorticoid receptor; HiNF-D, Histone nuclear factor D; HiNF-M, Histone nuclear factor M; HiNF-P, Histone nuclear factor P; HLF, Hepatic leukaemia factor; HNF-1, Hepatocyte nuclear factor 1; HNF-3α, Hepatocyte nuclear factor alpha; HNF-3β, Hepatocyte nuclear factor 3 beta; MAX, Myc-associated factor X; NF-Y, Nuclear transcription factor Y; NKX3-1, NK 3 homeobox 1; p53, Tumour suppressor protein 53; RXRα, Retinoid X receptor alpha; RXRγ, Retinoid X receptor gamma; SOX9, SRY (sex determining region Y)-box-9; SP1, Specificity protein 1; SPI1, spleen focus forming virus (SFFV) proviral integration oncogene 1; SPIB, SPI-B transcription factor; SRF, Serum response factor; SRY, Sex-determining region Y; T3Rα, Thyroid hormone receptor alpha; TFIID, Transcription factor IID; TMF, TATA element modulatory factor; USF1, Upstream transcription factor 1; UTR, untranslated region; YY1, Ying Yang 1.

3.3.2 *In Silico* analysis of exonic variants

The exonic variants identified in this study included three novel (T83, V246A and G633) and four previously described (V233, Y425, R367C and D544E) variants. The ESEfinder (ESE-Exonic Splice Element) program was used to analyse these variants identified in the coding regions of the *CP* gene to determine their possible effect on splicing (Cartegni *et al*. 2003). Results are tabulated in Table 3.8.

Table 3.8 Predicted SR protein binding sites in the coding region of the *CP* gene.

Variant	SR PROTEINS				
	SF2/ASF2	SF2/ASF2 (BRCA1)	SRp40	SRp55	SC35
T83					-
V223	+				
V246A	+	+			
R367C			+		-
Y425			-		+
D544E					-
G633					+

-, SR protein binding site abolished; +, SR protein binding site created. Abbreviations: SR, serine/arginine-rich proteins.

DISCUSSION

This study employed the techniques of HEX-SSCP analysis, RFLP analysis and semi-automated DNA sequencing analysis to identify variations in the DNA sequence of the *CP* gene. Fourteen previously described and four novel variants were identified, with the majority proven to be single nucleotide polymorphisms (SNPs).

SNPs are defined as DNA sequence variations that occur when a single nucleotide is altered. In order to be classified as a SNP, the least commonly occurring allele must be present in at least 1% of the general population. They occur at a frequency of approximately 1 in every 1000 bp (Cooper *et al.* 1985) and are responsible for the majority of the genetic variation discernible between two individuals (Taillon-Miller *et al.* 1998). Identification of SNPs is one of the most common methods employed in the study of genetic variation and they are frequently used as markers to identify genes associated with conferring susceptibility to genetic disease. SNPs themselves are not always considered to be disease-causing. However, they may indicate a risk for the development of a certain genetic disease and are therefore invaluable in the study of genetic disease (Kruglyak 1999).

Promoter region

The cellular environment is in a constant state of flux. Genes have the capability to react to these changes *via* regulation of transcription. Transcription factors (TFs) and other cellular proteins bind to regulatory promoter elements in the genes and subsequently affect the rate of transcription (Butler and Kadonaga 2002). The regulatory elements to which TFs bind are termed transcription factor binding sites (TFBSs).

TFBSs are *cis*-acting elements that are generally comprised of a 6-25 bp recognition sequence and are highly conserved in eukaryotes (Lettice *et al.* 2002, Cooper and Sidow 2003). The distribution frequency of TFBSs is not the same throughout the promoter regions of genes; the frequency appears to increase in regions close to the transcription initiation site (Guo and Jamison 2005). TFBSs are termed as enhancers if the binding of the relevant TF to the regulatory region results in activation of transcription. Conversely, TFBSs are termed as silencers if TF binding results in repression of transcription (Wang and Giaever 1988). It is important to note that although many genes may share identical sequences for a particular TFBS, the importance of each site differs with tissue type and the physiological environment. This, therefore, means that the presence of a particular TFBS in the regulatory region of a gene does not guarantee that the corresponding TF will bind *in vivo* (Li and Johnston 2001).

A large variety of TFs are known to exist within eukaryotic cells. To date, more than 1400 TFs have been described (Lander *et al.* 2001). SNPs that alter the expression of genes have been identified in the promoter regions of genes and are responsible for affecting the rate of transcription (Buckland 2004). These SNPs are termed regulatory SNPs (rSNPs) and they are usually responsible for altering a particular TFBS (Montgomery *et al.* 2007). This may result in a change in the binding affinity of the TF normally associated with this site and a subsequent change in the transcriptional activity of the gene.

In silico analysis of the region of the *CP* promoter spanning the variants identified in this study demonstrated the presence of a number of putative TFBSs (Table 3.7). Only TFs that are responsible for the regulation of *CP* or that are biologically relevant to this study (involved in iron metabolism or liver-specific) will be discussed further.

One putative TFBS that was of particular interest in this study was the tumour suppressor protein 53 (p53) TFBS predicted to be abolished in the presence of the novel 5'UTR-308G→A variant. The *p53* gene encodes a protein that possesses sequence-specific DNA binding properties. It is thought that when cells are exposed to damaging agents, p53 is able to activate target genes involved in cell repair and apoptosis. This therefore prevents cells from replicating after damage or induces apoptosis of cells containing damaged DNA (Kastan *et al*. 1991, Zhan *et al.* 1993). The p53 TF therefore functions indirectly as a tumour suppressor. Mutations in the DNA binding domain of p53 are common in a variety of cancers and may prevent the activation of genes involved in DNA repair. This could enhance the genomic instability of tumour cells (Cho *et al.* 1994). It is feasible therefore to consider that variants in p53 TFBSs could result in decreased levels of gene expression of target genes and possibly result in subsequent tumour development.

The novel 5'UTR-308G→A was found to be statistically associated (P=0.01) with OC in this study (Table 3.1). The fact that the binding motif for the p53 TF in patients with OC is abolished is indicative of the importance of this variant in the Black South African OC population screened in this study. However, this TFBS was only detected in one of the software programs utilised for *in silico* analysis (Table 3.7), and functional studies are therefore required to determine if p53 does interact with this region of the *CP* promoter.

In a study by Fleming and Gitlin (1992), the entire promoter region of the rat (*Rattus norvegicus*) *CP* gene was characterised. An area of approximately 300 bp was determined to be critical for gene expression to occur and has been shown to be highly conserved in humans. The region from -393 to -348 bp upstream of the initiating ATG was found to exert a positive effect on gene expression and also showed sequence homology to the rat albumin D

site, known to bind to CCAAT/enhancer binding protein alpha (C/EBPα). In this study, the 5'UTR-350C→T variant, which lies within the recognition site, was predicted to create a binding site for this TF (Table 3.7). *CP* levels have been shown to be up-regulated in response to certain aggressive cancers which may be mediated by C/EBP binding (Linder *et al.* 1981). 5'UTR-282A→G was predicted by two of the databases to abolish a C/EBPα binding domain (Table 3.7). C/EBP TFs are part of the bZIP family of TFs which have an important role in genes involved in the inflammatory response pathway. CP is an acute-phase protein that is involved in this inflammatory response pathway (Rice 1961). Therefore, disruption to C/EBPα binding domains may result in the down-regulation of *CP* expression leading to impaired inflammatory response.

In addition to C/EBPα, many TFBSs for the GATA-binding protein (GATA) family of TFs were predicted for the *CP* promoter region (Table 3.7). The 5'UTR-567C→G variant was predicted by the *in silico* software programs to create a GATA1 and a GATA3 TFBS. The 5'UTR-282A→G variant was also predicted to create GATA2 and GATA3 TFBSs.

GATA1 has been shown to play a role in the regulation of genes involved in the haem biosynthesis pathway (Ferreira *et al.* 2005). GATA2 is involved in the control of haematopoietic progenitor cells (Tsai *et al.* 1994). The function of GATA 3 is less understood but it has been shown to play a role in erythroid development (Ferreira *et al.* 2005). Genes expressed in the liver, such as *CP*, are known to be regulated by the GATA TFs, although the exact role of these TFs in *CP* expression is unknown. It is therefore important not to disregard the effect that these TFs may exert on *CP* regulation. Additional research will aid in determining the role (if any) of the GATA TFs in *CP* expression.

Hepatocyte nuclear factors (HNFs) regulate genes that are expressed in the cells of the liver. Correct expression of these genes is determined by the binding of a wide variety of HNFs to the regulatory regions of these genes (reviewed by Costa *et al*. 2003). HNFs are members of the steroid/thyroid nuclear receptor family which are expressed predominately in the liver and acts as an essential regulator of liver metabolism and development. The 5'UTR-567C→G variant creates a putative HNF-1 site in the *CP* promoter (Table 3.7). Creation of this TFBS may result in increased expression of the *CP* gene. The 5'UTR-282A→G variant abolishes a putative HNF-1 motif (Table 3.7). The deletion of the T nucleotide at position -364 is predicted to abolish a HNF-3α and HNF-3β TFBS (Table 3.7).

As previously mentioned, CP is an acute-phase protein that plays an important antioxidant role protecting cells against damage that may be caused by oxidative stress (Mukhopadhyay *et al.* 1998). It has been shown that members of the HNF family show an increase in expression in the presence of oxidative stress as a result of hepatitis C virus infection (Qadri *et al*. 2006). Excess iron catalyses the conversion of hydrogen peroxide to free radicals and leads to oxidative stress in various tissues (Andrews 1999). Disruptions to HNF TFBSs may perturb *CP* expression and therefore result in incorrect functioning of the antioxidant mechanism. It is possible that oxidative damage from the presence of excess iron could lead to cell damage and the subsequent formation of tumour cells.

The presence of the 5'UTR-567C→G and the 5'UTR-563T→C variants are predicted to abolish a putative TFBS for YY1 (Table 3.7). YY1 is a zinc finger protein that has been shown to play a role in iron metabolism, particularly by interacting with transferrin (Adrian *et al.* 1996). It has also been implicated in gene silencing in the liver (Yan *et al.* 2001). CP is known to interact with transferrin during the transfer of iron molecules across the basolateral

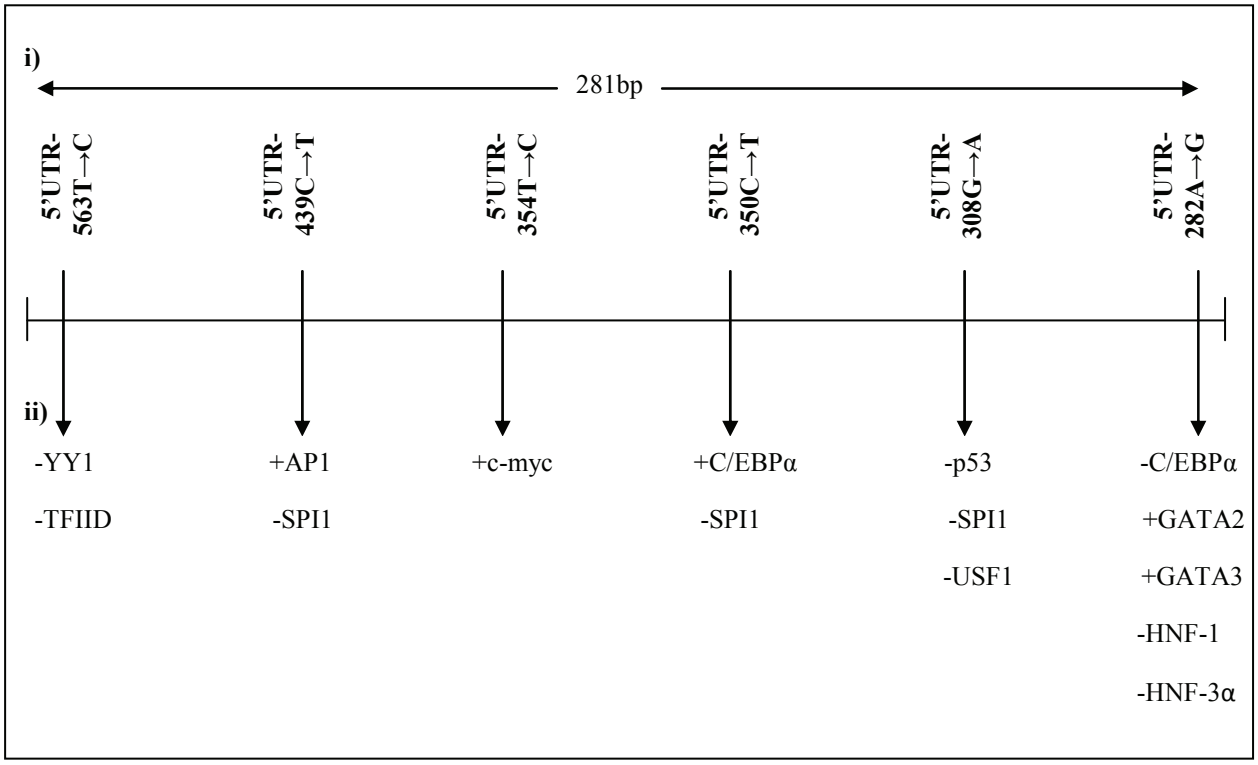
membrane of the enterocyte. This TF could be involved in simultaneously mediating expression of these two proteins.

A number of other putative TFBSs were predicted to be altered following *in silico* analysis of the variants identified in the *CP* promoter (Table 3.7). These include the binding motif for activator protein 1 (AP1) which has shown to be involved in cell proliferation, differentiation and tumourigenesis (Hilberg *et al.* 1993), upstream regulatory factor 1 (USF1) which has defined roles in regulation of genes expressed in the liver (Vallet *et al.* 1998) and spleen focus forming virus proviral integration oncogene 1 (SPI1) which plays a role in the development of myeloid and β-lymphoid cells which are important during an immune response (Tondravi *et al.* 1997). The functional importance of these TFs in regulation of *CP* expression should not be overlooked.

A haplotype, consisting of six of the promoter variants detected in this study, was predicted to exist following analysis using the Haploview 4.0 program (Figure 3.19). The variants comprising the haplotype span a genomic region of 281 bp that was identified as a critical promoter region in *R. norvegicus* (Fleming and Gitlin 1992). This critical region is highly conserved in humans and is thought to be involved in gene expression. This haplotype was shown to be significantly associated with OC (P=0.01) in this study. TFBSs that are disrupted by the variant alleles of this haplotype have already been discussed and are summarised in Figure 3.20. It is interesting to note that three of these haplotype variants abolish SPI1 TFBSs. SPI1 is known to play a role in regulation of genes involved in the inflammation response pathway (Tondravi *et al.* 1997). CP is an acute-phase protein which is up-regulated during times of cellular stress, and may therefore be controlled to a degree by this TF. Inadequate *CP* expression and an impaired immune response could potentially occur as a

result of disruptions to the binding motif for SPI1 caused by the variant alleles. However, no studies have demonstrated an involvement between *CP* and SPI1 to date.

Figure 3.20 Schematic representation of the predicted *CP* promoter haplotype and the TFBSs that are abolished or created in the presence of the variant alleles.



Legend to Figure 3.20

Schematic representation of the promoter haplotype predicted by Haploview 4.0. i) Gene annotation indicating the position of variants used as markers for haplotype assembly. Gene annotation is not drawn to scale and only indicates the positions of the variants. ii) Putative TFBSs that are created or abolished in the presence of the variant allele. -, TFBS abolished; +, TFBS created. **Abbreviations:** 5', 5 prime end; A, Adenine; AP-1, Activator protein 1; C, Cytosine; C/EBPα, CCAAT/enhancer binding protein alpha; c-myc, Myelocytomatosis viral oncogene c; G, Guanine; GATA1, GATA-binding protein 1; GATA2, GATA-binding protein 2; GATA3, GATA-binding protein 3; HNF-1, Hepatocyte nuclear factor 1; HNF-3α, Hepatocyte nuclear factor alpha; p53, Tumour suppressor protein 53; SPI1, spleen focus forming virus (SFFV) proviral integration oncogene 1; T, Thymine; TFBS, Transcription factor binding site; TFIID, Transcription factor IID; USF1, Upstream transcription factor 1; UTR, untranslated region; YY1, Ying Yang 1.

Coding region

The exonic variants identified in this study were all subjected to *in silico* analysis using the ESEfinder program (Cartegni *et al.* 2003) (see Table 3.8). Exonic splicing enhancers (ESEs) are elements that function as binding sites for serine/arginine-rich (SR) proteins in a *cis*-acting manner. SR proteins are responsible for defining exon/intron boundries by recruiting spliceosomal components. They bind to the ESE through their RNA-binding domain and recruit other essential components *via* protein-protein interactions and by antagonizing the action of splicing silencers in the vicinity (Cartegni *et al.* 2003). These SR proteins are therefore essential in the regulation of alternative splicing and exon inclusion.

Accurate splicing of pre-mRNAs requires multiple *cis* elements which include the splice sites and polypyrimidine tracts (Freyer *et al*. 1989, Zahler *et al.* 1992). ESEs have been identified in exons associated with regulated splicing adjacent to introns with weak intronic splicing signals (Watakabe *et al.* 1993). Disruption of these vital elements can result in exon skipping or the suppression of intron removal, leading to aberrant mRNA formation (Lejeune *et al*. 2001). There have been several human genetic diseases associated with defects in the alternative splicing mechanism in genes (Cáceres and Kornblihtt 2002, Zhang *et al.* 2004).

Only those variants that demonstrated statistical significance with OC as well as those that were detected exclusively in the OC patient cohort will be discussed further.

A synonymous variant, G633, was detected after HEX-SSCP analysis of exon 11 of the *CP* gene. This variant was shown to be statistically significantly (P=0.0004) associated with OC in this study. Synonymous SNPs in the coding exons of genes are traditionally regarded as

silent mutations. This is due to the fact that they do not result in a change in amino acid and subsequently the structure of the protein is not altered in any way. They can however have an effect on alternative splicing by inactivating or creating a splice site, interfering with other elements critical to the splicing mechanism, or by activating a cryptic splice site (Cartegni *et al.* 2003). These factors can result in the mutant exon being skipped by the splicing mechanism and the subsequent production of an aberrant protein.

The novel variant, G633, was the only exonic variant to demonstrate significant association with OC following statistical analysis (Table 3.2) in this study. Analysis of this variant using the ESEfinder program demonstrated that a SC35 binding motif is created. This could mean that improper splicing occurs in the presence of this variant. The results from a homology search of CP in different species showed that this glycine amino acid was conserved in all the species tested, except for the mouse (*Mus musculus*) where a tryptophan was indicated in this position. This could indicate a level of evolutionary conservation at this amino acid and changes at this position may therefore not be tolerated.

The G633 variant was identified in 13 of the 88 OC patients. The serum ferritin levels of the patients displaying the G633 variant differed markedly from each other. Seven of the patients showed normal SF levels [one male: 177.6 µg/l (reference range: 20-300 µg/l); six females: 100.6 µg/l, 141.3 µg/l, 169.24 µg/l, 124.3 µg/l, 171.8 µg/l and 97.5 µg/l (reference range: 20-200 µg/l)], three patients had elevated SF levels (one male: 409.9 µg/l and two females: 229.6 µg/l and 714.39 µg/l), one patient demonstrated iron deficiency (one female: 13.5 µg/l) and one patient had extremely elevated SF levels (one male: 2152 µg/l). It is important to note that at the time of OC diagnosis, the majority of patients may be suffering from malnutrition due to the location of the tumours in the upper region of the digestive tract and the

subsequent inability to ingest sufficient nutrients. As previously discussed, iron is absorbed predominately from the diet and the body iron levels of OC patients may be compromised as a result of insufficient iron uptake. The relationship between iron parameters, measured at this stage of the disease, and the effects of variations in the iron metabolising genes are therefore difficult to correlate.

It is possible that the G633 variant could be acting in conjunction with other variants identified in the *CP* gene, or other genes in the iron metabolism pathway, to produce the differences in iron levels observed in each of the patients. These data, along with the fact that the G633 variant was not detected in the control individuals, could indicate that this variant is associated with OC in the Black South African population. Further studies using the minigene system are warranted in order to determine what effect this variant has on protein expression in patients presenting with OC (Baralle *et al.* 2003).

Another synonymous variant, Y425 (RefSNP ID: rs34237139), was detected in exon 7 of the *CP* gene following HEX-SSCP analysis. It was identified in the heterozygous state in only one female OC patient who had moderately elevated SF levels [266.6 µg/l (reference range: 20-200 µg/l)]. Interestingly, this patient was also heterozygous for a number of other variants identified in this study (5'UTR-563T→C, 5'UTR-354T→C, 5'UTR-282A→G and IVS7+9T→C). It is possible that the combination of the variants could be playing a role in the iron overload demonstrated by this patient. ESEfinder results showed that the Y425 variant abolished a SRp40 binding motif, as well as creating a SC35 site (Table 3.8). It is therefore possible that this silent mutation could exert a negative effect on exon splicing, a hypothesis that should be researched further. The sequence alignment indicated that the tyrosine amino acid at position 425 is evolutionarily conserved amongst all the different

species utilised. Changes at this position may therefore not be tolerated. Statistical analysis did not show any significant results (P=0.328) (Table 3.2). These results could indicate that this amino acid is essential for protein function and that any type of disruption to ESE binding motifs may result in incorrect gene product function, and therefore lead to the iron overload phenotype observed in this patient.

A previously described intronic variant, IVS7+9T→C (RefSNP ID: rs35272481), was identified in the same patient as the Y425 variant. Two other previously identified (IVS4-14C→T and IVS15-12T→C) intronic variants were also detected in the non-coding region of the *CP* gene following HEX-SSCP analysis.

All genes are flanked by regions of non-coding DNA that were once considered to be genetic waste and subsequently termed "junk DNA". However, these intronic regions between exons are now considered to have an important role in transcriptional regulation. Wang *et al.* (2006) have demonstrated the importance of the role of non-coding RNA molecules in a variety of cellular processes including the regulation of the production of ribosomal RNA (rRNA). These non-coding RNAs were shown to originate from non-coding DNA regions called intergenic spaces (IGS). SNPs in the introns of genes have the ability to interfere with intron-exon splice sites, thereby creating variations in the splicing of certain genes. In eukaryotes, the accuracy of splicing is determined by the 5' splice site, the 3' splice site and the branch site (Cartegni *et al*. 2003). Variants that have an effect on the process of alternative splicing may not only be found in the regions of splice sites, but may also be found in regions that function as regulatory elements within exons or introns. Examples of these regions are termed enhancers or silencers (Pagani and Baralle 2004).

IVS7+9T→C was identified in a single patient with moderately elevated SF levels [one female: 266.6 µg/l (reference range: 20-200 µg/l)] in the heterozygous state. No significant (*P*=0.328) associations were observed following statistical analysis (Table 3.2) which indicates that this variant is possibly not involved in the development of iron overload or OC in the Black South African population. However, the patient with the IVS7+9T→C variant was also heterozygous for a number of other *CP* variants (5'UTR-563T→C, 5'UTR-354T→C, 5'UTR-282A→G and Y425) identified in this study. This could infer that the interplay between the different variants and their effect on the regulation of gene expression could be responsible for the iron overload observed in this patient. IVS7+9T→C could potentially disrupt a splice donor site. However, it lies 3 bp beyond the published consensus sequence for 5' (donor) splice sites (Padgett *et al.* 1986). Functional analysis of this variant using the minigene assay for intronic variants described by Cartegni *et al.* (2003) is necessary to determine if this variant is contributing to the disease phenotype.

A novel missense mutation, V246A, was detected in exon 4 of the *CP* gene following HEX-SSCP analysis. Non-synonymous SNPs in the coding regions of genes are called missense mutations and result in amino acid changes that subsequently lead to a protein that differs in its amino acid composition from that of the wild-type. Often this has no effect on the biological activity of the protein within the cell as many amino acid changes are tolerated. However, some amino acid substitutions, such as those at the active site of an enzyme, have a greater impact on the functionality of the protein. These changes can result in incorrect protein folding or protein-protein interactions that are compromised (de Pouplana *et al.* 1998).

The novel V246A variant results in an amino acid change from valine to alanine at position 246. As both of these amino acids are classified as non-polar hydrophobic residues, this variant is not expected to result in any structural changes within the protein. ESEfinder results showed that this variant is expected to create an SF2/ASF2 binding motif in the presence of the alanine amino acid, which may result in inaccurate exon splicing (Table 3.8). Protein homology comparison between species showed that the valine residue is conserved amongst the species. The macaque (*Macaca mulatta*) was the only one that demonstrated variation with an isoleucine in place of the valine. The V246A variant was detected only in the heterozygous state in two of the OC patients. Both individuals were female and had SF levels within the normal parameters [124.3 µg/l and 95.7 µg/l (reference range: 20-200 µg/l)]. No statistical significance (P=0.179) with OC was observed following χ^2 analysis (Table 3.2).

This evidence suggests that the novel V246A variant does not play a significant role in disease susceptibility. However, the fact that this variant is novel and that it was only detected in individuals presenting with OC, means that functional studies should be completed in order to conclusively exclude it from any association with OC in the Black South African population.

Variants that could possibly result in the dysregulation of CP may lead to the formation of an aberrant protein that is unable to oxidise Fe^{2+} efficiently. This would lead to an increase in Fe^{2+}, which is unable to bind to transferrin, in the serum. In addition, when CP levels are low or *CP* expression is inhibited, it is thought that a feedback mechanism may exist that prevents SLC40A1 from releasing Fe^{2+} from the cells. A resulting accumulation of iron within the intracellular compartment leads to oxidative damage of the cells. This damage is further exacerbated by the lack of CP, which in addition to acting as a ferroxidase enzyme, also

exhibits antioxidant properties. A study by Bosio *et al.* (2002) lends support to this hypothesis. In this study, a patient who was a compound heterozygote for two novel *CP* mutations demonstrated hepatic iron overload. An excess of iron has been previously described as a risk factor for the development of cancers, including OC. Variants detected in *CP* in patients presenting with OC should therefore be carefully considered with regard to disease pathogenesis. Further functional studies would provide vital information regarding the effect of these variants on gene expression.

CHAPTER 4

CONCLUSIONS AND FUTURE PROSPECTS

OC is one of the most common forms of cancer found in the South African population, particularly in the Transkei region. Like the majority of cancers, it demonstrates an extremely complicated aetiology, with both genetic and environmental factors involved in disease development and progression. Previous studies have attempted to identify possible risk factors for OC as well as candidate genes involved in disease development. However, contributing factors leading to OC are still poorly understood on a molecular level and many genetic aberrations that could infer disease risk or resistance remain to be identified. OC is insidious in onset, as symptoms often present late in the development of the disease. As a result, conventional cancer treatment is highly ineffectual in the majority of cases, leading to the increase in mortality rate observed in OC patients. Early detection and screening programmes are therefore essential if this disease is to be effectively treated and managed in the future. An increase in iron has previously been identified as a potential risk factor in the development of many forms of cancer, including OC. This study therefore attempted to investigate any possible associations with iron dysregulation and OC in the Black South African population.

The first objective of this study was to identify known and novel variants and/or polymorphisms within the *CP* gene, involved in the iron regulatory pathway, in patients with OC. This was achieved by PCR amplification of the regulatory, coding and flanking intronic regions of the *CP* gene, and subsequent analysis using HEX-SSCP analysis, RFLP analysis (where applicable) and semi-automated bidirectional DNA sequencing analysis.

A total of fourteen previously described (5'UTR-567C→G, 5'UTR-563T→C, 5'UTR-439C→T, 5'UTR-364delT, 5'UTR-354T→C, 5'UTR-350C→T, 5'UTR-282A→G, V223, Y425, R367C, D544E, IVS4-14C→T, IVS7+9T→C and IVS15-12T→C) and four novel

(5'UTR-308G→A, T83, V246A and G633) variants were identified, demonstrating that the first objective of this study was successfully achieved.

The samples that were screened in this study comprised gDNA from 96 OC patients and 88 population-matched, healthy control individuals. PCR amplification of the *CP* gene was successful in the majority of the samples. Restrictions on DNA quality and quantity however, meant that amplification optimisation for every sample was not successful for each fragment screened. This is reflected in the varying numbers of individuals (n) scored for each locus (see Tables 3.1 and 3.2). In addition, the PCR amplification of exon 13 was unsuccessful despite many attempts at optimisation. The primer set was redesigned using the Primer3 v0.2 software however amplification of this fragment was still not achieved. This could possibly be due to the presence of polymorphisms within the primer binding sites that could inhibit the primer adhering correctly to the target DNA. Future research should aim to optimise the PCR technique of this exon, as possible variants related to OC may be present in this region.

The size of the population screened in this study was sufficient to allow the detection of the aforementioned variants. In addition, allele and genotype frequencies of the variants detected were able to be calculated within the Black South African population. Screening of a larger patient cohort in the future could possibly allow for the detection of other previously described and novel variants not detected in this study. An increase in sample size may also aid in obtaining more statistically significant data.

The patient cohort demonstrated an equal distribution of males (n=48) and females (n=48). However, the population-matched control group was comprised of an unequal number of male and female individuals. This was unfortunate as it made it impossible to compare the prevalence of the variants identified between males and females with any accuracy.

A marked variation in iron parameters within the patient and control groups meant that it was extremely difficult to draw any substantial conclusions about the effect of the individual variants on iron regulation. This was further complicated by the fact that the iron levels of OC patients may be severely compromised as a result of malnutrition. The diagnosis of OC frequently occurs at a late stage in disease development. At this point, the size and location of the tumours of many patients may have already impeded their ability to ingest sufficient dietary nutrients, including iron. It would be beneficial for future studies to obtain the iron parameters of OC patients before severe malnutrition becomes a factor. However, this is a challenging task due to the complexity of OC symptoms and the subsequent delay in disease diagnosis.

HEX-SSCP analysis, RFLP analysis and semi-automated bidirectional DNA sequencing analysis were the detection methods employed in this study. With respect to the first objective of this study, to detect known and/or novel polymorphisms in the *CP* gene, these methods proved to be effective. However, previously described polymorphisms in this gene were not detected in this study. This occurrence could be due to a variety of factors, including the fact that these polymorphisms may not be present in the Black South African population, these variants do not contribute to OC susceptibility and are therefore not generally present in the OC population or the detection method of choice was not sensitive enough to identify these variants.

The lack of detection sensitivity was demonstrated during the screening of exon 2. In order to distinguish clearly between homozygous and heterozygous individuals for the T83 variant, RFLP analysis had to be employed. It is therefore possible that the HEX-SSCP technique was not sensitive enough to detect other variants that may be present in the *CP* gene in the patient and control cohorts used in this study.

The combination of single-strand conformation (SSCP) analysis with heteroduplex (HEX) methodology, as utilised in this study, offers the advantages of being inexpensive and relatively simple to use. It demonstrates 70-95% mutation detection sensitivity when DNA fragments smaller than 150-250 bp in size are analysed (Xiao and Oefner 1992, Hayashi and Yandell 1993), but is unable to detect large genomic rearrangements (Fujita and Silver 1994). It is however preferable to SSCP analysis alone, which is expected to demonstrate only 70% specificity in detecting single base pair changes (Sheffield 1993). The majority of the fragments analysed in this study were greater than 250 bp which could account for some previously described *CP* variants being undetected in this study. Sensitivity of this method is further influenced by the position of the variant within the screened fragment, pH, temperature and gel composition.

Further optimisation of this technique could contribute to future work involving the screening of a larger population. In addition, other detection techniques that demonstrate increased sensitivity could be employed. Denaturing high performance liquid chromatography (dHPLC) is currently considered to be superior to SSCP analysis as it demonstrates high SNP detection sensitivity (92-100%) and larger fragments (198-732 bp) are also able to be effectively analysed (Underhill *et al.* 1997). To date, the most accurate mutation detection method is considered to be semi-automated DNA sequencing which has the ability to detect all variants present within a DNA sequence (Kristensen *et al.* 2001). However, this technique

is relatively expensive and is therefore not considered to be effective for the initial screening of a population.

The second objective of this study was to determine if any significant association exists between the variants identified and OC. This was achieved using various statistical methods. The patient and control cohorts were tested for HWE at the variant loci. All groups were found to conform to HWE in this study (Tables 3.4 and 3.5). HWE is imperative for the accurate determination of the significance of the findings of the study.

Genotype and allele frequencies of the variants were determined by allele counting and were subsequently compared between the OC patients and the population-matched controls to determine any significant association with OC. Statistically significant association was observed for the novel 5'UTR-308G→A (P = 0.012) and G633 (P = 0.0004) variants identified in the *CP* promoter and coding region respectively. It is therefore possible that the presence of these variants may contribute to OC susceptibility. Further studies are required to elucidate the role of these potential disease-contributing mutations.

The third objective of this study involved the bioinformatic analysis of variants identified in this study using the appropriate *in silico* programs. *In silico* analysis of the region of the promoter spanning the identified variants sought to identify putative TFBSs that could possibly regulate the expression of *CP*. The programs employed for this analysis used different algorithms and criteria to establish if a TFBS is abolished or created in the presence of the variant allele.

Several TFBSs were found to be altered by the variants in the *CP* promoter region which could indicate changes in gene expression in patients with OC. *In silico* analysis using these particular programs is a predictive method only and provides a cost-free method of analysing regulatory regions. In no way does it serve to specify the presence of *in vivo* TFBSs absolutely. It is however an informative tool in the process of developing future functional studies to elucidate the function of identified variants.

With regard to future studies, transfection experiments using luciferase reporter constructs should be conducted for the promoter variants identified in this study. Particular attention should be paid to the variants in the 281 bp genomic region that constitutes the significant haplotype identified in this study (Figure 3.20). Luciferase experiments should be carried out to determine if this region is functionally important in driving *CP* expression by cloning normal promoters and those lacking the 281 bp region into suitable vectors and observing any difference in transcriptional activity. This information could enhance our understanding of the role that this gene plays in iron regulation in the Black South African population and may aid in establishing the *in vivo* functionality of these variants.

Bioinformatic analysis of the variants identified in the *CP* coding regions was conducted using the ESEfinder program. All of the novel variants identified in the exonic regions (T83, V246A and G633) as well as the previously described exonic variants (V233, R367C, Y425 and D544E) were predicted to affect splicing by altering the binding sites for various SR proteins. Therefore, the role of these variants in the development of OC should not be disregarded. However, this program does not present conclusive evidence for the functional impact that these variants may have on splicing. Functional studies should therefore be

performed in future to determine what role, if any, these variants play in exon splicing in the *CP* gene.

In order to further investigate the potential role that the regulation of iron homeostasis plays in the development of OC, other genes involved in this pathway should be studied. The coding regions of *HFE*, *HMOX1*, *SLC40A1*, *HAMP*, *CYBRD1* and *HJV* have already been screened for novel and/or previously described polymorphisms in the existing population (Human 2007). Analysis of the promoter regions of these genes is currently being performed in our laboratory. All of these data combined could contribute significantly to our knowledge of how iron overload acts as a risk factor for OC, and may serve to identify possible modifier genes for this complex disease. Future research should also be extended to include the White and Coloured populations. Data accumulated from these genetically distinct South African population groups will no doubt further our understanding of iron dysregulation and OC pathophysiology.

It is also important that future researchers aim to collect as much clinical information on the patients and controls as possible (ie: smoking status, alcohol consumption, nutrition and workplace) to enable more in-depth analysis of the effects of the variants on OC aetiology. It will also be imperative in the future to compare iron parameters (serum ferritin and transferrin saturation) between patients and control individuals to try and elucidate the effect (if any) that these variants have on iron regulation.

Novel and previously described variants in the Black South African OC population were successfully identified in this study. Statistically significant associations between two of the novel variants were observed with OC. Although no other statistically significant associations

were observed with any of the other variants identified in this study, the role that they may play in the aetiology of OC should not be disregarded. The effect of various combinations of variants in *CP*, and other regulatory genes, could potentially be a factor in disease development. Future research may enhance our knowledge of how these variants contribute to OC pathogenesis as well as providing a greater understanding of how dysregulation of the iron metabolism pathway can act as a risk factor for this disease. This study constitutes an integral part of a much larger long-term analysis of the correlation between aberrations in iron regulatory genes and disease. The data generated from this study, in conjunction with that of the larger study, may potentially lead to the identification of molecular markers for OC. These markers could ultimately serve as an effective tool in the pre-symptomatic diagnosis and subsequent treatment of this disease.

CHAPTER 5

REFERENCES

5.1 GENERAL REFERENCES

Abnet CC, Borkowf CB, QiaoY-L, Albert PS, Wang E, Merrill Jr AH, Mark SD, Dong Z-W, Taylor PR, Dawsey SM (2001) Shingolipids as biomarkers of fumonisin exposure and risk of esophageal squamous cell carcinoma in China. Cancer Causes and Control **12**: 821-828

Adams PC and Chakrabarti S (1998) Genotypic/phenotypic correlation in genetic haemochromatosis: evolution of diagnostic criteria. Gastroenterology **114**: 319

Adelaide J, Monges G, Derderian C, Seitz JF, Birnbaum D (1995) Oesophageal cancer and amplification of the human cyclin D gene CCND1/PRAD1. Brit J Cancer **71**: 64-68

Aisen P, Enns C, Wessling-Resnick M (2001) Chemistry and biology of eukaryotic iron metabolism. Int J Biochem Cell Biol **33**: 940-959

Ala A, Walker AP, Ashkan K, Dooley JS, Schilsky ML (2007) Wilson's disease. Lancet **369**: 397-408

Amer MH, El-Yazigi A, Hannan MA, Mohamed ME (1990) Water contamination and esophageal cancer at Gassim Region, Saudi Arabia. Gastroenterology **98**: 1141-1147

Anderson GJ, Frazer DM, McKie AT, Vulpe CD, Smith A (2005) Mechanisms of haem and non-haem iron absorption: lessons from inherited disorders of iron metabolism. Biometals **18**: 339-348

Andrews NC (1999) Disorders of iron metabolism. N Engl J Med **341**: 1986-1995

Andrews NC (2008) Forging a field: The golden age of iron biology. Blood **112**: 219-230

Andrews NC Fleming MD, Levy JE (1999) Molecular insights into mechanisms of iron transport. Curr Opin Hematol **6**: 61-64

Baralle M, Baralle D, De Conti L, Mattocks C, Whittaker J, Knezevich A, French-Constant C, Baralle FE (2003) Identification of a mutation that perturbs *NF1* gene splicing using genomic DNA samples and a minigene assay. J Med Genet **40**: 220-222

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics **21**: 263-265

Barrett MT, Galipeau PC, Sanchez CA, Emond MJ, Reid BJ (1996) Determination of the frequency of loss of heterozygosity in esophageal adenocarcinoma by cell sorting, whole genome amplification and microsatellite polymorphisms. Oncogene **12**: 1873-1878

Blot WJ (1994) Esophageal cancer trends and risk factors. Semin Oncol **21**: 403-410

Boll M-C, Sotelo J, Otero E, Alcaraz-Zubeldia M, Rios C (1999) Reduced ferroxidase activity in cerebrospinal fluid from patients with Parkinson's disease. Neurosci Lett **265**: 155-158

Bosio S, De Gobbi M, Roetto A, Zecchina, Leonardo E, Rizzetto M, Lucetti C, Petrozzi U, Binyccelli U, Camaschella C (2002) Anemia and iron overload due to compound heterozygosity for novel ceruloplasmin mutations. Blood **100**: 2246-2248

Bothwell TH, Seftel HC, Jacobs P, Torrance JD, Baumslag N (1964) Iron overload in Bantu subjects. Studies on the availability of iron in Bantu beer. Am J Clin Nutr **14**: 47-51

Bottomley SS, May BK, Cox TC, Cotter PD, Bishop DF (1995) Molecular defects of erythroid 5-aminolevulinate synthase in X-linked sideroblastic anemia. J Bioenerg Biomembranes **27**: 161-168

Boz A, Evliyaoğlu O, Yildirim M, Erkan N, Karaca B (2005) The value of serum zinc, copper, ceruloplasmin levels in patients with gastrointestinal tract cancers. Turk J Gastroenterol **16(2)**: 81-84

Bradshaw E, Schonland M (1969) Oesophageal and lung cancers in Natal African males in relation to certain socio-economic factors. An analysis of 484 interviews. Br J Cancer **23**: 275-284

Brittenham GM, Weiss G, Brissot P, Lainé F, Guillygomarc'h A, Guyader D, Moirand R, Deugnier Y (2000) Clinical consequences of new insights in the pathophysiology of disorders of iron and heme metabolism. Hematology Am Soc Hematol Educ Program: 39-50

Buckland PR (2004) Allele-specific gene expression differences in humans. Hum Mol Genet **13**: 255-260

Butler JEF, Kadonaga JT (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev **16**: 2583-2592

Cáceres JF, Kornblihtt AR (2002) Alternative splicing: multiple control mechanisms and involvement in human disease. Trends Genet **18**: 186-193

Cai L, Mu L-N, Lu H, Lu Q-Y, Yuko You N-C, Yu S-Z, Le AD, Zhao J, Zhou X-F, Marshall J, Heber D, Zhang Z-F (2006) Dietary selenium intake and genetic polymorphisms of the *GSTP1* and *p53* genes on the risk of esophageal squamous cell carcinoma. Cancer Epidemiol Biomarkers Prev **15**: 294-300

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR (2003) ESEfinder: a web resousrc to identify exonic splicing enhancers. Nucleic Acids Res **31**: 3568-3571

Chen PL, Chen YM, Bookstein R, Lee WH (1990) Genetic mechanisms of tumour suppression by the human p53 gene. Science **250**: 1576-1580

Chen X, Yang GY, Ding YW, Bondoc F, Curtis SK, Yang CS (1999) An esophagogastroduodenal anastomosis model for esophageal adenocarcinogenesis in rats and enhancement by iron overload. Carcinogenesis **20**: 1801-1808

Chen X, Ding YW, Yang G-Y, Bondoc F, Lee M-J, Yang CS (2000) Oxidative damage in an esophageal adenocarcinoma model with rats. Carcinogenesis **21**: 257-263

Chen X, Yang CS (2001) Esophogeal adenocarcinoma: a review and perspectives on the mechanism of carcinogenesis and chemoprevention. Carcinogenesis **22**(8): 1119-1129

Cherukuri S, Potla R, Sarkar J, Nurko S, Leah Harris Z, Fox PL (2005) Unexpected role of ceruloplasmin in intestinal iron absorption. Cell Metabolism **2**: 309-319

Cho Y, Gorina S, Jeffrey PD, Pavletich NP (1994) Crystal structure of a p53 tumour suppressor-DNA complex: understanding tumorigenic mutations. Science **265**: 346-355

Chow WH, Blot WJ, Vaughan TL, Risch HA, Gammon MD, Stanford JL, Dubrow R, Schoenberg JB, Mayne ST, Farrow DC, Ahsan H, West AB, Rotterdam H, Niwa S, Fraumeni Jr JF (1998) Body mass index and risk of adenocarcinomasof the esophagus and gastric cardia. J Natl Cancer Inst **90**: 150-155

Connor JR, Tucker P, Johnson M, Snyder B (1993) Ceruloplasmin levels in the human superior temporal gyrus in aging and Alzheimer's disease. Neurosci Lett **159**: 88-90

Conrad ME, Umbreit JN, Moore EG (1999) Iron absorption and transport. Am J Med Sci **318**: 213-229

Cooper DN, Smith BA, Cooke HJ, Niemann S, Schmidtke J (1985) An estimate of unique DNA sequence heterozygosity in the human genome. Hum Genet **69**: 201–205

Cooper GM, Sidow A (2003) Genomic regulatory regions: insights from comparative sequence analysis. Curr Opin Genet Dev **13**: 1-7

Costa RH, Kalinichenko VV, Holterman A-XI, Wang X (2003) Transcription factors in liver development, differentiation and regeneration. Hepatology **38**: 1331-1347

Crespi M, Bogomoletz VW, Munoz N (1994) Cancer of the esophagus. Gastroenterol Int **7**: 24-35

Daly JM, Fry WA, Little AG (2000) Esophageal cancer: results of an American College of Surgeons Patient Care Evaluation Study. J Am Coll Surg **190**: 562-572

Danks DM (1989) Hereditary disorders of copper metabolism in Wilson's disease and Menke's disease. The Metabolic basis of inherited disease 5th ed.: 1251-1268. New York : McGraw-Hill

Davies M, Parry JE, Sutcliffe RG (1981) Examination of different preparations of human placental plasma membrane for the binding of insulin, transferrin and immunoglobins. J Reprod Fertil **63**: 315-324

Day GL, Blot WJ, Shore RE, McLauglin JK, Austin DF, Greenberg RS, Liff JM, Preston-Martin S, Sarkar S, Schoenberg JB, Fraumeni JF (1994) Second cancers following viral and pharyngeal cancers: role of tobacco and alcohol. J Natl Cancer Inst **86**: 131-137

De Pouplana LR, Turner RJ, Steer BA, Schimmel P (1998) Genetic code origins: tRNAs older than their synthetases? Proc Natl Acad Sci **95**: 11295

Deiss A (1983) Iron metabolism in reticuloendothelial cells. Semin Hematol **20**: 81-90

Donovan A, Brownlie A, Zhou Y, Shepard J, Pratt SJ, Moynihan J, Paw BH, Drejer A, Barut B, Zapata Z, Law TC, Brugnara C, Lux SE, Pinkus GS, Pinkus JL, Kingsley PD, Palis J, Fleming MD, Andrews MC, Zon LI (2000) Positional cloning of zebrafish ferroportin 1 identifies a conserved vertebrate iron exporter. Nature **403**: 776-781

Dorner MH, Abel U, Fritze D, Manke H-G, Drinos P (1983) Serum ferritin in relation to the course of Hodgkin's diseases. Cancer **52**: 2308-2312

Doustjalali SR, Yusof R, Govindasamy GK, Bustam AZ, Pillay B, Hashim OH (2006) Patients with nasopharyngeal carcinoma demonstrate enhanced serum and tissue ceruloplasmin expression. J Med Invest **53**: 20-28

Ellis A, Field JK, Field EA (1994) Tylosis associated with carcinoma of the oesophagus and oral leukoplakia in a large Liverpool family – a review of six generations. Eur J Cancer B Oral Oncol **30B**: 102-112

Emerit J, Beaumont C, Trivin F (2001) Iron metabolism, free radicals and oxidative injury. Biomed Pharmacother **55**: 333-339

Eng C, Spechler SJ, Ruben R, Li FP (1993) Familial Barrett esophagus and adenocarcinoma of the gastroesophageal junction. Cancer Epidemiol Biomarkers Prev **2**: 397-399

Enzinger PC, Mayer RJ (2003) Esophageal cancer. N Engl J Med **349**: 2241-2252

Farrow DC, Vaughan TL, Sweeney C (2000) Gastroesophageal reflux disease, use of H2 receptor antagonists, and risk of esophageal and gastric cancer. Cancer Causes Control **11**: 231-238

Feder JN, Tsuchihashi Z, Irrinki A, Lee VK, Mapa FA, Morikang E, Prass CE, Starnes SM, Wolff RK, Parkkila S, Sly WS, Schatzman RC (1997) The haemochromatosis founder mutation in HLA-H disrupts β_2-microglobulin interaction and cell surface expression. J Biol Chem **272**: 14025-14028

Freyer G A, O'Brien JP, Hurwitz J **(**1989) Alterations in the polypyrimidine sequence affect the in vitro splicing reactions catalyzed by HeLa cell-free preparations. J Biol Chem **264**: 14631–14637

Fein R, Kelsen DP, Geller N, Bains M, McCormack P, Brennan MF (1985) Adenocarcinoma of the esophagus and gastroesophageal junction. Prognostic factors and results of therapy. Cancer **56**: 2512-2518

Fernandes ML, Seow A, Chan Y-H, Ho K-Y (2006) Opposing trends in incidence of esophageal squamous cell carcinoma and adenocarcinoma in a multi-ethnic Asian country. Am J Gastroenterol **101**: 430-1436

Ferreira R, Ohneda K, Yamamoto M, Philipsen S (2005) GATA1 function, a paradigm for transcription factors in haematopoiesis. Mol Cell Biol **25**: 1215-1227

Finch C (1994) Regulators of iron balance in humans. Blood **84**: 1697-1702

Fleming MD, Trenor CC III, Su MA, Foernzler D, Beier DR, Dietrich WF, Andrews NC (1997) Microcytic anaemia mice have a mutation in Nramp2, a candidate iron transporter gene. Nat Genet **16**: 383-386

Fleming MD, Romano MA, Su MA, Garrick LM, Garrick MD, Andrews NC (1998) Nramp2 is mutated in the anemic Belgrade b rat: evidence of a role for Nramp2 in endosomal iron transport. Proc Natl Acad Sci USA **95**: 1148-1153

Fleming RE, Gitlin JD (1992) Structural and functional analysis of the 5'-flanking region of the rat ceruloplasmin gene. Biol Chem **267**: 479-486

Fujita K and Silver J (1994) Single-strand conformational polymorphism. PCR Methods Appl **4**: 137-140

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science **296**: 2225-2229

Ganz T (2003) Hepcidin, a key regulator of iron metabolism and mediator of anemia of inflammation. Blood **102**: 783-788

Ganz T and Nemeth E (2006) Regulation of iron acquisition and iron distribution in mammals. Biochem Biophys Acta **1763**: 690-699

Goldblum JR (2003) Barrett's esophagus and Barrett's-related dysplasia. Mod Pathol **16**(4): 316-324

Goldstein SR, Yang GY, Chen X, Curtis SK, Yang CS (1998) Studies of iron deposits, inducible nitric oxide synthase and nitrotyrosine in a rat model for esophageal adenocarcinoma. Carcinogenesis **19**: 145-1449

Greene FL, Page GL, Fleming ID (2002) AJCC cancer staging manual. 6th ed. New York: Springer-Verlag.
Gunshin H, Mackenzie B, Berger UV, Gunshin Y, Romero MF, Boron WF, Nussberger S, Gollan JL, Hediger MA (1997) Cloning and characterisation of a mammalian proton-coupled metal-ion transporter. Nature **388**: 482-488

Guo B, Yu Y, Leibold EA (1994) Iron regulates cytoplasmic levels of a novel iron-responsive element-binding protein without aconitase activity. J Biol Chem **269**: 24252-24260

Guo Y, Jamison DC (2005) The distribution of SNPs in human regulatory regions. BMC Genomics **6**: 140

Gutteridge JMC (1992) Iron and oxygen free radicals in the brain. Ann Neurol **32**: S16-S21

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser **41**: 95-98

Hallendorff MA (2008) Ironing out Haemochromatosis: a study of an Indian family. March 2008. University of Stellenbosch.

Hann HW, Stahlhut MW, Hann CL (1990) Effect of iron and desferoxamine on cell growth and in vitro ferritin synthesis in human hepatoma cell lines. Hepatology **11**: 566-569

Hann H-W L, Stahlhut MW, Rubin R, Maddrey WC (1992) Antitumor effect of deferoxamine on human hepatocellular carcinoma growing in athymic nude mice. Cancer **70**: 2051-2056

Harris LZ, Takahashi Y, Miyajima H, Serizawa M, MacGillivray RTA, Gitlin J (1995) Aceruloplaminemia: Molecular characterization of this disorder of iron metabolism. Proc Natl Acad Sci USA **9**: 2539-2543

Harris ZL, Klomp LW, Gitlin JD (1998) Aceruloplasminemia: an inherited neurodegenerative disease with impairment of iron homeostasis. Am J Clin Nutr **67**: 972-977

Harris ZL, Durley AP, Man TK, Gitlin JD (1999) Targeted gene disruption reveals an essential role for ceruloplasmin in cellular iron efflux. Proc Natl Acad Sci USA **96**: 10812–10817

Hayashi K, Yandell DW (1993) How sensitive is PCR-SSCP? Hum Mutat **2**: 338-346

Hentze MW, Kuhn LC (1996) Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. Proc Natl Acad Sci USA **93**: 8175-8182

Hilberg F, Aguzzi A, Howells N, Wagner EF (1993) c-Jun is essential for normal mouse development and hepatogenesis. Nature **365**: 179-181

Hoefkens P, Smit MH, deJe-Jaspars NM, Hijskes-Heins MI, deJong G, vanEijk HG (1996) Isolations, renaturation and partial characterization of recombinant human transferrin and its half molecules form Escherichia coli. Int J Biochem Cell Biol **28**: 975-982

Hollstein M, Shormer B, Greenblatt M, Soussi T, Hovig E, Montesano R, Harris CC (1996) Somatic point mutations in the p53 gene of human tumors and cell lines. Nucleic Acids Res **24**: 141-146

Hsing AW, McLaughlin JK, Olsen JH, Mellemkjar L, Wacholder S, Fraumeni JF Jr (1995) Cancer risk following primary hemochromatosis: a population-based cohort study in Denmark. Int J Cancer **60**: 160-162

Huang HJS, Yee JK, Shew JY, Chen PL, Bookstein R, Freidmann T, Lee EYHP, Lee WH (1988) Suppression of the neoplastic phenotype by replacement of the RB gene in human cancer cells. Science **242**: 1563-1566

Huang Y, Boynton RF, Blount PL, Silverstein RJ, Yin J, Tong Y, McDaniel TK, Newkirk C, Resau JH, Sridhara R, Reid BJ, Meltzer SJ (1992) Loss of heterozygosity involves multiple tumor suppressor genes in human esophageal cancers. Cancer Res **54**: 6094-6096

Huang Y, Meltzer SJ, Yin J, Tong Y, Chang EH, Srivastava S, McDaniel T, Boynton RF, Zou ZQ (1993) Altered messenger RNA and unique mutational profiles of p53 and Rb in human esophageal carcinomas. Cancer Res **53**: 1889-1894

Human VR (2007) Molecular analysis of genes involved in iron overload implicated in oesophageal cancer. March 2007. University of Stellenbosch.

Huot AE, Gundel RM, Hacker MP (1990) Effect of erythrocytes on alveolar macrophage cytostatic activity induced by bleomycin lung damage in rats. Cancer Res **50**: 2351-2355

Isaacson C, Bothwell TH, MacPhail AP, Simon M (1985) The iron status of urban Black subjects with carcinoma of the oesophagus. S Afr Med J **67**: 591-593

Kakushima N, Fujishiro M (2008) Endoscopic submucosal dissection for gastrointestinal neoplasms. World J Gastroenterol **14**: 2962-2967

Kastan MB, Onyekwere O, Sidransky D, Vogelstein B, Craig RW (1991) Participation of p53 protein in the cellular response to DNA damage. Cancer Res **51**: 6304-6311

Kawabata H, Yang R, Hirama T, Vuong PT, Kawano S, Gombart AF, Koeffler HP (1999) Molecular cloning of transferrin receptor 2. A new member of the transferrin receptor-like family. J Biol Chem **274**: 20826-20832

Ke Y, Wu J, Leibold EA, Walden WE, Theil EC (1998) Loops and bulge/loops in iron-responsive element isoforms influence iron regulatory protein binding. Fine-tuning of mRNA regulation. J Biol Chem **273**: 23637-23640

Kel A, Kel-Margoulis O, Borlak J, Tchekmenev D, Wingender E (2005) Databases and tools for *in silico* analysis of regulation of gene expression. Handbook of Toxiogenomics: Strategies and Applications. Ed J Borlak, WILEY-VCH p253-290

Kotze MJ, Theart L, Callis M, Peeters A, Thiart R, Langenhoven E (1995) Nonradioactive multiplex PCR screening strategy for the simultaneous detection of multiple low-density lipoprotein receptor gene mutations. PCR Methods Applic **4**: 352-356

Kristensen VN, Kelefiotis D, Kristensen T, Borrensen-Dale AL (2001) High-throughput methods for detection of genetic variation. Biotechniques **30**: 318-322

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet **22**: 139-144

Lam AKY (2000) Molecular biology of esophageal squamous cell carcinoma. Crit Rev Oncol Hematol **33**: 71-90

Lagergren J, Bergstr□m R, Lindgren A, Nyrén O (1999) Symptomatic gastoesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med **340**: 825-831

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, *et al*. (2001) Initial sequencing and analysis of the human genome. Nature **409**: 860-921

Larsen B (1977) On linkage relations of ceruloplasmin polymorphism (Cp) in cattle. Anim Genet **8**: 111-113

Lasko D, Cavenee W, Nordenskjold M (1991) Loss of constitutional heterozygosity in human cancer. Ann Rev Genet **25**: 281-314

Lejeune F, Cavaloc Y, Stevenin J (2001) Alternative splicing of intron 3 of the serine/arginine-rich protein 9G8 gene: identification of flanking exonic splicing enhancers and involvement of 9G8 as a trans-acting factor. J Biol Chem **276**: 7850-7858

Lettice LA, Horikoshi T, Heaney SJH, van Baren MJ, van der Linde HC, Breedveld GJ (2002) Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc Natl Acad Sci USA **99**: 7548-7553

Levine MS, Chu P, Furth EE, Rubesin SE, Laufer I, Herlinger H (1997) Carcinoma of the esophagus and esophagogastric junction: sensitivity of radiographic diagnosis. Am J Roentgenol **168**: 1423-1426

Li Q, Johnston SA (2001) Is all DNA binding and transcription regulation by an activator physiologically relevant? Mol Cell Biol **21**: 2467-2474

Lieu P, Heiskala M, Peterson PA, Yany Y (2001) The roles of iron in health and disease. Mol Asp Med **22**: 1-87

Lijinsky W, Reuber MD, Riggs CW (1981) Dose response studies of carcinogenesis in rats by nitrosodiethylamine. Cancer Res **41**: 4997-5003

Lim CW, Parker HM, Versonder RF, Haschek WM (1996) Intravenous fumonisin B1 induces cell proliferation and apoptosis in the rat. Nat Toxins **4**: 34-41

Linder MC, Moor JR, Wright K (1981) Ceruloplasmin assays in diagnosis and treatment of human lung, breast, and gastrointestinal cancers. J. Natl. Cancer Inst. **67**: 263–275

Lock K, Pomerleau, J, Causer L, McKee M. (2004) Low fruit and vegetable consumption. In: Ezzati M, Lopez AD, Rodgers A, Murray CJ. Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors. Geneva: World Health Organization. 597–728.

Loeffler DA, Connor JR, Juneau PL, Snyder BS, Kanaley L, DeMaggio AJ, Nguyen H, Brickman CM, LeWitt PA (1995) Transferrin and iron in normal, Alzheimer's disease, and Parkinson's disease brain regions. J Neurochem **65**: 710-716

Looker AC and Johnson CL (1988) Prevalence of elevated serum transferrin saturation in adults in the United States. Ann Int Med **129**: 940-945

Lu SH, Hsieh LL, Luo FC, Weinstein IB (1988) Amplification of the EGF receptor and *c-myc* genes in human esophageal cancers. Int J Cancer **42**: 502-505

Lu SH (2000) Alterations of oncogenes and tumor suppressor genes in esophageal cancer in China. Mut Res **462**: 343-353

MacPhail AP, Simon MO, Torrance JD, Charlton RW, Bothwell TH, Isaacson C (1979) Changing patterns of dietary iron overload in Black South Africans. Am J Clin Nutr **32**: 1272-1278

Maesawa C, Tamura G, Suzuki Y, Ogasawara S, Ishida K, Saito K, Satodate R (1994) Aberrations of tumor-suppressor genes (*p53, apc, mcc and Rb*) in esophageal squamous-cell carcinoma. Int J Cancer **57**: 21-25

Majuri R, Grasbeck R (1987) A rosette receptor assay with haem-microbeads. Demonstration of a haem receptor on K562 cells. Scand J Haemotol **38**: 21-25

Makaula AN, Marasas WF, Venter FS, Badenhorst CJ, Bradshaw D, Swanevelder S (1996) Oesophageal and other cancer patterns in four selected districts of the Transkei, Southern Africa: 1985-1990. Afr J Health Sci **3**: 11-15

Mandishona E, MacPhail AP, Gordeuk VR, Kedda MA, Paterson AC, Rouault TA, Kew MC (1998) Dietary iron overload as a risk factor for hepatocellular carcinoma in Black Africans. Hepatology **27**: 1563-1566

Marasas WFO, van Rensburg SJ, Mirocha CJ (1979) Incidence of Fusarium species and the mycotoxins, deoxynivalenol and zearalenone, in corn produced in esophageal cancer areas in Transkei. J Agric Food Chem **27**: 1108

Marasas WF, Jaskiewicz K, Venter FS, Van Schalkwyk DJ (1988) Fusarium moniliforme contamination of maize in oesophageal cancer areas in Transkei. S Afr Med J **74**: 110-114

Matsha T, Erasmus R, Kafuko AB, Mugwanya D, Stepien A, Parker MI (2002) Human papillomavirus associated with oesophageal cancer. J Clin Pathol **55**: 587-590

Mayne ST, Risch HA, Dubrow R, Chow W-H, Gammom MD, Vaughan TL, Farrow DC, Schoenberg JB, Stanford JL, Ahsan H, West AB, Rotterdam H, Blot WJ, Fraumeni JF (2001) Nutrient intake and risk of subtypes of esophageal and gastric cancer. Cancer Epidemiol Biomarkers Prev **10**: 1055-1062

McKie AT, Marciani P, Rolfs A, Brennan K, Wehr K, Barrow D, Miret S, Bomford A, Peters TJ, Farzaneh F, Hediger MA, Hentze MW, Simpson RJ (2000) A novel duodenal iron-regulated transporter, *IREG1*, implicated in the basolateral transfer of iron to the circulation. Mol Cell **5**: 299-309

Miller MP (1997) Tools for population genetic analysis (TFPGA) 1.3. A windows program for the analysis of allozyme and molecular population genetic data.

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucl Acids Res **16**(3): 1215

Mohebbi M, Mahmoodi M, Wolfe R, Nourijelyani K, Mohammad K, Zeraati H, Fotouhi A (2008) Geographical spread of gastrointestinal tract cancer incidence in the Caspian Sea region of Iran: Spatial analysis of cancer registry data. BMC Cancer **8**(137): 1-12

Montgomery SB, Griffith OL, Schuetz JM, Brooks-Wilson A, Jones SJM (2007) A survey of genomic properties for the detection of regulatory elements. Plos Comput Biol **3**: 106

Mukhopadhyay CK, Attieh ZK, Fox PL (1998) Role of ceruloplasmin in cellular iron uptake. Science **279**: 714-717

Myburg RB, Dutton MF, Chuturgoon AA (2002) Cytotoxicity of fumonisin B_1, diethylnitrosamine, and catechol on the SNO esophageal cancer cell line. Environ Health Pers **110**: 813-815

Naylor SL, Yang F, Cutshaw S, Barnett DR, Bowman BH (1985) Mapping ceruloplasmin cDNA to human chromosome 3. Cytogenet Cell Genet **40**: 711

Nemeth E, Rivera S, Gabayan V, Keller C, Taudorf S, Pedersen BK, Ganz T (2004) IL-6 mediates hypoferremia of inflammation by inducing the synthesis of iron regulatory hormone hepcidin. J Clin Invest **113**: 1271-1276

Nicolas G, Bennoun M, Devaux I, Beaumont C, Grandchamp B, Kahn A, Vaulont S (2001) Lack of hepcidin gene expression and severe tissue iron overload in upstream stimulatory factor 2 (USF2) knockout mice. Proc Natl Acad Sci USA **98**: 8780-8785

Nicolas G, Chauvet C, Viatte L, Danan JL, Bigard X, Devaux I, Beaumont C, Kahn A, Vaulont S (2002) The gene coding the iron regulatory peptide hepcidin is regulated by anemia, hypoxia and inflammation. J Clin Invest **110**: 1037-1044

Oates PS, Thomas C, Freitas E, Callow MJ, Morgan EH (2000) Gene expression of divalent metal transporter 1 and transferrin receptor in duodenum of Belgrade rats. Am J Physiol Gastrointest Liver Physiol **278**: G930-G936

O'Connell M, Halliwell J, Moorehouse B, Aruorna R, Baum OI, Peters TJ (1986) Formation of hydroxyl radicals in the presence of ferritin and haemosiderin. Is haemosiderin formation a biological protective mechanism? Biochem J **234**: 727-731

Ojala K, Sorri M, Jokinen K, Kairaluoma M (1982) Symptoms of carcinoma of the oesophagus. Med J Aust **1**: 384-385

Osaki S, Johnson DA, Frieden E (1966) The possible significance of ferrous oxidase activity of ceruloplasmin in normal human serum. J Biol Chem **241**(12): 2746-2751

Padgett R, Grabowski P, Konarska M, Seilers S, Sharp P (1986) Splicing of messenger RNA precursors. Annu Rev Biochem **55**: 1119-1150

Pagani F, Baralle FE (2004) Opinion: Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet **5**: 389-396

Pantopoulos K (2005) Regulation of Iron Metabolism. In: Weiss G VRG, Hershko C (eds) Anemia of chronic disease, Vol 30. CRC Press, Taylor & Francis Group, Boca Raton, p 1

Papanikolaou G, Pantopoulos K (2005) Iron metabolism and toxicity. Toxicol Appl Pharmacol **202**: 199-211

Parent M-E, Siemiatycki J, Fritschi L (2000) Workplace exposures and oesophageal cancer. Occup Environ Med **57**: 325-334

Parkin DM, Pisani P, Lopez AD, Masuyer E (1994) At least one in seven cases of cancer is caused by smoking. Global estimates for 1985. Int J Cancer **59**: 494-504.

Parkin DM, Bray FI, Devessa SS (2001) Cancer burden in the year 2000. The global picture. Eur J Cancer **37**: 4-66

Parkin DM, Bray F, Ferlay J, Pisani P (2005) Global cancer statistics, 2002. CA Cancer J Clin **55**: 74-108

Parkkila S, Waheed A, Britton RS, Bacon BR, Zhou XY, Tomatsu S, Fleming RE, Sly WS (1997) Association of the transferrin receptor in human placenta with HFE, the protein defective in hereditary haemochromatosis. Proc Natl Acad Sci USA **94**: 13198-13202

Patel BN, Dunn RJ, Jeong SY, Zhu Q, Julien J-P, David S (2002) Ceruloplasmin regulates iron levels in the CNS and prevents free radical injury. J Neurosci **22**(15): 6578-6586

Patel BN, David S (1997) A novel glycosylphosphatidylinositol-anchored form of ceruloplasmin is expressed by mammalian astrocytes. J Biol Chem **272**: 20185-20190

Paull A, Trier JS, Dalton MD, Camp RC, Loeb P, Goyal RK (1976) The histologic spectrim of Barrett's esophagus. N Engl J Med **295**: 476-480

Pietrangelo A (2004) Non-HFE haemochromatosis. Hepatology **39**: 21-29

Pisani P, Parkin DM, Bray F, Ferlay J (1999) Estimates of the worldwide mortality from 25 cancers in 1990. Int J Cancer **83**: 870-873

Pousset D, Piller V, Bureaud N, Piller F (2001) High levels of ceruloplasmin in the serum of transgenic mice developing hepatocellular carcinoma. Eur J Biochem **268**: 1491-1499

Purchase IF, Tustin RC, van Rensburg SJ (1975) Biological testing of food grown in the Transkei. Food Cosmet Toxicol **13**: 639-647

Qadri I, Iwahashi M, Kullak-Ublick GA, Simon FR (2006) Hepatocyte nuclear factor (HNF) 1 and HNF4 mediate hepatic multidrug resistance protein 2 up-regulation during hepatitis C virus gene expression. Mol Pharmacol **70**: 627-636

Qui A, Jansen M, Sakaris A, Hee Min S, Chattopadhyay S, Tsai E, Sandoval R, Akabas MH, Goldman ID (2006) Identification of an intestinal folate transporter and the molecular basis for hereditary folate malabsorption. Cell **127**: 917-928

Ragan HA, Nacht S, Lee GR, Bishop CR, Cartwright GE (1969) Effect of ceruloplasmin om plasma iron in copper-deficient swine. Am J Physiol **217**: 1320-1323

Rice EW (1961) Evaluation of the role of ceruloplasmin as an acute-phase reactant. Clin. Chim. Acta. **6**: 652–655

Risk JM, Mills HS, Garde J (1999) The tylosis esophageal cancer (TOC) locus: more than just a familial cancer gene. Dis Esophagus **12**: 173-176

Rogers AE, Zeisel SH, Groopman J (1993) Diet and carcinogenesis. Carcinogenesis **14**: 2205-2217

Rosati G, Riccardi F, Tucci A (2000) Use of tumor markers in the management of head and neck cancer. Int J Biol Markers **15**: 179-183

Rose EF (1982) Esophageal cancer in Transkei: The pattern and associated risk factors. In: Pfeiffer CJ, editor, Cancer of the Esophagus. Florida: CRC Press. 19–28.

Rozen S and Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds). Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386

Sammon AM (1992) A case-control study of diet and social factors in cancer of the oesophagus in Transkei. Cancer **69**: 860-865

Sammon AM, Iputo JE (2006) Maize meal predisposes to endemic squamous carcinoma of the oesophagus in Africa: Breakdown of esterified linoleic acid to the free form in stored meal leads to increased intragastric PGE2 production and a low-acid reflux. Med Hypot **67**: 1430-1435

Sammon AM (2007) Carcinogens and endemic squamous cancer of the oesophagus in Transkei, South Africa. Environmental initiation is the dominant factor; tobacco and other carcinogens of low potency or concentration are sufficient for carcinogenesis in the predisposed mucosa. Med Hypot **69**: 125-131

Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucl Acids Res **32**: 91-94

Sargent PJ, Farnaud S, Evans RW (2005) Structure/Function Overview of Proteins Involved in Iron Storage and Transport. Current Med Chem **12**: 2683-263

Segal M, Reinach SG, de Beer M (1988) Factors associated with oesophageal cancer in Soweto, South Africa. Br J Cancer **58**: 681-686

Seidell JC (1997) Time trends in obesity: an epidemiological perspective. Horm Metab Res **29**: 155-158

Selikoff IJ, Hammond EC, Seidman H (1979) Mortality experience of insulation workers in the United States and Canada. Ann NY Acad Sci **330**: 91-116

Shayeghi M, Latunde-Dada GO, Oakhill JS, Laftah AH, Takeuchi K, Halliday N, Khan Y, Warley A, McCann FE, Hider RC, Frazer DM, Anderson GJ, Vulpe CD, Simpson RJ, McKie AT (2005) Identification of an intestinal haem transporter. Cell **122**: 789-801

Sheffield VC, Beck JS, Kwitek AE, Sandstrom DW, Stone EM (1993) The sensitivity of single-stranded conformation analysis for the detection of single base substitutions. Genomics **16**: 325-332

Shephard GS, Marasas WFO, Yazdanpanah H, Rahimian H, Safavi N, Zarghi A, Shafaati A, Rasekh HR (2002) Fumonisin B_1 in maize harvested in Iran during 1999. Food Addit Contam **19**: 676-679

Shimada M, Yanagisawa A, Kato Y, Inoue M, Shiozaki H, Monden M, Nakamura Y (1996) Genetic mechanisms in esophageal carcinogenesis: frequent deletion of 3p and 17p in premalignant lesions. Genes Chromosomes Cancer **15**: 165-169

Siah CW, Ombiga J, Adams LA, Trinder D, Olynyk JK (2006) Normal iron metabolism and the pathophysiology of iron overload disorders. Clin Biochem Rev **27**: 5-16

Siewert JR, Stein HJ, Feith M, Bruecher BL, Bartels H, Fink U (2001) Histological tumor type is an independent prognostic parameter in esophageal cancer: lessons from more than 1,000 consecutive resections at a single center in the Western world. Ann Surg **234**: 360-369

Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL (1987) Human breast cancer: correlation of relapse and survival with amplification of the *HER-2/neu* oncogene. Science **235**: 177-182

Sofic E, Reiderer P, Heinsen H, Beckmann H, Reynolds GP, Hebenstreit G, Youdim MB (1988) Increased iron (III) and total iron content in post mortem substantia nigra of parkinsonian brain. J Neural Transm **74**: 199-205

Sundelöf M, Ye W, Dickman PW, Lagergren J (2002) Improved survival in both histologic types of oesophageal cancer in Sweden. Int J Cancer **99**: 751-754

Syrjanen KJ (1982) Histological changes identical to those of condylomatous lesions found in esophageal squamous cell carcinomas. Arch Geschwulstforsch **52**: 283-292

Takahashi N, Ortel TL, Putman FW (1984) Single-chain structure of human ceruloplasmin: The complete amino acid sequence of the whole molecule. Proc Natl Acad Sci USA **81**: 390-394

Tarmin L, Yin J, Zhou X, Suzuki H, Jiang HY, Rhyu MG, Abraham JM, Krasna MJ, Cottrell J, Meltzer SJ (1994) Frequent loss of heterozygosity on chromosome 9 in adenocarcinoma and squamous cell carcinoma of the esophagus. Cancer Res **54**: 6094-6096

Testa U, Kuhn L, Petrini M, Quaranta MT, Pelosi E, Peschle C (1991) Differential regulation of iron regulatory element-binding protein(s) in cell extracts of activated lymphocytes versus monocyte-macrophages. J Biol Chem **266**: 13925-13930

The International HapMap Consortium (2003) The International HapMap Project. Nature **426**: 789-796

Theil EC (1998) The iron responsive element (IRE) family of mRNA regulators. Regulation of iron transport and uptake compared in animals, plants and microorganisms. Metal Ions Biol Syst **35**: 403-434

Thein SL, Wallace RB (1986) The use of synthetic oligonucleotides as specific hybridization probes in the diagnosis of genetic disorders. Human genetic diseases: a practical approach. Davies KE (ed). Oxford, IRL Press, pp 33-50

Togawa K, Jaskiewicz K, Hiroshi T (1994) Human papilloma virus sequences in esophagus squamous cell carcinoma. Gastroenterology **107**: 128-36

Tondravi MM, McKercher SR, Anderson K, Erdmann JM, Quiroz M, Maki R, Teitelbaum SL (1997) Osteoporosis in mice lacking haematopoietic transcription factor PU.1. Nature **386**: 81-84

Trinder D, Olynyk JK, Sly WS, Morgan EH (2002) Iron uptake from plasma transferrin by the duodenum is impaired in the Hfe knockout mouse. Proc Natl Acad Sci USA **99**: 5623-5626

Tsai FY, Keller G, Kuo FG, Weiss M, Chen J, Rosenblatt M, Alt FW, Orkin SH (1994) An early haemtaopoietic defect in mice lacking transcription factor GATA 2. Nature **371**: 221-226

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res **7**: 996-1005

Vallet VS, Casado M, Henrion AA, Bucchini D, Raymondjean M, Kahn A, Vaulont S (1998) Differentiatial roles of upstream stimulatory facotrs 1 and 2 in the transcriptional response to liver genes. J Biol Chem **273**: 20175-20179

Van Rensburg SJ, Bradshaw ES, Bradshaw D, Rose EF (1985) Oesophageal cancer in Zulu men, South Africa: a case-control study. Br J Cancer **51**: 399-405

Waheed A, Parkkila S, Saarnio J, Fleming RE, Zhou XY, Tomatsu S, Britton RS, Bacon BR, Sly WS (1999) Association of the HFE protein with transferrin receptor in crypt enterocytes of human duodenum. Proc Natl Acad Sci USA **96**: 1579-1584

Wang LD, Shi ST, Zhou Q, Goldstein S, Hong JY, Shao P, Qiu SI, Yang CS (1994) Changes in p53 and cyclin D1 protein levels and cell proliferation in different stages of human esophageal and gastric-cardia carcinogenesis. Int J Cancer **59**: 514-519

Wang JC, Giaever GN (1988) Action at a distance along a DNA. Science **240**: 300-304

Wang M and Marín A (2006) Characterization and prediction of alternative splice sites. Gene **366**:219-227

Wang Z, Xiao X, Van Nostrand E, Burge CB (2006) General and specific functions of exonic splicing silencersin splicing control. Mol Cell **23**: 61-70

Watakabe A, Tanaka K, Shimura Y (1993) The role of exon sequences in splice site selection. Genes Dev. **7:** 407–418

Wei W-Q, Abnet CC, Qiao Y-L, Dawsey SM, Dong Z-W, Sun X-D, Fan J-H, Gunter EW, Taylor PR, Mark SD (2004) Prospective study of serum selenium concentrations and esophageal and gastric cardia cancer, heart disease, stroke, and total death. Am J Clin Nutr **79**: 80-85

Weinberg ED (1999) Iron loading and disease surveillance. Emer Infectious Dis **5**: 346-352

Weiss G, Fuchs D, Hausen A, Reibnegger G, Werner ER, Werner-Felmayer G, Wachter H (1992) Iron modulates interferon-gamma effects in the human myelomonocytic cell line THP-1. Exp Hematol **20**: 605-610

Weitkamp LR (1983) Evidence for linkage between the loci for transferrin and ceruloplasmin in man. Ann Hum Genet **47**: 293-297

Williamson AL, Jaskiesicz K, Gunning A (2001) The detection of human papillomavirus in oesophageal lesions. Anticancer Res **11**: 263-265

Wingender E, Chen X, Fricke E, Geffers R, Hehl R, Liebich I, Krull M, Matys V, Michael H, Ohnauser R, Pruss M, Schacherer F, Thiele S, Urbach S (2001) The TRANSFAC system on gene expression regulation. Nucl Acids Res **29**: 281-283

Xiao W and Oefner PJ (1992) Denaturing high-performance liquid chromatography: a review. Hum Mutat **17**: 439-474

Yang F, Lum JB, McGill JR, Moore CM, Naylor SL, vanBragt PH, Baldwin WD, Bowman, BH (1984) Human transferrin: cDNA characterization and chromosomal localization. Proc Natl Acad Sci USA **81**: 2752-2756

Yang CS, Wang ZY (1993) Tea and cancer. J Natl Cancer Inst **85**: 1038-1049

Yang F, Naylor SL, Lum JB, Cutshaw S, McCombs JL, Naberhaus KH, McGill JR, Adrian GS, Barnett DR, Bowman BH (1986) Characterization, mapping, and expression of the human ceruloplasmin gene. Proc Natl Acad Sci USA **83**: 3257-3261

Yang PC, Davis S (1988) Incidence of cancer of the esophagus in the US by histological type. Cancer **61**: 612-617

Yoshizawa T, Yamashita A, Luo Y (1994) Fumonisin occurrence in corn from high- and low-risk areas for human esophageal cancer in China. Appl Environ Microbiol **60**: 1626-1629

Yu MC, Garabrant DH, Peters JM, Mack TM (1988) Tobacco, alcohol, diet, occupation, and cancer of the esophagus. Cancer Res **48**: 3843-3848

Zahler AM, Lane WS, Stolk JA, Roth MB (1992) SR proteins: a conserved family of pre-mRNA splicing factors. Genes and Dev **6**: 837-847

Zahringer J, Baliga BS, Munro HN (1976) Novel mechanism for translational control in regulation of ferritin synthesis by iron. Proc Natl Acad Sci USA **73**: 857-861

Zhan Q, Carrier F, Fornace AJ (1993) Induction of cellular p53 activity by DNA-damaging agents and growth arrest. Mol Cell Biol **13**: 4242-4250

Zhang H, Chen S-H, Li Y-M (2004) Epidemiological investigation of esophageal carcinoma. World J Gastroenterol **10**: 1834-1835

Zhang L, Vincent GM, Baralle M, Baralle FE, Anson BD, Benson DW, Whiting B (2004) An intronic mutation causes long QT syndrome. J Am Coll Cardiol **44**: 1283-1291

Zhou X, Tarmin L, Yin J, Jiang HY, Suzuki H, Rhyh MG, Abraham JM, Meltzer SJ (1994) The MTS1 gene is frequently mutated in primary human esophageal tumors. Oncogene **9**: 3737-3741

Zucker S, Freidman S, Lysik RM (1974) Bone marrow erythropoiesis in the anemia of infection, inflammation, and malignancy. J Clin Invest **53**: 1132-1138

5.2 ELECTRONIC DATABASE INFORMATION

BLAST: http://www.ncbi.nlm.nih.gov/blast

Ensembl: http://www.ensembl.org

ESEfinder: http://exon.cshl.edu/ESE

GenAtlas: http://www.genatlas.org

HapMap: http://www.hapmap.org

JASPAR CORE: http://www.jaspar.genereg.net

PATCH: http://www.gene-regulation.com/cgi-bin/pub/programs/patch/bin/patch.cgi

Primer3: http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi,2002

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

STATISTICA [StatSoft, Inc (2003) STAT (data analysis software system)], version 6

APPENDIX 1: LIST OF CHEMICALS/REAGENTS USED IN THIS STUDY AND THEIR SUPPLIERS

CHEMICAL/REAGENT	SUPPLIER
AA	Sigma-Aldrich
Agarose	Laboratory Specialist Services
APS	Merck
BAA	Sigma-Aldrich
$C_{13}H_{28}N_2Na_4O_{13}S$	Fluka
$C_{19}H_{10}Br_4O_5S$	Merck
Cresol Red	Merck
ddH_2O	Adcock-Ingram
dNTPs (dATP, dTTP, dCTP, dGTP)	Fermentas
EDTA	Merck
EtBr	Merck
EtOH	Merck
H_3BO_3	Kimix
H_2NCHO	Sigma-Aldrich
$MgCl_2$	Bioline
NaCl	Fluka
$(NH_2)_2CO$	Sigma-Aldrich
SfcI	New England Biolabs
Sucrose	Associated Chemical Enterprises
Taq polymerase	Bioline
TEMED	Fluka
Tris-HCl	Fluka

APPENDIX 2: PROMOTER AND CODING REGIONS OF THE *CP* GENE INDICATING PRIMER BINDING POSITIONS AND VARIANTS IDENTIFIED IN THIS STUDY

The promoter (5' to 3') and coding regions (5' to 3') of the *CP* gene were acquired from Ensembl and GenAtlas (reference numbers indicated). The primers designed for the amplification of the promoter region are highlighted in various colours which are described in the key. Primers designed for the amplification of the coding regions are indicated using various text colours. Red text indicates the forward primer of the first primer set of a particular fragment and blue text indicates the reverse primer. In cases where a particular fragment has been split in two for amplification purposes (see Table 2.4), green text indicates the forward primer of the second primer set and purple denotes the reverse, with the first primer pair the same as described above. Arrows flank the primers with arrowheads indicating the orientation of each primer. The translation initiation site (ATG) and the stop codon (TGA) are indicated in bold red text. The variations that were detected in the current study are in bold text, highlighted in grey and the superscript number indicates the nucleotide position (described in the key). The 5'UTR and the introns are indicated in lowercase text while the mRNA encoding regions (transcribed regions) are indicated in uppercase.

_CP_ promoter sequence

[ENSG00000047457 (Ensembl)]

aagactaatctttgcctctgaagacaggtttagttatggtttgtttatttgatgattaaaacaaagttattagcc
cctgttaggctctgctagttatttgcatatagcaatacagaatgatctggtaaagtttttatgcagatttggtcc
aagctcataaaacagataaattcttccaaattaaattcttatttttgtttacttaaatttggcatgctagatatc
atgtttcatatcataaagctgcaattaatcatttgttagacttttagaaatagtcatgcaccacataataatgtt
tgagtaaatgacaaactgcatatgtgatggtggtcccataagattataataacatatttttactgtacattttct
atgtttgcatatgtatagatacagaaatatcattgtgttacaattgtctatagcattcagcatagttatagactg
tacaggtttgtagcctaggagcaataggtatacc[1]atat[2]agcctacctagatgtatagtaggctatatcatata
agtctgtgtaagtacactctgtgatagtcacacaatgacaaaatagcctaacaatgcatttctcagaacgtatcc
ctgtcattaagaac[3]gactgtagtaaaaagctacatgttttgattcattttacataattcttatcctgatttact
aaaagtgcactgttt[4]gcaggacact[5]gtgc[6]tgtgacatattgtcagaggaacaaagggagcatttagacacg[7]t
tctctgccctcctggaatttacaca[8]atgaatggaaaaaggcacagagttatgcacaccctaatgcctccaacaa
taactgttgacttttttatttcagtcagagaagcctggcaaccaagaactgtttttttggtggtttacgagaact
taactgaattggaaaatatttgctttaatgaaacaatttactcttgtgcaacactaaattgtgtcaatcaagcaa
ataaggaaagtcttatttataaaattgcctgctcctgatttacttcatttcttctcaggctccaagaaggggaa
aaaaatgaagattttgatacttggtatttttctgtttttatgtagtaccccagcctgggcgaaagaaaagcatta
ttacattggaattattgaaacgacttgggattATGAAGATTTTGATACTTGGTATTTTTCTGTTTTTA
TGTAGTACCCCAGCCTGGGCGAAAGAAAAGCATTATTACATTGGAATTATTGAAACGACTTG
GGATTATGCCTCTGACCATGGGGAAAAGAAACTTATTTCTGTTGACACgtaagtcactattt
ttattgtttatagaccaa

_CP_ promoter region key:

Primers	
Amplicon 1	CPP1F & R
Amplicon 2	CPP2F & R
Amplicon 3	CPP3F
Amplicon 4	CPP4F
Amplicon 5	CPP5R
Amplicon 6	CPP6R
Amplicon 3 & 5	CPP3R & CPP5F
Amplicon 4 & 6	CPP4R & CPP6F

Variant	
1	5'UTR-567C→G
2	5'UTR-563T→C
3	5'UTR-439C→T
4	5'UTR-364delT
5	5'UTR-354T→C
6	5'UTR-350C→T
7	5'UTR-308G→A
8	5'UTR-282A→G

_CP_ coding regions

[NM_000096 (GenBank)]

Exon 1

ttttttggtggtttacgagaacttaactgaattggaaaatatttgctttaatgaaacaatttactcttgtgcaac
actaaattgtgtcaatcaagcaaataaggaagaaagtcttatttataaaattgcctgctcctgattttacttcat
ttcttctcaggctccaagaaggggaaaaaaaATGAAGATTTTGATACTTGGTATTTTTCTGTTTTTATGTAGTACC
CCAGCCTGGGCGAAAGAAAAGCATTATTACATTGGAATTATTGAAACGACTTGGGATTATGCCTCTGACCATGGG
GAAAAGAAACTTATTTCTGTTGACACgtaagtcactattttattgtttatagaccaaaattttaagttatttttt
aaaggatagagccaataagatgaaatatttttaaaatttcttatatacagaaactgcaaaatgtgggcttgaaac
ccaagacacttgaatcaatctcctttatattcctgtatatttttttatttggtggac

Exon 2

taaaataacagggctaactaccagccctccacttcaatttggaggcatccctacaacaggcaaacattcaatgt
gatgtgctcattaatttctctaattataatgttattaaattataatgataatgtcatccctagaaatgatggctt
ctgatagagttgtcttgtttttctttgcagGGAACATTCCAATATCTATCTTCAAAATGGCCCAGATAGAATTGG
GAGACTATATAAGAAGGCCCTTTATCTTCAGTACACAGATGAAACCTTTAGGACAACT[1]ATAGAAAAACCGGTCT
GGGCTTGGGTTTTTAGGCCCTATTATCAAAGCTGAAACTGGAGATAAAGTTTATGTACACTTAAAAAACCTTGCCT
CTAGGCCCTACACCTTTCATTCACATGGAATAACTTACTATAAGGAACATGAGGgtaagttcagctcatgtagct
ggaatttcctcttactttcagtagaagtgccttttagcttctaaaattgtggacttctctcctggataagacat
ataactgccattcagctcttaactatttttttttatttttttttaattttttttgacttctgtcacccaggctg
gagtgcagt

Exon 3

cttctacacaccctagccattagcctttcattttcctgtgtccaaataggataacacatccccaaggatcacata
gtagatgtttaataaatgtttgctgaatgaattaatggactatctattctctcatttaagttcaaacacattcac
tttgcaccttcattgcatgttgcttcctagGGGCCATCTACCCTGATAACACCACAGATTTTCAAAGAGCAGATG
ACAAAGTATATCCAGGAGAGCAGTATACATACATGTTGCTTGCCACTGAAGAACAAAGTCCTGGGGAAGGAGATG
GCAATTGTGTGACTAGGATTTACCATTCCCACATTGATGCTCCAAAAGATATTGCCTCAGGACTCATCGGACCTT
TAATAATCTGTAAAAAAGgtacatcttctccttattgcacatgctatatgaccaaaagacaggggcagggcagtt
tgtcttttaaaaagaaatcaagtcttctgtgctgagattgagcaggtaaggtagagaggtggtaagtagaaaaaa
aggaagggaagagccaggcaggactaggagaagaagggataaatagaa

Exon 4

tactgatcataaaaattccctgagtgtttctgtctgtaaaacaggagtgactggtccagacttgtgatttattcg
aaaaaataaatactaacatttttaaaaactataaatcacctgaaaacataaatcaaaatacatgctaaaagttta
gtcttgctaaaatagtaactttaactccag**ATTCTCTAGATAAAGAAAAAGAAAAACATATTGACCGAGAATTTG**
TGGTGATGTTTTCTGTG[2]GTGGATGAAAATTTCAGCTGGTACCTAGAAGACAACATTAAAACCTACTGCTCAGAA
CCAGAGAAAGT[3]TGACAAAGACAACGAAGACTTCCAGGAGAGTAACAGAATGTATTgtaagattacagcagacat
gttaactttacttttttttcccttgatccattaattctgctaagcttattccctcacataaataagtactgtgtg
tctagtctgtggtccttataacaaagcatattggatgtgcaatttgatattcagctaagtaagtcagaataaaaa
ctctaagccag

Exon 5

ttccaaaatgaaagagtgctgtgattgattagcaatgtctgacatgagtatagggggcaagaataccagcatgtg
tgccagatatttgccattccaaattgattctatttatataggaggtaactctcactatctcaagcattaatgactc
ttagtcacaaactctgc[4]tcttgacttacag**CTGTGAATGGATACACTTTTGGAAGT**CTCCCAGGACTCTCCATG
TGTGCTGAAGACAGAGTAAAATGGTACCTTTTTGGTATGGGTAATGAAGTTGATGTGCACGCAGCTTTCTTTCAC
GGGCAAGCACTGACTAACAAGAACTACCGTATTGACACAATCAACCTCTTTCCTGCTACCCTGTTTGATGCTTAT
ATGGTGGCCCAGAACCCTGGAGAATGGATGCTCAGCTGTCAGAATCTAAACCATCTGAAAGgtaggacatcttta
cctgaaattagcagtcagctgaaaaaggtggttactgggaatcagaatctaaaagctaaatatctttgaaataat
ctgagctttgaactggcatcttagttttatccctatgatggtatatataataatgtaataaaaaaagactgtaag
acaggaatattataaa

Exon 6

ttttaaacttgagttttccatcattaatttatttacagctgtctgccctagtttatagtactgttgctatctgtc
tattccttcaatttctacttgtcatcacccgagcagtgtttacagggcaataggagagagaaacattgtcttcat
taatcaatgcctctttctgtttcatttcag**CCGGTTTGCAAGCCTTTTTCCAGGTCCAGGAGTGTAACAAGTCTT**
CATCAAAGGATAATATCC[5]GTGGGAAGCATGTTAGACACTACTACATTGCCGCTGAGGAAATCATCTGGAACTAT
GCTCCCTCTGGTATAGACATCTTCACTAAAGAAAACTTAACAGCACCTGGAAGgtaatttacaaaaaaagagtac
tatttcaaaatatgctctcccccgcacaaaggaactacaaaaggctaattcagcaattaaaattttaattggtat
ctgaaattgagtaactagactgctaaatatttattaaggtatttgatgtaagagtctgagcttgagaactatgag
ttttgagt

Exon 7

ctatgagtggactggaactgtctgctttgtttgagtaactggtacggattgctataaatcatggataatattcat
ctctaagcactggtcagaatttttctcttcaattagcaccttaagtgaatacctgctttctgttctacttcaacc
ccagcattaaacacttttttccccctgcag**TGACTCAGCGGTGTTTTTTGAACAAGGTACCACAAGAATTGGAGG**
CTCTTATAAAAAGCTGGTTTAT[6]CGTGAGTACACAGATGCCTCCTTCACAAATCGAAAGGAGAGAGGCCCTGAAG
AAGAGCATCTTGGCATCCTGGgtgagttct[7]tgcagaatatatatggctggtttactcttcccatgggctacatc
aacctcagaaagaatagtaacttcaggacctactgcatgatttctatgtgttaagtaggcgtgtggaagataaaa
caataaaatgcagcagctaagtactttccaaatgtaaattaccaatggtaa

Exon 8

ccaacctcgacatatcagattgattatggacttctcaaaaaccaagctttccattcaccacaaataagcgtgttg
ctcctagtagttcttgccagagaacacagtcaaagaagataactatcttatacattctgcatgttacataccaag
gaaggataataattctctctttgttcccag**GTCCTGTCATTTGGGCAGAGGTGGGAGACACCATCAGAGTAACCT**
TCCATAACAAAGGAGCATATCCCCTCAGTATTGAGCCGATTGGGGTGAGATTCAATAAGAACAACGAGGGCACAT
ACTATTCCCCAAATTACAACCCCCAGAGCAGAAgtgagtatacagcactgtgtaacaaatcaacaatgtttaata
acccactgtactggtcactgacagtatgctgcaaaaactgaaggcaggaactcccaaggaaaccgctcatatatc
atttctaggcatcatgccaggaaacaaataaacccttctcctacctccaattggtttacgga

Exon 9

ctatctggatgcaatcacatcggtctgttatctgacaccagagcataagtgggattccattctaatttataacaa
agaactcagaagtgggctattctagactaaacttagtctactttgtggctcaaatgaccacgttagaaagcacac
ttcacctctctaatgtgacctttctcacag**GTGTGCCTCCTTCAGCCTCCCATGTGGCACCCACAGAAACATTCA**
CCTATGAATGGACTGTCCCCAAAGAAGTAGGACCCACTAATGCAGATCCTGTGTGTCTAGCTAAGATGTATTATT
CTGCTGTGGAT[8]CCCACTAAAGATATATTCACTGGGCTTATTGGGCCAATGAAAATATGCAAGAAAGGAAGTTTA
CATGCAAATGGGAGACAGgtaagtccaactggggaaaaaaaaattgttcaatctctttgaaaatgatcatttgac
agaaatagaaatatctaacctcatcattatctccaaaaatactttagtatataccttaaaagataaggacttct
ttaaaaatgcccacaattatcacaccttaaaataataataacacacat

Exon 10

taccaaagaaagctctggtggtgaactgaaatatttcaggaaccctagtttacaaagagcaacagaaacttctgt
tctcattctttcctttaggtgtgtttgtgcataaatagtattgtgcacatggaagtcttctgctttattcattct
gtctccatctcattttgttttgatttacag**AAAGATGTAGACAAGGAATTCTATTTGTTTCCTACAGTATTTGAT**
GAGAATGAGAGTTTACTCCTGGAAGATAATATTAGAATGTTTACAACTGCACCTGATCAGGTGGATAAGGAAGAT
GAAGACTTTCAGGAATCTAATAAAATGCACTgtaagtactgcatcatccaccaatctaccaagttattttagtac
tgcatatgtaatgcttttaactaaacaaaatgtgaagagctttgctcattgattatttttcaataaacgattac
aatagtgtcctgaaggtaatctcactggttatttgatgtgtctgtcagttatccattattt

Exon 11

tgctctgaaatgccaggaaccacaaggagttgcaaggatactcctgacacataaagcgtgggcaagcctgaatag
gtcttttaaaaattaaaaaaacaaagttggtcctggaaagtctgtgaggttgtcatgtagtagaatctaaaccac
ttccaaaatgttatttcccaacttttacag**CCATGAATGGATTCATGTATGGGAATCAGCCGGGT[9]CTCACTATG**
TGCAAAGGAGATTCGGTCGTGTGGTACTTATTCAGCGCCGGAAATGAGGCCGATGTACATGGAATATACTTTTCA
GGAAACACATATCTGTGGAGAGGAGAACGGAGAGACACAGCAAACCTCTTCCCTCAAACAAGTCTTACGCTCCAC
ATGTGGCCTGACACAGAGGgtatatttacttaaagccaagtctctattggatggatttgtgtagttttggtttaa
aatcctgtagcagaatgtgttgggtgataaacttatcccttccccaagcctctgaagtgcccttcctatctttga
aagagcaagtctaggccactatcagctgaagtatgtaccagataggtct

Exon 12

gcacacacttgatctgctctaagaacagcaaaaggatggatggagcagggtgagcagaaggtgtggaaagctaga
tcacctgaggccatgataaggacttcagcattaattctgaggaaggtggagagcccagagggcactagcaaaga
agtgacatgttttcttatttaccactccag**GGACTTTTAATGTTGAATGCCTTACAACTGATCATTACACAGGCG**
GCATGAAGCAAAAATATACTGTGAACCAATGCAGGCGGCAGTCTGAGGATTCCACCTTCTACCTGGGAGAGAGGA
CATACTATATCGCAGCAGTGGAGGTGGAATGGGATTATTCCCCACAAAGGGAGTGGGAAAAGGAGCTGCATCATT
TACAAGAGCAGAAgtaattctccagactctgattcccaatttcagcagtaattttaggtttccttttcagagaag
taagaattacaccaaattttagaaaaaaaaaaaaagttgcattaaacaataacaacaacaacaacaacaacaacaa
aacacttcctaggaacttgtccttattcatttccgcttctctc

Exon 13

ttctaccacacagacttttaaagtcttattatgcattcctacttggatagaagagtgatgatttaaaactactat
aattactctaataaaaactatatttttattcaaaatttaaagtttaaaacttaagttacatttcattaagacaa
atactaaccttagttgtttttttttttaag**TGTTTCAAATGCATTTTTAGATAAGGGAGAGTTTTACATAGGCTC**
AAAGTACAAGAAAGTTGTGTATCGGCAGTATACTGATAGCACATTCCGTGTTCCAGTGGAGAGAAAAGCTGAAGA
AGAACATCTGGGAATTCTAGgtatgttaatatctccagtttatcaggtgtgggttaatgatgtgatgctacctgc
agtaaccgtcaattcatgtctatgggtttcattttcttggttcatatatcaaagttgttcaaataaaaatttaaa
aattagttatctaatgaaataagactgatgtttattaaataatatgacat

Exon 14

aaatggtctccatttttatttccagtgtattgattccaaattccctgggacgcagtataaatccaagagtattat
accacaaaccccatgggggagagaaacaagttatctttgctgatctttgcttttattactaagttagaccttctga
ttggacactcttcctcttcacttttgccag**GTCCACAACTTCATGCAGATGTTGGAGACAAAGTCAAAATTATCT**
TTAAAAACATGGCCACAAGGCCCTACTCAATACATGCCCATGGGGTACAAACAGAGAGTTCTACAGTTACTCCAA
CATTACCAGgtactcacgggggcgggggtggggggtggtgcattttaacaggctgtacagcctttccacagattac
tctcatcatagagagtctggaccaagaaagtgaaaagggaggtagttgtgaagaggggatgcaaggaggtgtctg
tgtgtatgcaagtacatgcacatgcttgtactaaaaaaa

Exon 15

aaactaaattcctgaaaacaatcttgagtaccacaatttttaaaaactaattacactgacatttccagttcctgg
gtcctagttatcaaggctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtacaatataaaccaaataca
accagatccattaacacttgatgaatttag**GTGAAACTCTCACTTACGTATGGAAAATCCCAGAAAGATCTGGAG**
CTGGAACAGAGGATTCTGCTTGTATTCCATGGGCTTATTATTCAACTGTGGATCAAGTTAAGgtaaaatacaaat
catttgttttgtgacacttcccaaaatagacacacaattgaaggaagacaaggttagtgttttcctgtggtccaa
ttctacgttaaaatcctaggaaaataaactgtttgcaatctactttagcaactatgctgctacctatgacacaca
agacaaaaatctaacat

Exon 16

ccattgttccaatttaactacttcaagtgaaaagttactaaattggcagtaccacattgggcaaagtctgttact
tcaattggtctctccaaaattcctgaggcctgggtgcaaagtctcagttctctgaaatcctgacctaattcaca
agggttactgaagattttt^{10}cttgtttccag**GACCTCTACAGTGGATTAATTGGCCCCCTGATTGTTTGTCGAA**
GACCTTACTTGAAAGTATTCAATCCCAGAAGGAAACTGGAATTTGCCCTTCTGTTTCTAGTTTTTGATGAGAATG
AATCTTGGTACTTAGATGACAACATCAAAACATACTCTGATCACCCCGAGAAAG**TAAACAAAGATGATGAGGAAT**
TCATAGAAAGCAATAAAATGCATGgtatgtcacattattctaaaacaatttttctacttttttatcattttctct
ttcaaataaaattattttggagaccattcatttgttttgtcttaaaataaaccacttctgcctcactttgtatgt
caggtagaatcctgttcaaaacaaaaaggggactatgttcaaaatcatttgcta

Exon 17

gcctagaaaagcaatgtgatattattgactttaaagtatccatttctctcatttgtagtaaaataagtcctatta
catcaaaaagattctttacattaacctaataatcatattttctttgaaatagctattggtttaaaatatcctgaa
aagtaacataaaaccatatgctttccctag**CTATTAATGGAAGAATGTTTGGAAACCTACAAGGCCTCACAATGC**
ACGTGGGAGATGAAGTCAACTGGTATCTGATGGGAATGGGCAATGAAATAGACTTACACACTGTACATTTTCACG
GCCATAGCTTCCAATACAAGgtaagagctatccatggcaattactcttgctctgtttgaaaatgttttaatacaa
gtgaagaaaatattttctcagagagacttatgaaaaaaagctgcaaagagtacagccggtgcttcatatccgtgg
attccccacccaggattcaatcatggattgaaaatattcaggaaaaaata

Exon 18

aagacttattaggaatacaaaaaccacattttaccatttgtattaaacacagaaacgagctcgatttttaaaaa
cagatttatccatttttcaatatccctgagctgaacattaacaaatgacatgagtgaactctcatatacaatat
tatacattatttgtcatttttctaattaag**CACAGGGGAGTTTATAGTTCTGATGTCTTTGACATTTTCCCTGGA**
ACATACCAAACCCTAGAAATGTTTCCAAGAACACCTGGAATTTGGTTACTCCACTGCCATGTGACCGACCACATT
CATGCTGGAATGGAAACCACTTACACCGTTCTACAAAATGAAGgtgaatatccaggtagtaattctagaagccat
atataaaagatataactaaaggaactactttgcctaaatacttcccccaatatgatgcatcaatatgaagcattt
gtgaatctgctgccctgatttcttccttctctacctccaaacaggatttttttcctgctaaatggtgagggat

Exon 19

atctttctttaaataaagcagcaatagaatgagagattttaaacatataaaaagcagaaaaatataaatgaatc
aggagtaaagtaacattaaaaattggaatacatttaaagaaccatctactaatttctaaccatatattatttta
ataatggtcttaaaatttctttttctatag**ACACCAAATCTGGCTGA**ATGAAATAAATTGGTGATAAGTGGAAAA
AAGAGAAAAACCAATGATTCATAACAATGTATGTGAAAGTGTAAAATAGAATGTTACTTTGGAATGACTATAAAC
ATTAAAAGAAGACTGGAAGCATacaactttgtacatttgtgggggaaaactattaatttttgcaaatggaaaga
tcaacagactatataatgatacatgactgacacttgtacactaggtaataaaactgattcatacagtctaatgat
atcaccgctgttagggttttataaaactgcatttaaaaaaagatctatgacc

***CP* coding region key:**

Primers	
Forward Primer 1	Reverse Primer 1
Forward Primer 2	Reverse Primer 2

Variant	
1	T83
2	V223
3	V246A
4	IVS4-14C→T
5	R367C
6	Y425
7	IVS7+9T→C
8	D544E
9	G633
10	IVS15-12T→C